Forecasting the Post-Pandemic Public Transit Recovery in Toronto

University of Toronto — Rotman School of Management

J. Rapson

Abstract

Necessary restrictions on movement and gatherings during the COVID-19 pandemic have had major ramifications for public transit providers. Transit agencies that predominately rely on user-fees generated by fare revenue will require external funding to maintain operations until public health measures are able to be retracted. These agencies have a vested interest in obtaining a highly accurate prediction of transit demand under different potential courses of the pandemic. This paper uses data from the popular mobile Transit app to produce several forecasting models of public transit demand in Toronto during the COVID-19 pandemic. The final model predicted changes in transit demand from pre-pandemic levels with a mean absolute error of 3.9% in an ex-post forecast, generating highly accurate predictions of transit behaviour.

The COVID-19 pandemic has significantly affected public transit operations worldwide, with agencies experiencing up to 90% drops in ridership during initial lockdowns (Liu, Miller, & Scheff, 2020). For transit systems that heavily depend on user fees for operational revenue, such as the Toronto Transit Commission (TTC), these drops render pre-pandemic financing schemes unsustainable and make such systems dependent on external financing to maintain operations (Tyler, 2021). In 2020 alone, the TTC estimates that account revenue losses and public health expenditure related to COVID-19 had a gross financial impact of approximately $745 million on the system, necessitating federal and provincial intervention (Toronto Transit Commission, 2020). It is highly probable that sustained public health measures to prevent the spread of COVID-19 will increase the financial burdens placed on public transit systems. Operators must plan for future dependence on external funding or alternative financing schemes until pandemic-related restrictions are lifted.

Changes in transit demand can be difficult to ascertain at a granular level as most agencies only publish ridership figures at the monthly or yearly level. Data from third party sources, such as transit planning applications, can be used to fill these data gaps. The popular phone-based application, Transit app, has published regularly updated data on hourly and daily changes in app usage for 95 metropolitan areas and 168 individual transit systems around the world, enabling researchers to estimate the impact of public health measures on transit demand (Transit App, 2020). Changes in transit demand from the pre-pandemic baseline are measured as a percent change in app opens compared to the same day of the week one year prior to the COVID-19 pandemic. Activity from these three days is averaged over three weeks and adjusted for year-over-year patterns such as yearly growth in the relevant agency, metropolitan region, or country.

Overall app usage for Transit app dropped to 77% below pre-pandemic levels in April 2020 (Transit App, 2020). Through a survey of 25,000 Transit app users, it was found that the vast majority of those continuing to use the app are essential workers.

These users are disproportionately women, black, and Latino. Prior to the crisis, Transit had an approximate even split between male and female users. In the immediate aftermath of the shutdowns, 56% of remaining users were female and 40% were male. The app has also seen a decrease of almost 50% of its white users. The app finds that of those still riding transit, very few (10%) report travelling for leisure reasons, and even fewer (2%) report traveling to get to school. Instead, 92% of users report using public transit to commute to work, including almost 20% in food preparation and almost 20% for healthcare. Further, the app found that over 70% of pandemic transit users make less than $50,000 per year, a substantial increase from previous years. Such data reveals a dramatic shift in the profile of transit users during the COVID-19 pandemic.

Transit app data has also allowed researchers to determine temporal trends in transit demand post-crisis (Liu, Miller, & Scheff, 2020). Liu, Miller, and Scheff (2020) identify base values that transit demand does not fall below, indicating the minimal level of transit use that would be expected regardless of pandemic circumstances. This demand, as support by the Transit app survey, likely comes from essential workers continuing to use transit for work-related purposes. Liu, Miller, and Scheff (2020) also identify a cliff point where discretionary public transit demand — mostly those who have the ability to work from home or access alternative means of transportation — begins to decline. Interestingly, the authors find that this decline is not synchronized with COVID-19 case rates, but with the date when a state of emergency was declared for each state. The authors further identify a base point, the time when transit decline stabilizes, and a decay rate that represents the speed of the decline process. The work done by Liu, Miller, & Scheff, 2020 indicate the presence of predicable trends in transit demand.

As the COVID-19 pandemic continues, there is a pressing need for fare-dependent transit agencies to have access to a reliable forecast of transit demand, conditional on various public health scenarios. Such a model would enable transit agencies to understand the impact of various pandemic measures on transit demand and predict how this demand will respond to different future states. This paper explores several forecasting models for public transit demand in Toronto using Transit app data.

Methodology

A forecast of public transit demand in Toronto was developed by modelling historic responses to various public health measures, unemployment rates, and COVID-19 case rates using single equation regression. Two forecast models were built using an optimal combination of these features and applied to a holdout set of the most recent 100 days of transit demand data to validate model accuracy in an ex post forecast. The first model was produced using all available Transit app data starting in February 15, 2020; while the second was estimated using only data from after the cliff point where Toronto public transit demand rapidly decreased on March 17, 2020. This was done to identify whether structural breaks following the rapid reduction in transit use harm forecast accuracy. A third regression-based model was also build using all available Transit app data for Toronto demand. Finally, the regression models were compared to a baseline forecast built using only autoregressive integrated moving average (ARIMA) techniques. All analysis was completed in R 4.0.5 (R Core Team, 2021).

This section details the variables used in the final transit demand forecasts, the processes of building and validating the single equation regression models, as well as the process of building and validating the ARIMA baseline forecast.

Correlations between changes in Toronto public transit demand from the pre-pandemic baseline were explored for 17 variables, 9 of which were categorical dummies, mostly representing various public health measures, and 9 of which were continuous measures of unemployment or COVID-19 case rates. A complete summary of the variables explored in the model-building process is given in Table 1.1.

Table 1.1 — Summary of Variables Explored

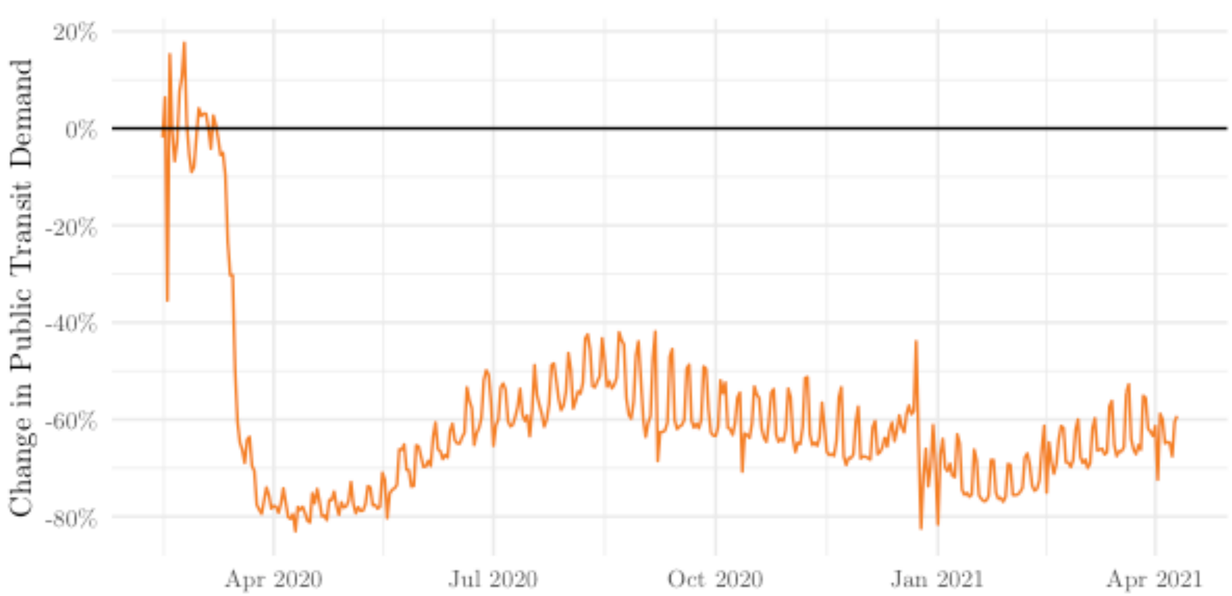
Variable Name	Type	Explanation
`Toronto`	Continuous	Percent change in Toronto public transit demand from pre-pandemic baseline
`State.of.Emergency`	Categorical	Presence of provincial State of Emergency
`School.Closures`	Categorical	Presence of Toronto school closures
`Gathering.Restrictions`	Categorical	Presence or degree of gathering restrictions in Toronto, three possible levels
`Travel.Restrictions`	Categorical	Presence of international travel restrictions
`Business.Closures`	Categorical	Presence or degree of mandated business closures in Toronto, three possible levels
`Lockdowns`	Categorical	Presence or degree of lockdowns in Toronto and Ontario, three possible levels
`Park.Closures`	Categorical	Presence of Toronto park closures
`Mandatory.Masks`	Categorical	Presence or degree of mandatory mask policies in Toronto and Ontario, two possible levels
`Weekend`	Categorical	Dummy variable for whether day was a weekend
`More.Than.Year`	Continuous	Percent change in Ontario workforce that has been unemployed for over a year from pre-pandemic baseline
`More.Than.3.Months`	Continuous	Percent change in Ontario workforce unemployed for over three months from pre-pandemic baseline

`Official.Rate`	Continuous	Percent change in official Ontario unemployment rate from pre-pandemic baseline
`Plus.Discouraged`	Continuous	Percent change in official Ontario unemployment rate from pre-pandemic baseline, plus discouraged searchers
`Plus.Waiting`	Continuous	Percent change in official Ontario unemployment rate from pre-pandemic baseline, plus those waiting on recall, replies, or long-term future starts
`Plus.Involuntary.Part.Time`	Continuous	Percent change in official Ontario unemployment rate from pre-pandemic baseline, plus those who are involuntarily employed as part-time workers
`Plus.All`	Continuous	Percent change in official Ontario unemployment rate from pre-pandemic baseline, plus discouraged searchers, waiting individuals, and involuntary part-time workers
`Total.Reported`	Continuous	Total COVID-19 case rates reported
`Total.Episode`	Continuous	Total COVID-19 cases contracted, produced through retroactive estimation from reports and contact tracing

Output Variable

Transit app data on changes in public transit demand from pre-pandemic rates, approximated by the change in the number of app opens, was used as a measure of transit de-

Figure 1.1 — Change in Toronto Public Transit Demand from Pre-Pandemic Baseline



mand in Toronto. As can be seen in Figure 1.1, the cliff point, or start of the drop in transit demand, occurs approximately on March 17, 2020. This date corresponds with the first state of emergency declared in Ontario, in line with transit demand responses observed by Liu, Miller, and Scheff (2020). Transit demand reaches a base value of 83% below pre-pandemic levels on April 10, 2020. Following this date, demand slowly increases until the fall of 2020. Throughout the entirety of the available data, there are observable weekly spikes in demand. These correspond to weekends, and will be discussed below. Finally, there is a notable spike on December 23, 2020 that corresponds with the day before Christmas Eve, when a second provincial lockdown was announced.

Categorical Input Variables

The most apparent impact on transit demand, supported by Liu, Miller, and Scheff's (2020) findings are the public health measures implemented at the provincial and municipal level. Eight of these measures are coded as dummy variables to represent their presence and severity, shown in Table 1.2 The effect of weekends is also explored.

As the COVID-19 pandemic is a rapidly evolving situation, it is very difficult to accurately forecast future cases rates and public health measure implementations. It is also beyond the score of this paper to make epidemiological claims about the progression of the pandemic. Thus in order to produce valid ex-ante forecasts, public health measures can be treated conditionally to enable transit agencies to forecast public transit demand under different scenarios. This enables forecasts to prepare for "worst case" public health outcomes with strict public health measures.

Table 1.2 — Coding Scheme for Public Health Measures

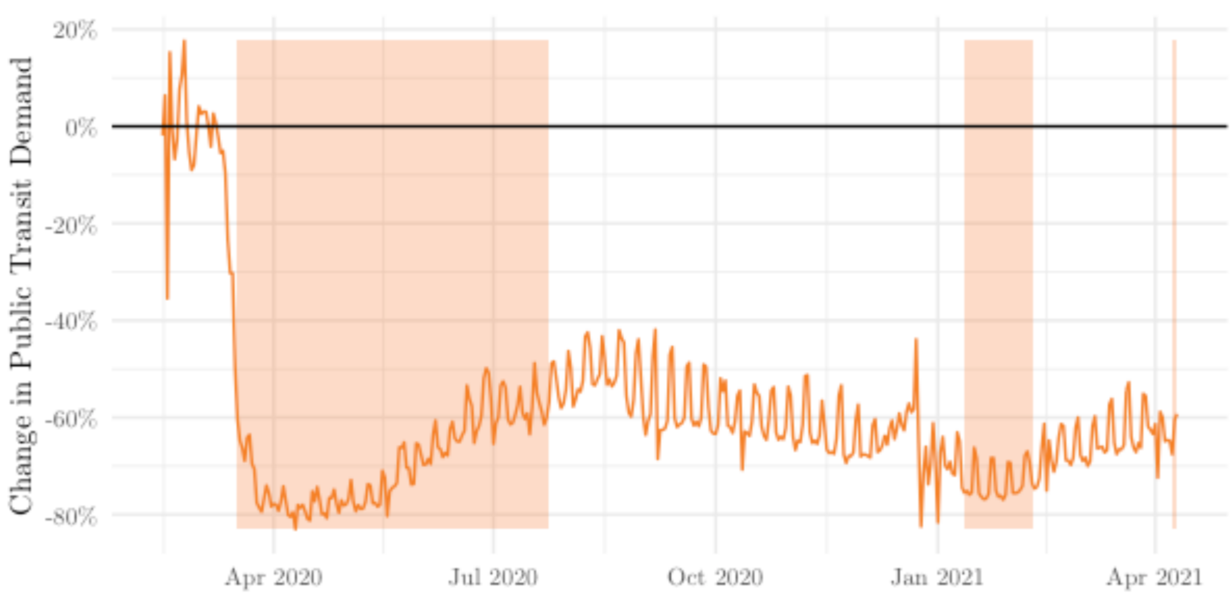
Public Health Measure	Coding
State of Emergency	0 = No state of emergency
	1 = Provincial state of emergency
School Closures	0 = No full municipal school closures
	1 = Full municipal school closures
Gathering Restrictions	0 = No gathering restrictions
	1 = Maximum 50-250 people indoors
	2 = Maximum 10 people indoors
	3 = Maximum 5 people or less indoors
Travel Restrictions	0 = No international travel ban
	1 = International travel ban

Business Closures	0 = No restrictions on business operations
	1 = Almost no restrictions on business operations
	2 = Multiple or severe restrictions on business operations
	3 = Mandatory closure of all non-essential businesses
Lockdowns	0 = No lockdowns
	1 = Either provincial or municipal lockdown
	2 = Both provincial and municipal lockdown
	3 = Mandatory stay-at-home order
Park Closures	0 = No provincial or municipal park closures
	1 = Provincial or municipal park closures
Mandatory Masks	0 = No provincial or municipal mask mandates
	1 = Either provincial or municipal mask mandates
	2 = Both provincial and municipal mask mandates

(a) State of Emergency

Liu, Miller, and Scheff's (2020) find that the most severe public transit demand drop occurs upon the declaration of a statewide state of emergency. This is largely true for Toronto, with the first major drop in transit demand occurring on March 17, 2020, the date when Ontario first declared a state of emergency. While there are drops in transit

Figure 1.2 — Impact of State of Emergencies on Toronto Public Transit Demand

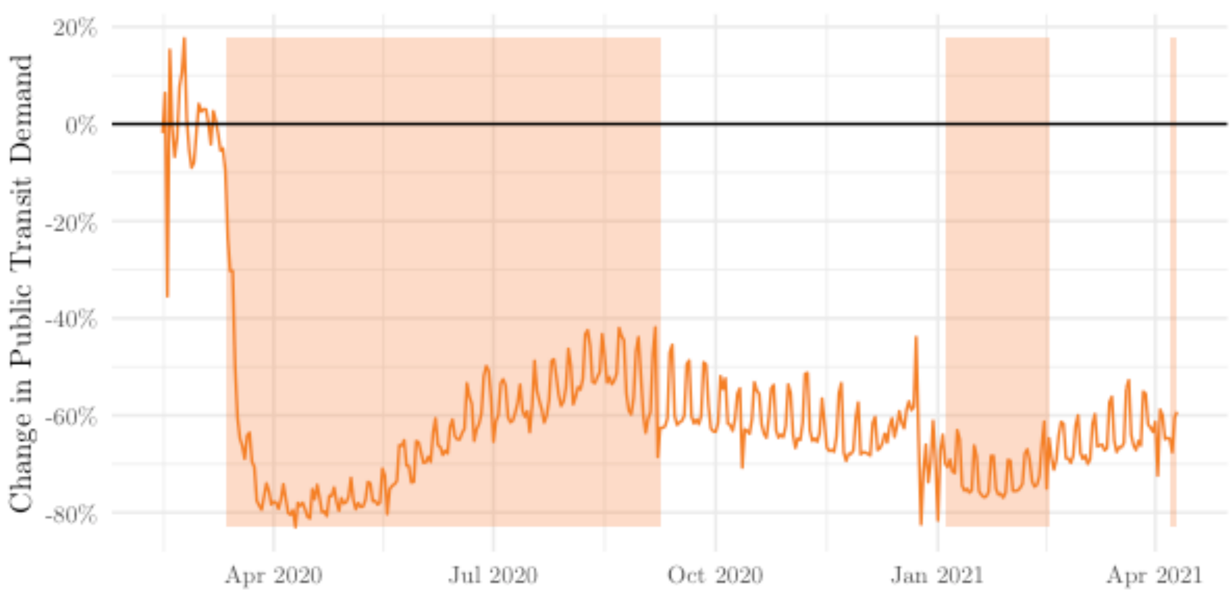


Demand prior to this date, none of them fall outside the February drops that are seemingly unrelated to pandemic measures and none of them are as large as the March 17 drop. Similar effects of the second state of emergency may also be seen in Figure 1.2. It is evident that state of emergencies alone cannot explain all variance in transit demand. There are trends within periods of emergency where transit demand is changing that require the inclusion of additional variables to explain.

(b) School Closures

In Figure 1.3, school closures also appear to be correlated with drops in transit demand, especially at the beginning of the pandemic when they capture some of the decrease in demand that falls outside the scope of provincial state of emergencies. However, the clear overlap between school closures and state of emergencies also posed a potential confound for the model, as it makes it difficult to identify unique effects of each measure on public transit demand. This issue is a concern for most health measures considered.

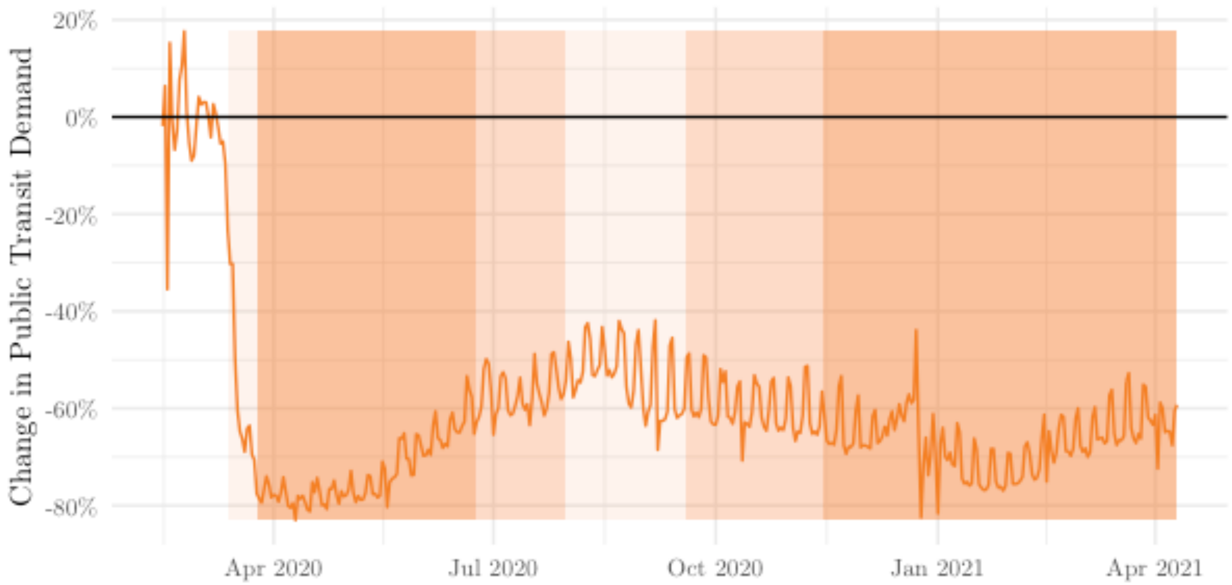
Figure 1.3 — Impact of School Closures on Toronto Public Transit Demand



(c) Gathering Restrictions

Figure 1.4 shows the impact of gathering restrictions on transit demand, with darker areas representing stricter restrictions. The image indicated that this variable may explain some of the trends within periods where a state of emergency was declared. It can be seen that looser gathering restrictions are correlated with periods of resurgence in transit use. However, there still remain trends in transit demand within periods of similar gather restrictions that are not explained by this variable alone. It was again a potential concern that gathering restrictions closely overlap with other public health measures.

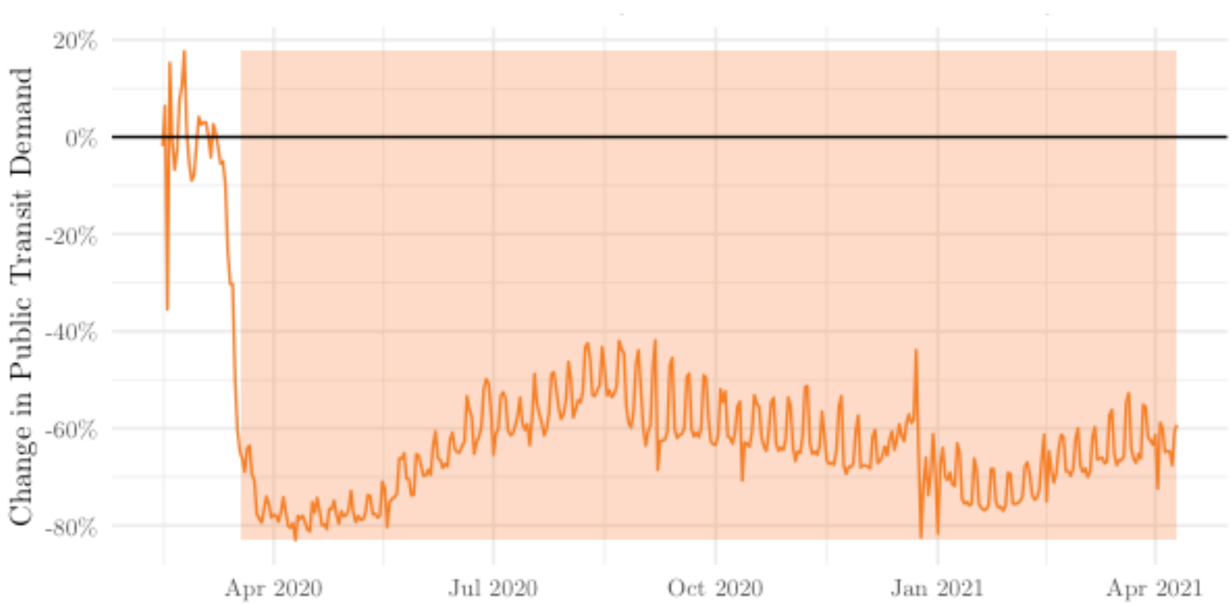
Figure 1.4 — Impact of Gathering Restrictions on Toronto Public Transit Demand



(d) Travel Restrictions

Figure 1.5 shows that travel restrictions have been in place for most of the pandemic. This presented both a potential explanatory variable for changes transit demand in major cities with substantial international travel like Toronto and a further confound for the effects of other public health measures that were implemented in similar timeframes.

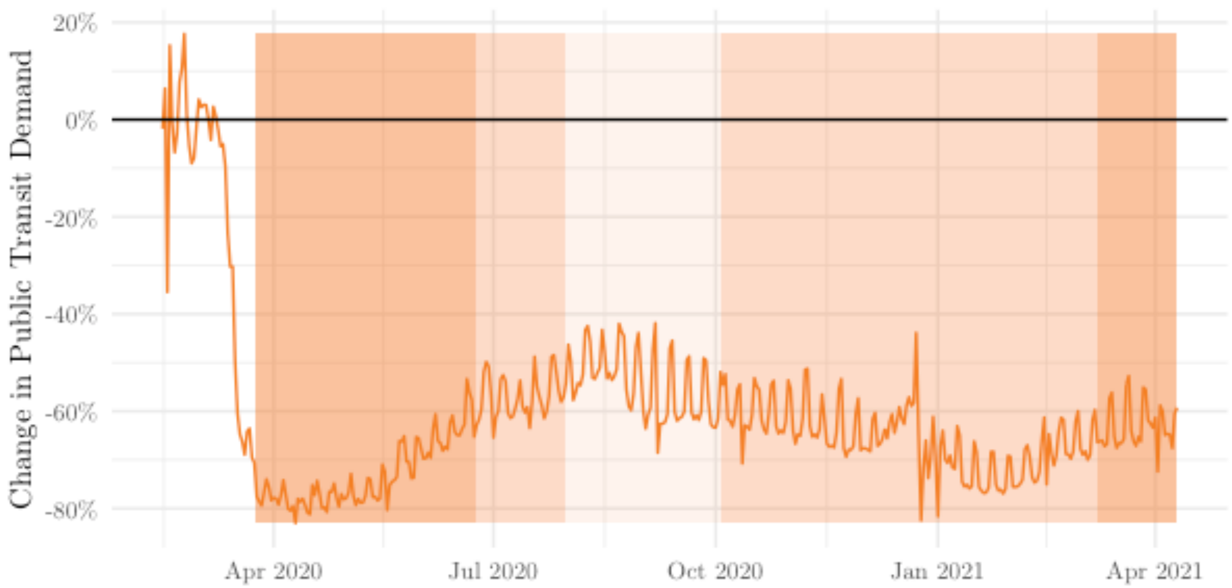
Figure 1.5 — Impact of Travel Restrictions on Toronto Public Transit Demand



(e) Business Closures

Mandated business closures presented another potential explanation for drops in transit demand. However, as they had a high overlap with gather restrictions, as seen in Figure 1.6, caution was used in applying these variables together in a forecasting model.

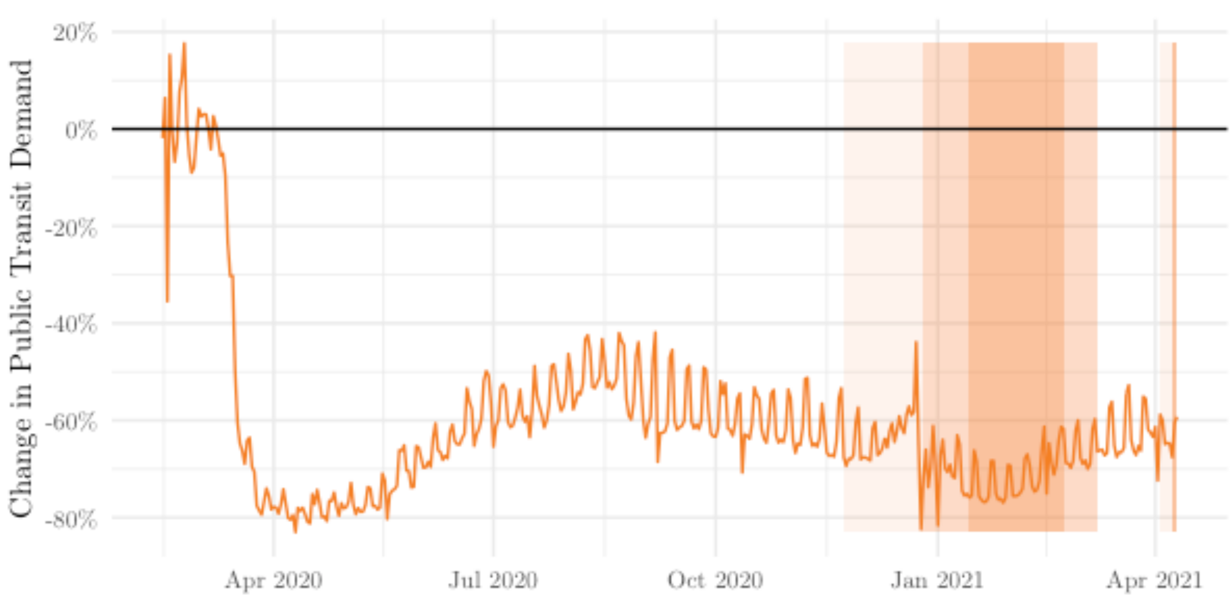
Figure 1.6 — Impact of Business Closures on Toronto Public Transit Demand



(f) Lockdowns

Toronto was began a four week lockdown on November 23, 2020. The province of Ontario also entered into a lockdown on December 26, 2020. A stay-at-home order was later declared on January 14, 2021. In Figure 1.7, there appear to be effects of these mea-

Figure 1.7 — Impact of Lockdowns on Toronto Public Transit Demand

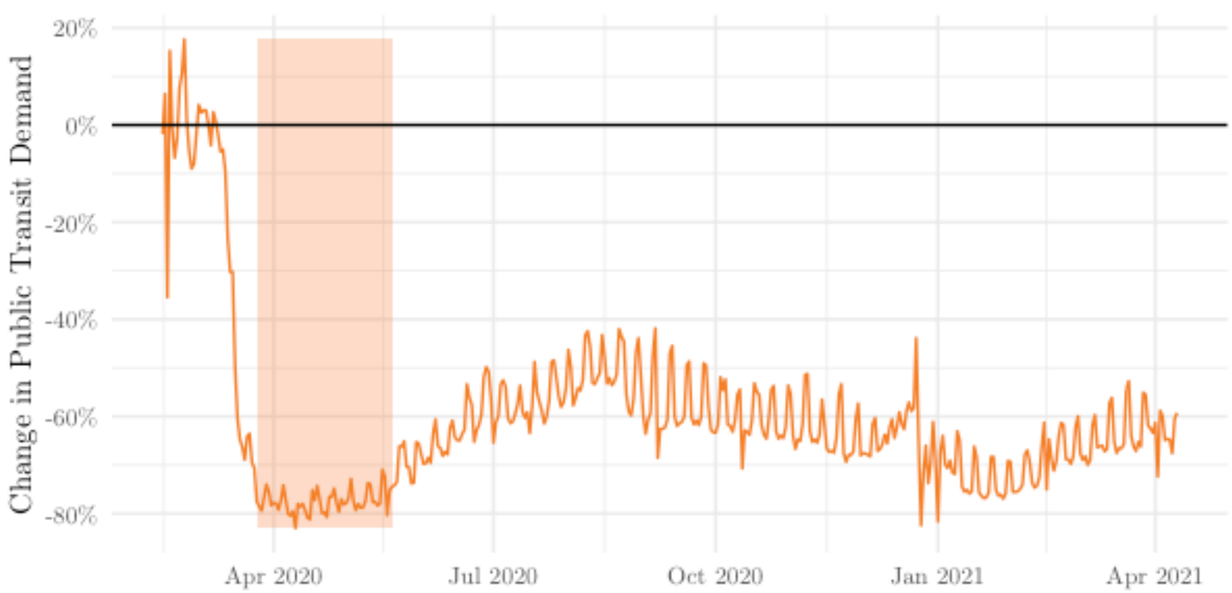


sures on transit demand, though lockdown measures substantially overlap with school closures. The provincial lockdown was also hypothesized to, at least in part, account for the spike in transit use on December 23, 2020 — though it is unclear what predictive value this fact would holds given that its annual circumstances are fairly unique to the interaction between the proximity to a major holiday and the lockdown. Further, as there is only substantial data available for a single cohesive period of lockdowns and stay-at-home orders, it was unclear how generalizable correlations between this period and decreases in Toronto public transit demand would be.

(g) Park Closures

Figure 1.8. shows a potential for an effect of park closures on transit demand, though similar caution must be applied as this overlaps with several other measures. As with lockdowns, there was only a single period of park closures in Toronto, which may make generalizing impacts of this public health measure difficult. Caution was thus applied if when including this measure in the final forecasting model.

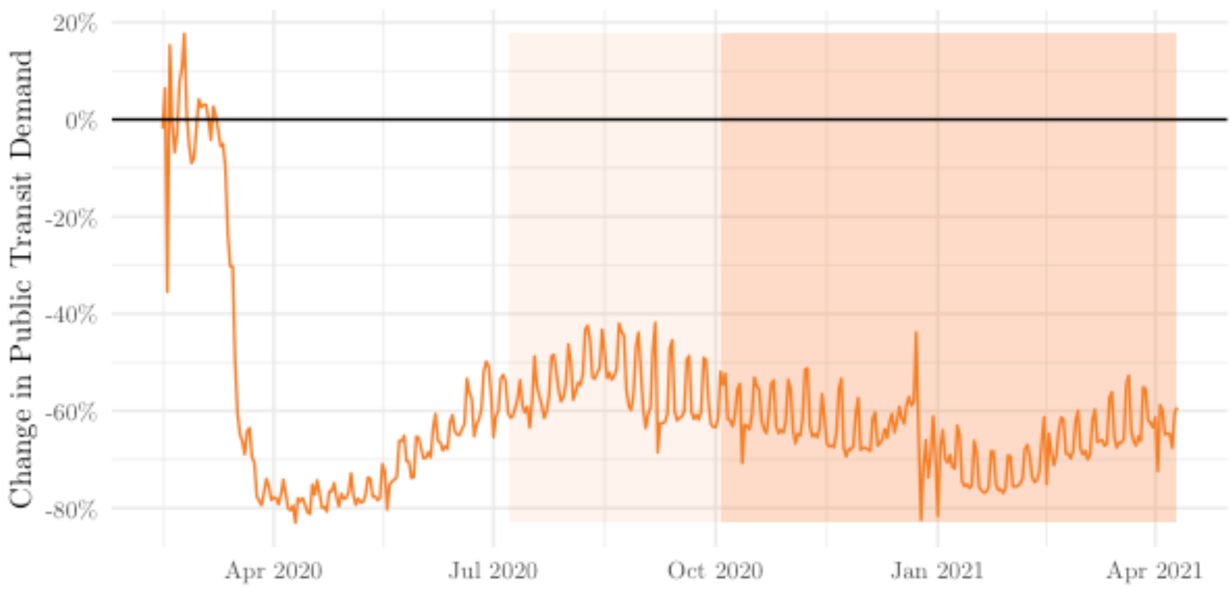
Figure 1.8 — Impact of Park Closures on Toronto Public Transit Demand



(h) Mandatory Masks

Mandatory mask policies are the last of the public health measures examined, shown in Figure 1.9. Unlike the other public health measures explored, there appeared to be little to no discernible effect of mask policies on transit use. This may be expected, as unlike other measures, mandatory mask policies do not provide a direct mechanism whereby movement would be restricted. However, it was still possible that there is some interaction effect between mandatory mask policies and transit demand.

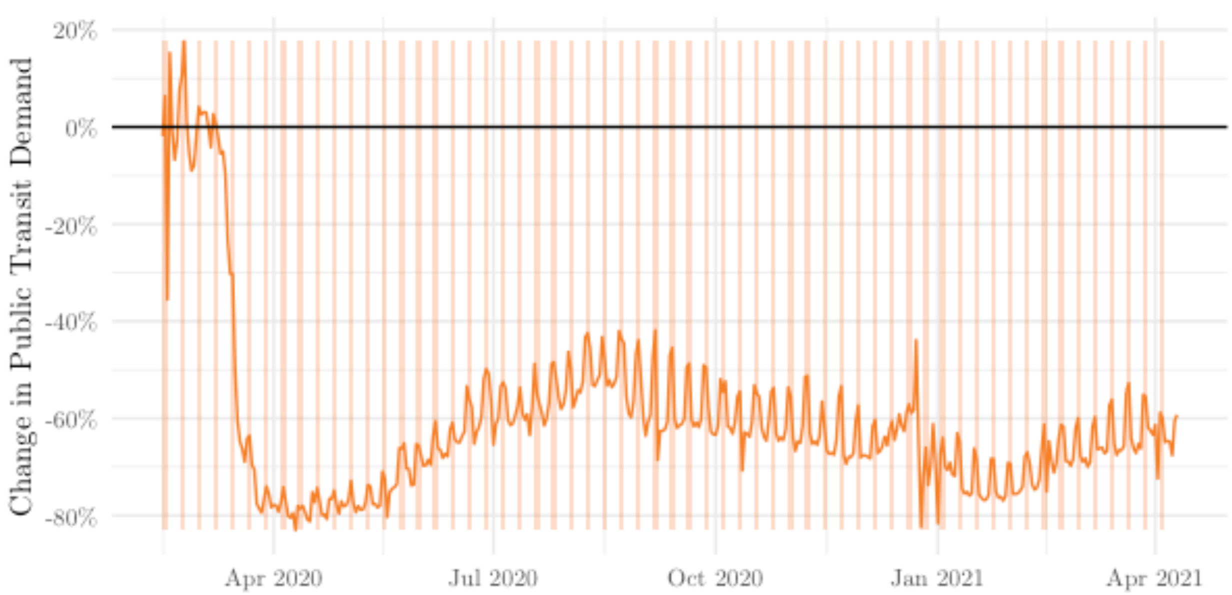
Figure 1.8 — Impact of Park Closures on Toronto Public Transit Demand



(i) Weekends

Though Transit app data adjusted for the day of the week, it was hypothesized that the weekly spikes in transit demand may nonetheless be attributable to weekend differences. This appears to be true, as seen in Figure 1.9. The apparent difference supported the claim that the bulk of transit demand decreases is caused by discretionary commuters who were able to establish alternative means of working during the pandemic. The remaining demand on weekends may thus be less affected by pandemic restrictions since there is a smaller drop in weekend travellers compared to pre-pandemic levels — who may have included more essential workers are recreational users — than on weekdays.

Figure 1.9 — Impact of Park Closures on Toronto Public Transit Demand

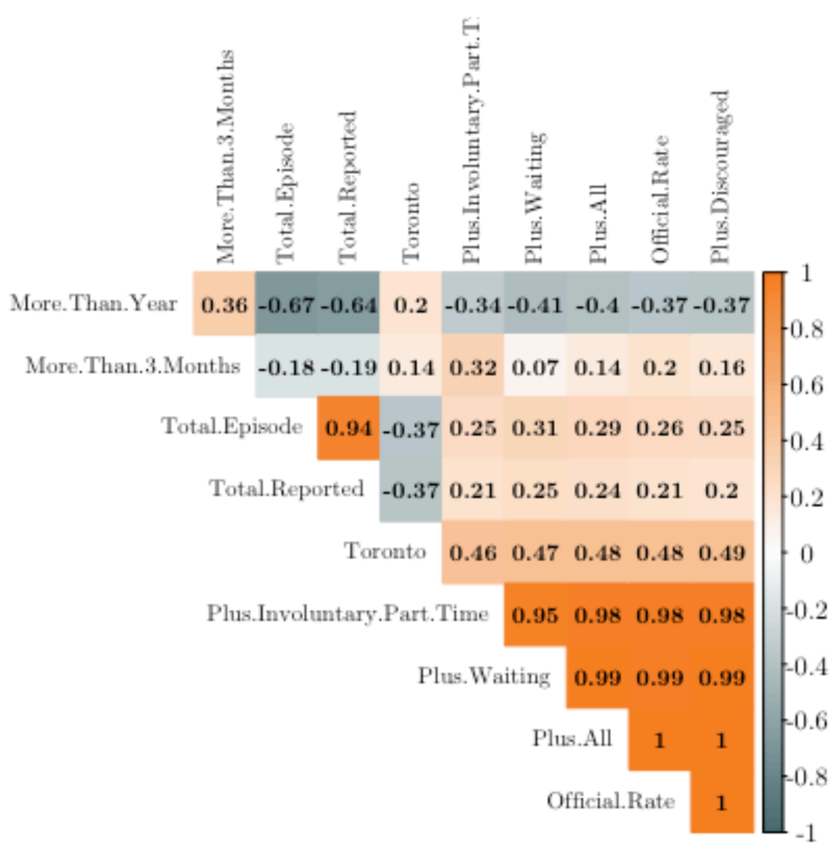


Continuous Input Variables

The two main continuous variables considered were provincial unemployment rates and municipal COVID-19 rates. Case rates were reported on a daily basis and unemployment rates were reported monthly, with no data available for April, 2021 at the time of writing. When possible, continuous variables were scaled to have the same interpretation as the output variable (i.e. change from the same-day pre-pandemic baseline). As there is are no pre-pandemic rates of COVID-19 case rates, this variable was left as a nominal value. Figure 1.10 shows the correlations for all continuous variables explored. In terms of relationships with the outcome variable, identified in the figure as "Toronto," the unemployment rate including discouraged searchers and the total reported COVID-19 case rates had the highest correlation out of all the unemployment and case rate variables, respectively. As the unemployment rates and COVID-19 case rates have high collinearity with other variables in the same category, the final model only included one of each.

When producing an ex ante forecast, there are several available unemployment and case rate forecasts to choose from. However, due to the volatile nature of the pandemic, it may be desirable for transit agencies to produce conditional forecasts dependent on different potential future scenarios, such as a worst case with high unemployment and high case rates. The models produced also allow for this flexibility.

Figure 1.10 — Correlation Matrix for Continuous Variables



ii. Single Equation Regression Models

Two single equation regression models were built using a subset of the available data that excluded the most recent 100 days in order to develop and evaluate an ex post forecast using the holdout data. The first of these models included data collected before the cliff point where Toronto transit demand rapidly decreases, while the second excluded the first 31 days of data. This was done to test whether limited the potential impact of structural breaks in the data following the onset of the pandemic harmed forecast accuracy. The risk of a structural break was weighted against the advantages of the first model which included pre-pandemic data and thus was potentially more capable of generating an accurate prediction of transit demand in post-pandemic scenarios. A third regression model was also build using the entirety of the data, including the most recent 100 days. This model included the same variables used in model that generated the best ex post forecast, and is intended for generating ex ante forecasts of transit demand.

Model Summaries

For all three models, an optimal combination of predictor variables was obtained by starting with a basic model that included only unemployment rates, COVID-19 rates, and weekends. From this, additional variables were added and kept if they improved the overall model fit, as measured by the adjusted R-squared, residual standard error, Akaike information criterion (AIC), and the Bayesian information criterion (BIC).

Care was taken to avoid using highly collinear or overlapping variables that risked producing multicolinearily in the model. Business closure dummy variables were excluded from the model for this reason. Even though lower levels of business closures were statistically significant, their inclusion reduced the significance of other variables in the model. Further, the coefficient estimates for higher levels of business closures were positive, meaning that the model was associating more severe business restrictions with higher transit use, which is conceptually unlikely. Even though removing this variable made the model fit slightly worse, it removed potential concerns that the weights being given to its coefficients were not being skewed by overlaps with other variables.

Of all the input variables included, only COVID-19 case rates made sense for including a lag. This was attempted, but produced multicolinearily issues with the same-day COVID-19 case rates, likely because both values were highly similar. For this reason, no lag on case rates was used. An attempt was also made to include a lag on the dependent variable, but this rendered most of the other predictors insignificant, likely because a large number of input variables did not change on a daily basis and thus would account for less variance in the model than would the lagged output variable. An attempt was also made to include a quadratic term for COVID-19 case rates, but resulted in highly inaccurate predictions for the ex post forecast and was thus removed.

Finally, each successive model was built using the successful variables used in the previous model. This enabled easy comparison of prediction accuracy and simplified the process of diagnosing the source of the forecasting model errors.

(a) Full Training Data Model

The model built using the training data set that excluded the most recent 100 days fit the data extremely well. All included terms were significant at the 0.1% level and had estimated coefficients in line with theoretical predictions. Both coefficient standard errors and overall residual standard errors were very small, and over 95% of the variation in Toronto public transit demand was explained by the model.

Table 2.1 — Summary of Full Training Data Model

Variable	Coefficient	Standard Error	t-Statistic	Probability
`(Intercept)`	-0.03724	0.00907	-4.104	*** 0.00005
`Plus.Discouraged`	-0.00781	0.00202	-3.865	*** 0.00014
`Total.Reported`	-0.00011	0.00003	-3.733	*** 0.00023
`Weekend`	0.08260	0.00632	13.072	*** 0.00000
`State.of.Emergency`	-0.07227	0.01020	-7.082	*** 0.00000
`Gathering.Restrictions1`	-0.38140	0.02252	-16.936	*** 0.00000
`Gathering.Restrictions2`	-0.38580	0.02526	-15.273	*** 0.00000
`Gathering.Restrictions3`	-0.42780	0.02826	-15.141	*** 0.00000
`Travel.Restrictions`	-0.17810	0.02226	-8.000	*** 0.00000
`Park.Closures`	-0.09976	0.00996	-10.012	*** 0.00000
`Weekend:Park.Closures`	-0.05659	0.01495	-3.787	*** 0.00000
Residual Standard Error				0.04572
Multiple R-Squared				0.9505
Adjusted R-Squared				0.9488
F-Statistic				576
p-Value				0.00000

Significance: *** $p < 0.001$ 0, ** $p < 0.01$, * $p < 0.05$

(b) Truncated Training Data Model

The model built using the training data that excluded the both most recent 100 days and all pre-pandemic data also fit the data fairly well, though somewhat less well than the full training data model. Notably, compared to the previous model travel restrictions no longer have a significant effect on transit demand. The second level of gathering restrictions also became insignificant. However, excluding these two variables did not improve model fit. Despite this, most included variables were still significant at the 0.1% level and had estimated coefficients in line with theoretical predictions. Coefficient standard errors and overall residual standard errors were still very small, and over 85% of the variation in transit demand was explained by the model.

Table 2.2 — Summary of Truncated Training Data Model

Variable	Coefficient	Standard Error	t-Statistic	Probability
`(Intercept)`	-0.55040	0.03694	-14.900	*** 0.00000
`Plus.Discouraged`	-0.00799	0.00164	-4.878	*** 0.00000
`Total.Reported`	-0.00010	0.00002	-4.040	*** 0.00007
`Weekend`	0.08401	0.00532	15.799	*** 0.00000
`State.of.Emergency`	-0.06556	0.00818	-8.017	*** 0.00000
`Gathering.Restrictions2`	-0.00814	0.00779	-1.044	0.297
`Gathering.Restrictions3`	-0.05374	0.01176	-4.571	*** 0.00000
`Travel.Restrictions`	-0.04907	0.03731	-1.315	0.190
`Park.Closures`	-0.10010	0.00785	-12.753	*** 0.00000
`Weekend:Park.Closures`	-0.05752	0.01189	-4.837	*** 0.00000
Residual Standard Error				0.0359
Multiple R-Squared				0.8630
Adjusted R-Squared				0.8585
F-Statistic				189
p-Value				0.00000

Significance: *** $p < 0.001$ 0, ** $p < 0.01$, * $p < 0.05$

(c) Full Data Model

The final model using all of available historic data fit the data extremely well. Like the first model, ever included term was significant at the 0.1% level and had estimated coefficients in line with theoretical predictions. Coefficients were also similar to the first model. Both coefficient standard errors and residual standard errors were very small, and over 93% of the variation in public transit demand was explained by the model.

Table 2.3 — Summary of Full Data Model

Variable	Coefficient	Standard Error	t-Statistic	Probability
`(Intercept)`	-0.03562	0.00929	-3.835	*** 0.00015
`Plus.Discouraged`	-0.00736	0.00170	-4.325	*** 0.00002
`Total.Reported`	-0.00008	0.00001	-5.398	*** 0.00000
`Weekend`	0.07773	0.00555	13.999	*** 0.00000
`State.of.Emergency`	-0.06432	0.00686	-9.380	*** 0.00000
`Gathering.Restrictions1`	-0.38250	0.02310	-16.557	*** 0.00000
`Gathering.Restrictions2`	-0.39410	0.02473	-15.936	*** 0.00000
`Gathering.Restrictions3`	-0.44790	0.02541	-17.629	*** 0.00000
`Travel.Restrictions`	-0.17770	0.02267	-7.836	*** 0.00000
`Park.Closures`	-0.09275	0.00963	-9.626	*** 0.00000
`Weekend:Park.Closures`	-0.05048	0.01503	-3.358	*** 0.00086
Residual Standard Error				0.04724
Multiple R-Squared				0.9360
Adjusted R-Squared				0.9344
F-Statistic				584.9
p-Value				0.00000

Significance: *** $p < 0.001$ 0, ** $p < 0.01$, * $p < 0.05$

Validation of Model Assumptions

Additional diagnostic tests were also run to validate key model assumptions.

(a) Heteroscedasticity

The White test for heteroscedasticity indicated that all three models may have heteroscedastic residuals. Future iterations of the model may wish to consider performing a Newey-West correction to adjust model standard errors accordingly. Figures 2.1-3 show the residuals versus predicted values plots for each model. It can be seen that the second model exhibits slightly less heteroscedastic than the other two models.

Figure 2.1 — Residuals Versus Predicted Values for Full Training Data Model

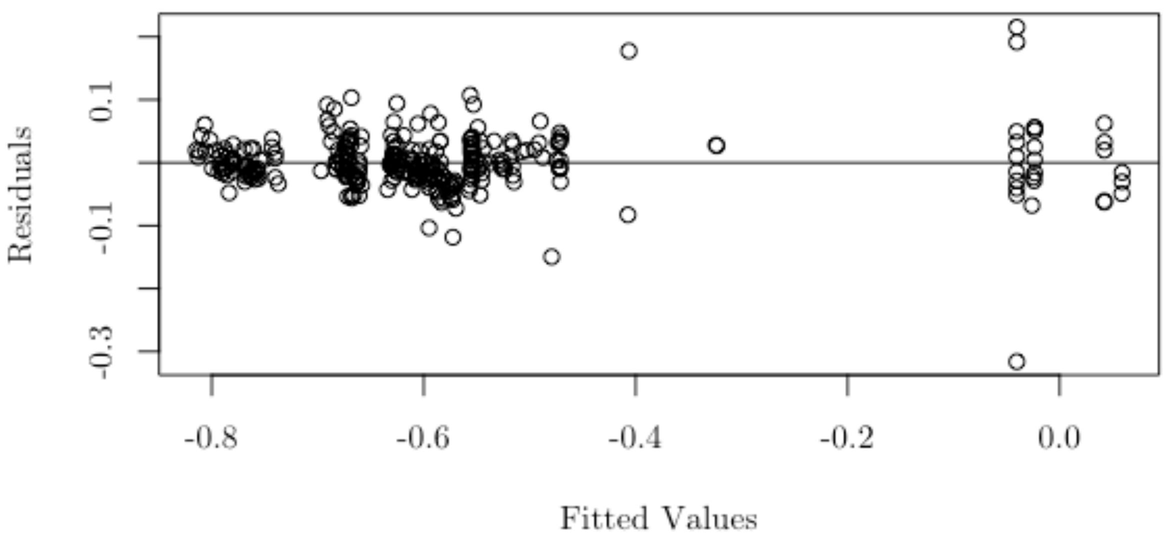


Figure 2.2 — Residuals Versus Predicted Values for Truncated Training Data Model

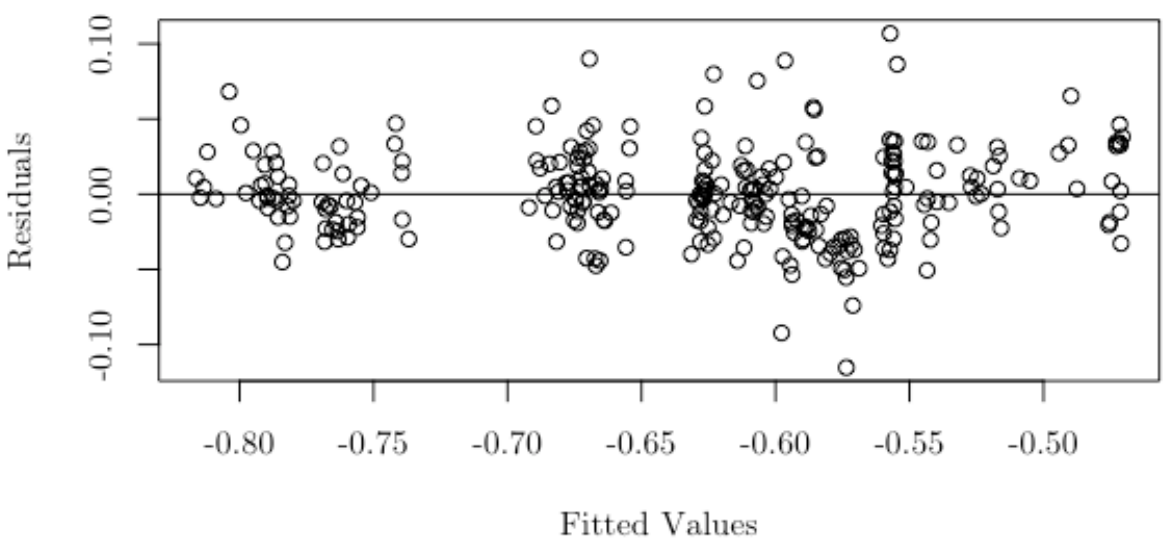
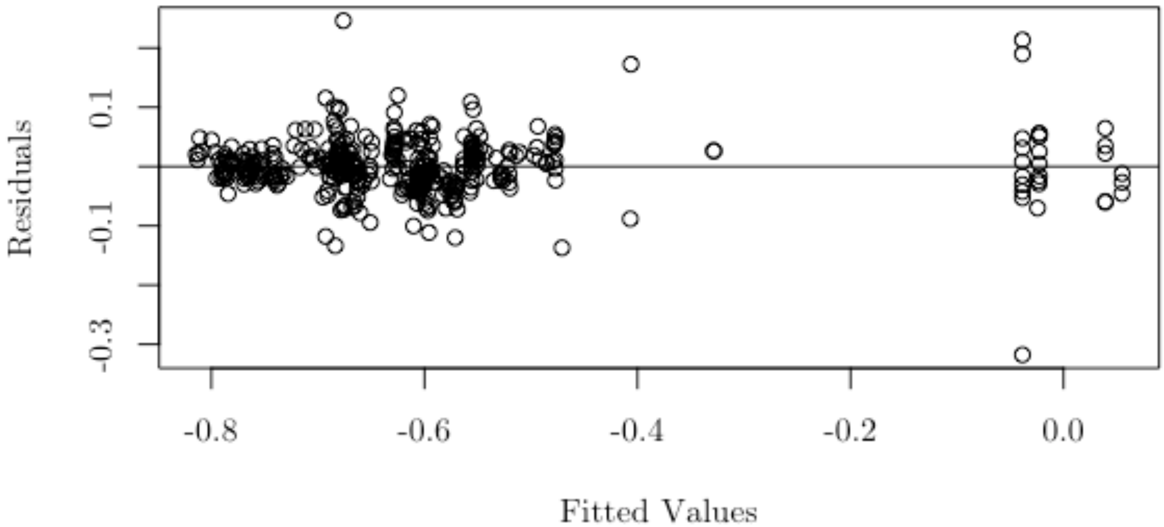


Figure 2.3 — Residuals Versus Predicted Values for Full Data Model



(b) Normality of Residuals

Shapiro-Wilk tests for normality also uncovered a high probability that all three models did not produce normally distributed residuals, violating one of the assumptions of ordinary least squares (OLS) regression. Figures 2.4-10 illustrate this issue. Though non-normally distributed residuals may prevent OLS regression from being the most efficient estimator, this violation does not seriously harm forecast accuracy and can likely be safely ignored for the final model. Note that the second model with the truncated training data appears to have the most normally distributed data of the three models.

Figure 2.4 — Residual Density Plot for Full Training Data Model

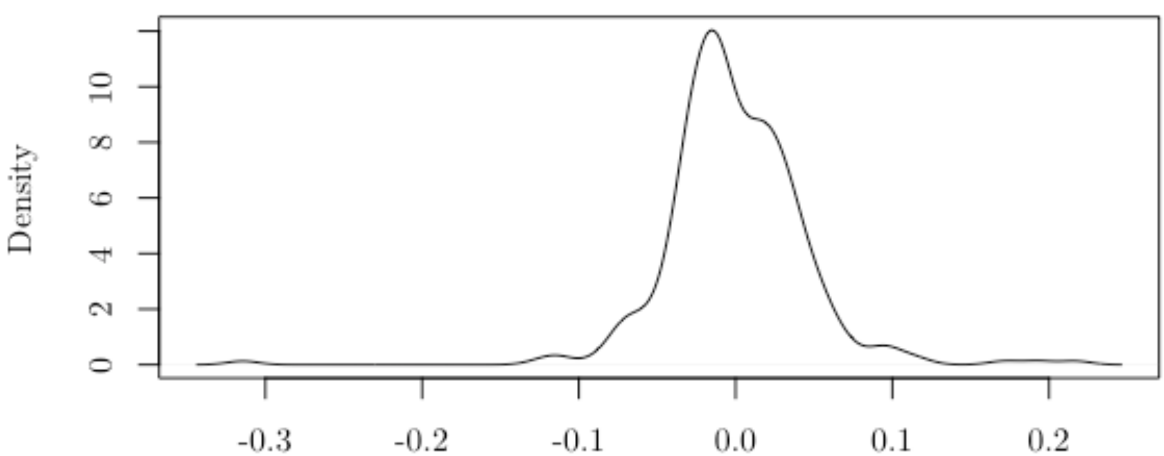


Figure 2.5 — Residual Density Plot for Truncated Training Data Model

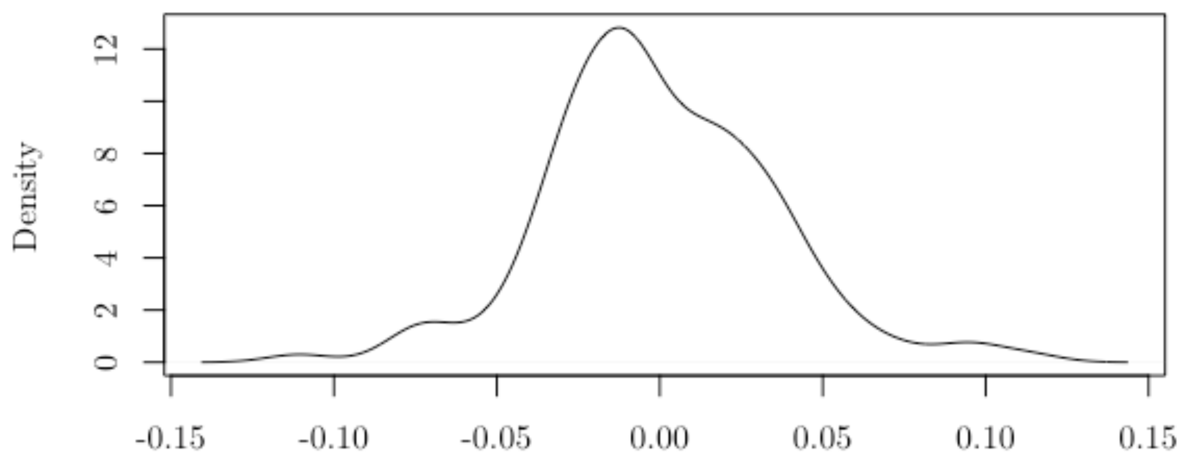


Figure 2.6 — Residual Density Plot for Full Data Model

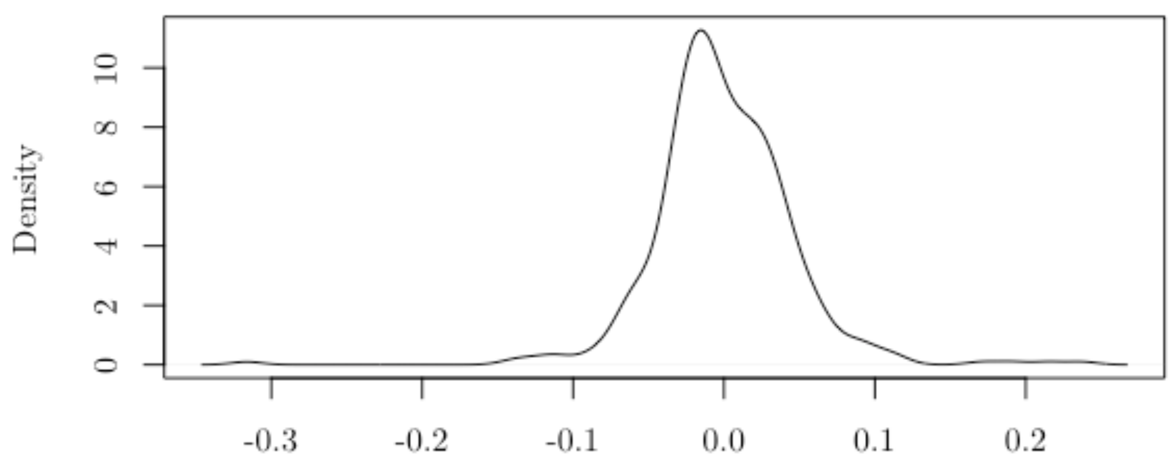


Figure 2.7 — Q-Q Plot for Full Training Data Model

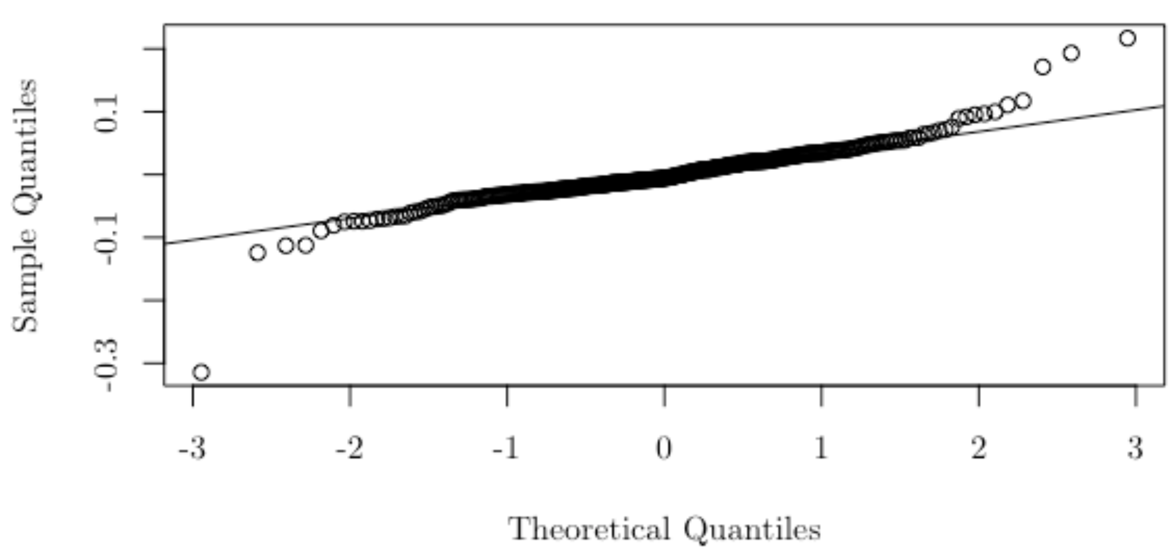


Figure 2.8 — Q-Q Plot for Truncated Training Data Model

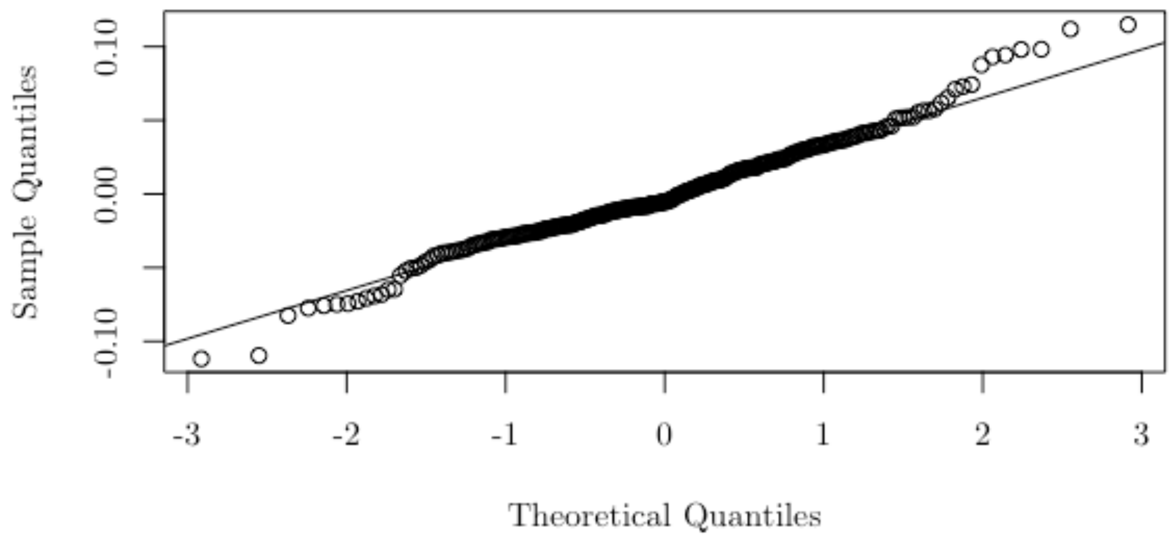


Figure 2.9 — Q-Q Plot for Full Data Model

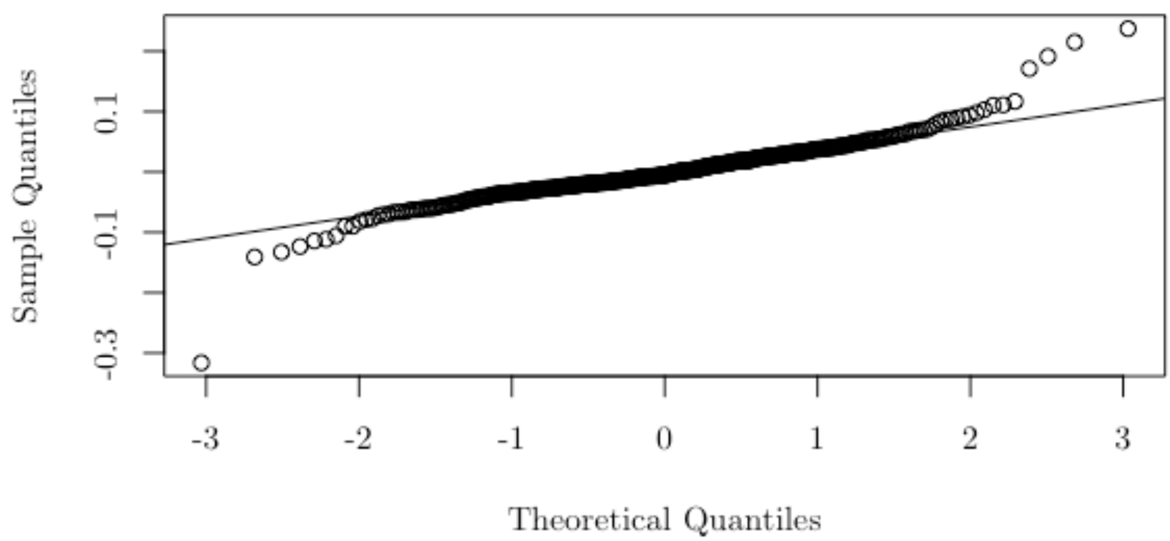
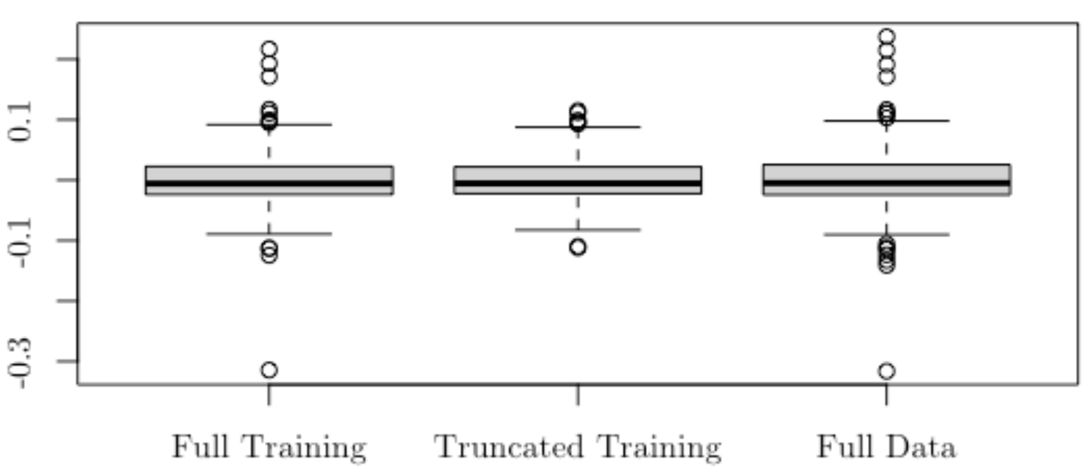


Figure 2.10 — Residual Interquartile Range for All Models



(c) Serial Correlation

Durbin-Watson statistics were computed for each model. Though serial correlation was detected for each model, the second model build using the second model had the statistic that was the furthest away from 2, indicated a lower probability of serial correlation. Future corrections to the model may wish to consider adding autoregressive terms.

(d) Specification Errors

Ramsey RESET tests for misspecified quadratic and cubic found potential misspecification in all three models. Future models may wish to more closely explore the use of higher polynomial values for COVID-19 case rates and unemployment rates.

(e) Structural Breaks

Chow tests for structural breaks also found significant structural breaks in all three models. This was likely to be expected from the nature of the data.

Forecast Accuracy

All three models were applied to the most recent 100 days of data to produce predictions. For the first two models, this was an ex post forecast, as the past 100 days were not included in the data used to build the model. For the third model, this data was used in the process of building the model, so errors are expected to be lower and less indicative of the accuracy that would be expected in an ex ante forecast.

As can be seen in Table 2.4, all three models performed exceedingly well. The table shows the mean error (ME), root mean squared error (RMSE), mean absolute error (MAE), mean percentage error (MPE), and mean absolute percentage error (MAPE). The mean error for all models was essentially zero, which was highly impressive.

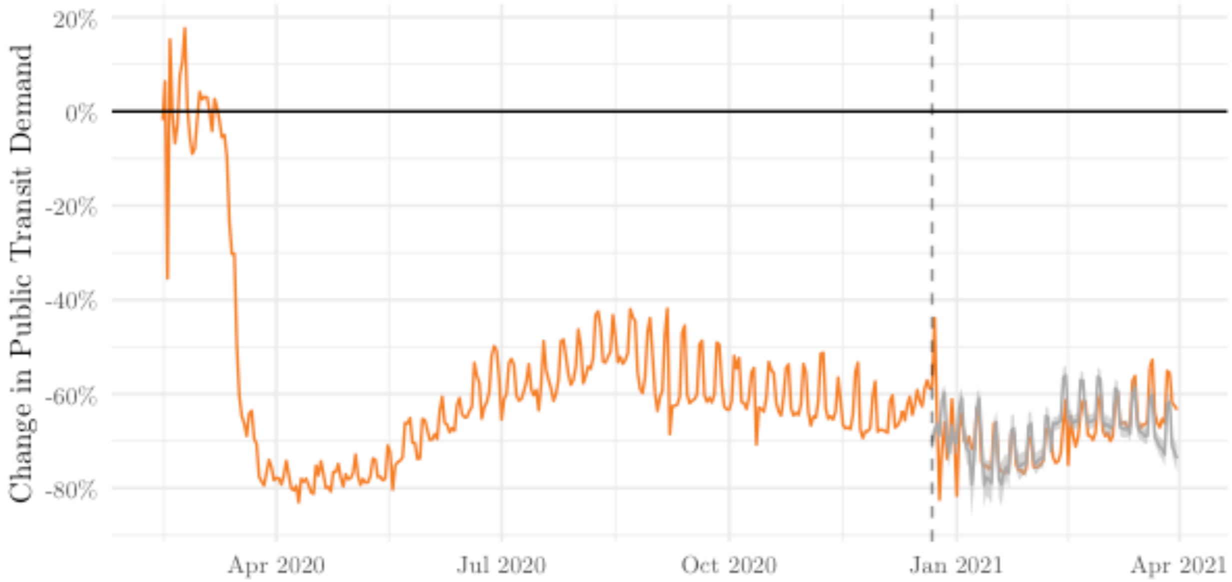
Table 2.4 — Forecast Error Comparison for All Models

Model	ME	RMSE	MAE	MPE	MAPE
Full Training Data Model	-0.00130	0.05340	0.03955	-0.28754	6.08002
Truncated Training Data Model	-0.00330	0.05251	0.03902	-0.02177	6.00349
Full Data Model	-0.00407	0.05071	0.03677	0.05833	5.69012

(a) Full Training Data Model

Figure 2.11 shows forecasted values for most recent 100 days of data produced by the first model. As expected by the forecast error metrics, the forecast is very close to the actual data values. There is a slight divergence at the very end of the ex post forecast, but this deviance is not severe compared to the range of transit demand data.

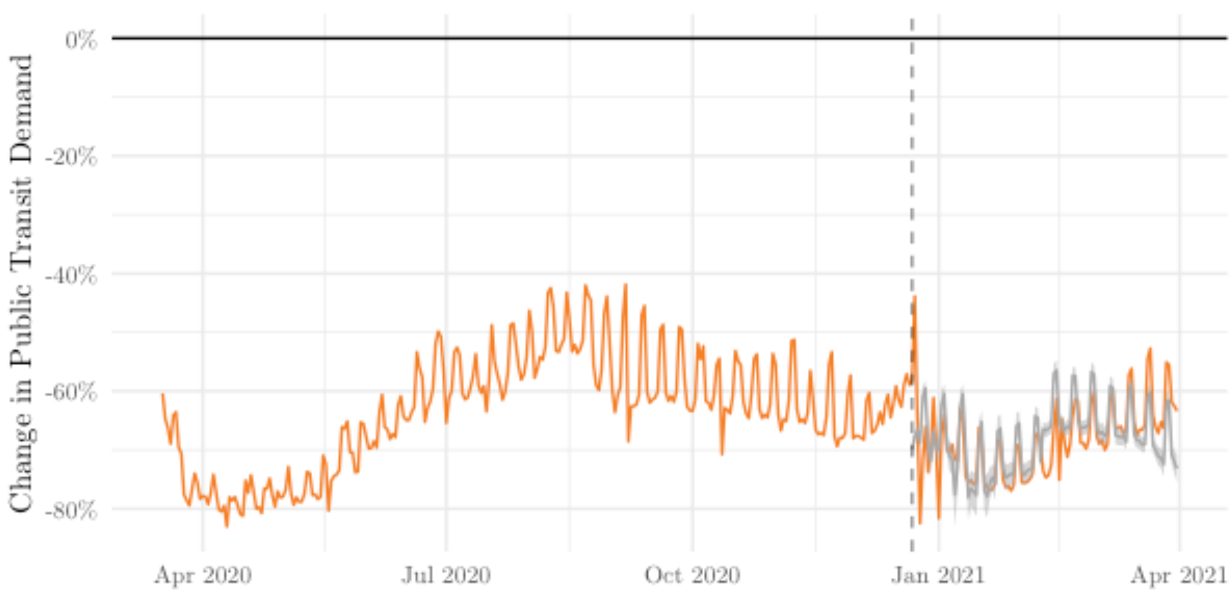
Figure 2.11 — Ex Post Forecast for Full Training Data Model



(b) Truncated Training Data Model

The second single equation regression model also performed exceedingly well in the ex post forecast. There again appears to be a slight divergence between predicted and actual values at the very end of the forecasted data, but again this does not appear to be severe compared to the data range. The prevalence of this diverge may have been caused by attrition in adherence to longterm public health measures. Further research is needed to confirm whether the model should be adjusted.

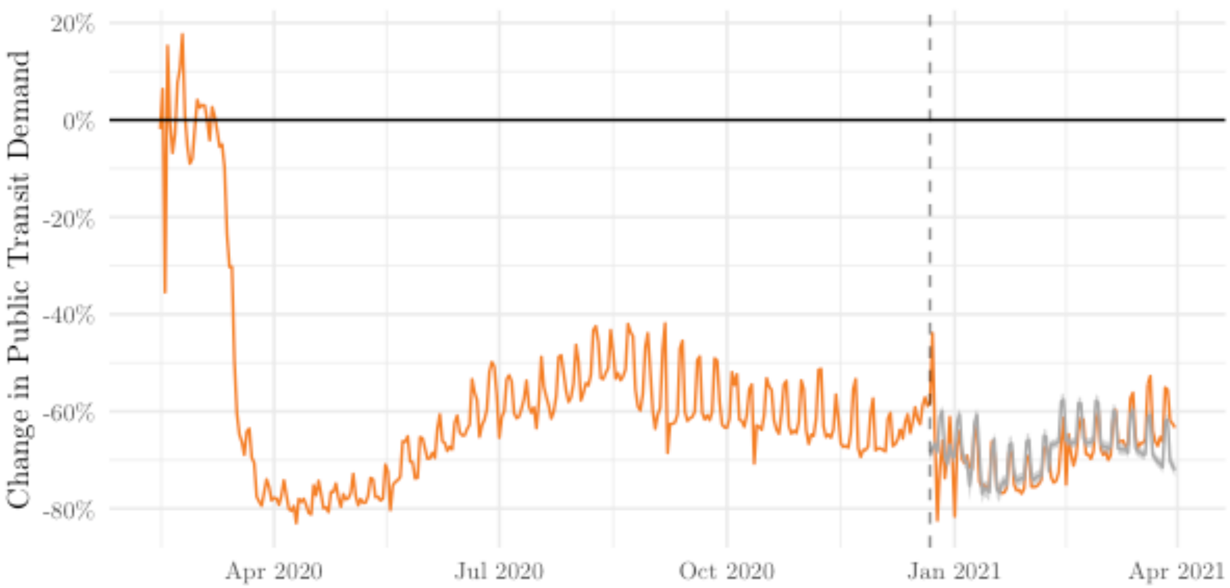
Figure 2.12 — Ex Post Forecast for Truncated Training Data Model



(c) Full Data Model

Finally, the full data model generated predictions highly similar to actual values. Even though this model was built using the past 100 days of data, the similarity of its forecast to the first model indicates that the inclusion of more recent data did not greatly alter the predictions made. This is evidence that an ex ante forecast would behave similarly.

Figure 2.11 — Forecast for Full Data Model



iii. ARIMA Models

An ARIMA forecast of Toronto public transit demand was also produced to serve as a baseline for the single equation regression forecasts. The correlogram shown in Figure 3.1 indicates that the autocorrelation function of the undifferenced output variable does not go to zero very quickly. The differenced variable in Figure 3.2, however, does.

Figure 3.1 — Correlogram for Undifferenced Change in Toronto Transit Demand

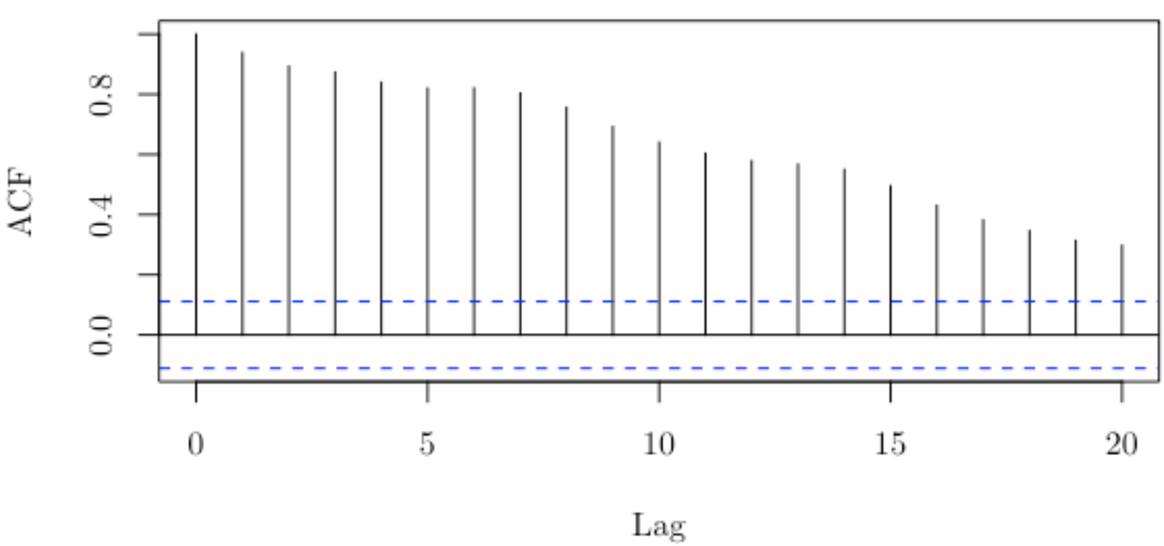
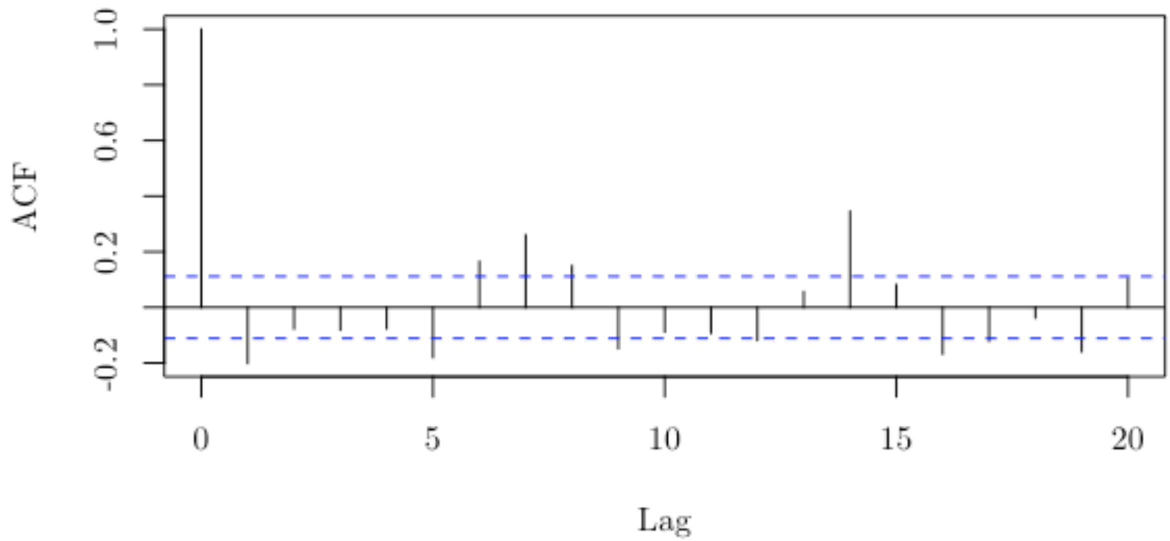


Figure 3.2 — Correlogram for First Difference Change in Toronto Transit Demand



Q-statistics for each lag of the differenced data are all statistically significant, indicating that the change in the output variable is not white noise, and can be forecast using autoregressive methods. An Augmented Dickey-Fuller test further reveals that the null hypothesis that the differenced series contains a unit root can be rejected, and is instead stationary, as indicated by the alternative hypothesis.

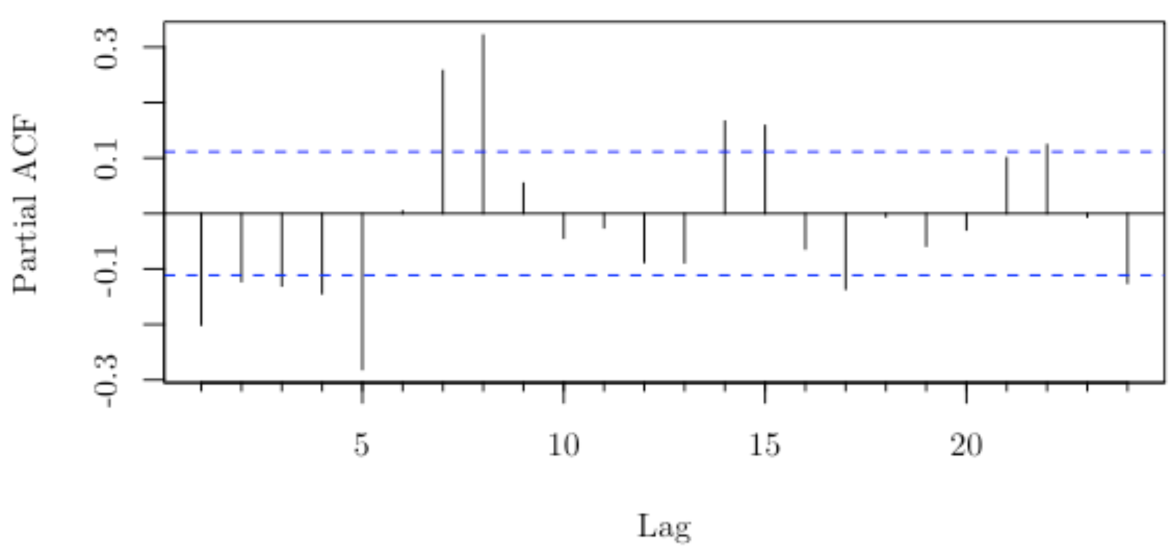
Model Summaries

Two ARIMA models were built using the differences series. The first was built manually, using visually determined autoregressive and moving average components. The second was created using a built-in autogenerated ARIMA function.

(a) Manual ARIMA Model

From the differenced series autocorrelation function in Figure 3.2, it appears as if there

Figure 3.3 — PAC for First Difference Change in Toronto Transit Demand



may be an autoregressive process at the first and fifth lag. From the partial difference series autocorrelation function in Figure 3.3, it appears as if there might be a moving average process at the fifth, seventh, and eighth lags, though this is somewhat unclear. When building the ARIMA function, convergence issues occurred when using moving average components greater than five, so only the first moving average lag was used. This resulted in a final ARIMA model that used a first-differenced series with a fifth order autoregressive process and a first order moving average process.

Table 3.1 — Summary of Manual ARIMA Model

Component	Coefficient	Standard Error
AR(1)	-0.3042	0.0538
AR(2)	-0.2221	0.0585
AR(3)	-0.2405	0.0658
AR(4)	-0.1555	0.0665
AR(5)	-0.3844	0.0668

(b) Autogenerated ARIMA Model

The autogenerated ARIMA function was very similar to the manually generated one, with the exception that no moving average process was recommended. Thus the final autogenerated ARIMA model used a first-differenced series with a fifth order autoregressive process and no moving average processes.

Table 3.2 — Summary of Autogenerated ARIMA Model

Component	Coefficient	Standard Error
AR(1)	-0.4296	0.1226
AR(2)	-0.2557	0.0668
AR(3)	-0.2624	0.0680
AR(4)	-0.1748	0.0686
AR(5)	-0.4107	0.0673
MA(1)	0.1395	0.1241

Validation of Model Assumptions

Additional diagnostic tests were also run to validate key ARIMA model assumptions.

(a) Sum of Coefficients

The sum of the manually generated ARIMA autoregressive coefficients is -1.3067, while the sum of the autogenerated ARIMA coefficients is -1.5332. This indicates that the series is stationary and the mean of the series is well-defined.

(b) Residual White Noise

A Ljung-Box test on the first 10 lags of both models produces a statistically significant Q-statistic. This indicates that residuals are not white noise, and may require adjustment in future models to reduce forecast error.

Forecast Accuracy

Both ARIMA models were applied to to the past 100 days of data to produce ex post forecasts. As ARIMA predictions are generated based on the degree of lag used in the model specification, the forecast will eventually revert to the mean value of the output variable. While this is obviously not ideal for an accurate forecast, it serves as a baseline from which the single equation regression forecasts can be compared.

(a) Manual ARIMA Model

Figure 3.4. shows that the manual ARIMA predictions had wide confidence intervals and the given forecast was essentially the data mean. This produced very high mean errors and mean percentage errors that indicate ARIMA modelling is not appropriate for a long-term forecast of Toronto public transit demand. Theil's U-statistic was also incredibly high, indicating both high variance and bias.

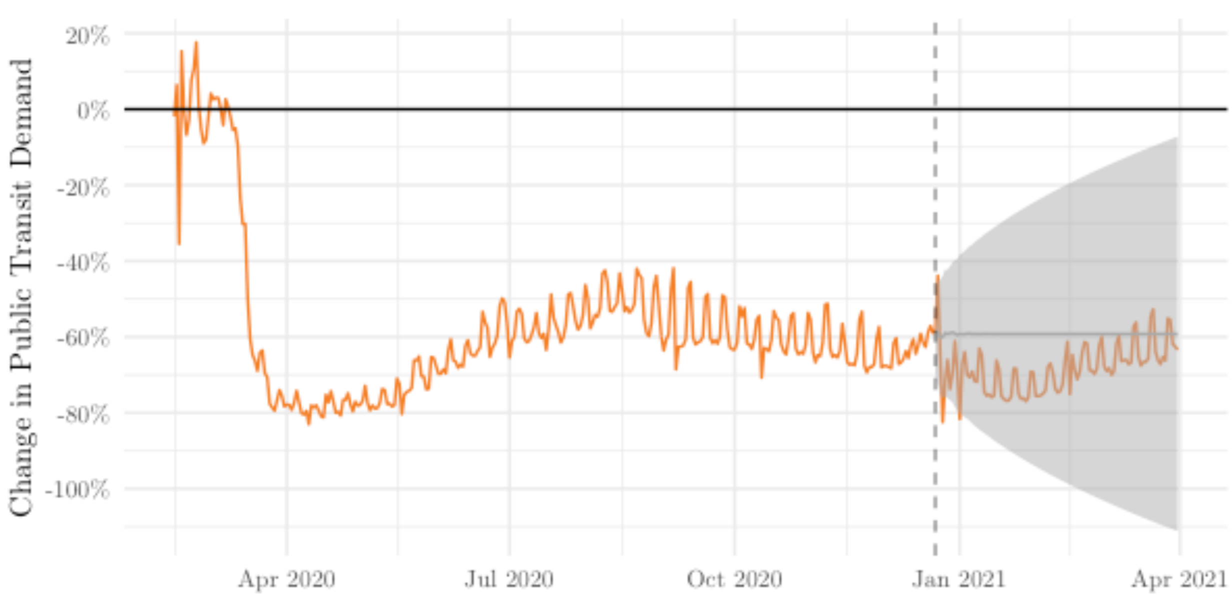
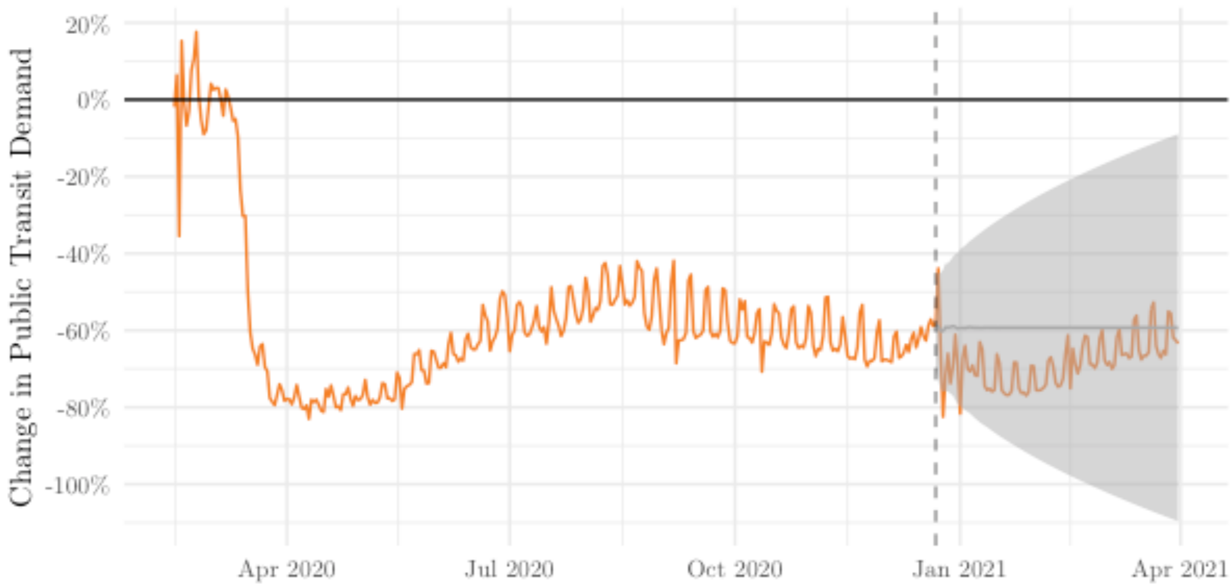


Figure 3.4 — Ex Post Forecast for Manual ARIMA Model

(b) Autogenerated ARIMA Model

The same issues with the manual ARIMA model were present in the autogenerated model, though mean errors and mean percentage errors were marginally lower. Theil's U-statistic was again incredibly high, indicating both high variance and bias.

Figure 3.5 — Ex Post Forecast for Autogenerated ARIMA Model



Results

In comparing the models generated, it was evident that the single equation regression predictions are highly superior to the baseline time series forecast. Of the three regression models produced, the two forecasts produced from the full dataset, including pre-pandemic data, were most accurate. To generate ex ante forecasts, the third regression model is thus most appropriate, as it utilizes all the available data in a manner that is validated by a highly accurate ex post forecast. Future improvements to the model could correct for heteroscedasticity, non-normal residuals, serial correlation, specification issues, and structural breaks. However, despite the fact that these issues were present, the ex post forecast generated were nonetheless very accurate.

With the full data single equation regression identified as the most accurate forecast model, there are nonetheless major caveats that would need to be considered before any ex ante forecast of Toronto transit demand is generated. As discussed in the methodology section, there is some concern as to whether this indicates the presence of an omitted variable or attrition effect where extended public health measures have a diminishing effect on behaviour over time. If this is the case, then the models generated may be biased. There is also concern over unmeasurable omitted variables, such as increases in the number of employees who are working from home. The exclusion of this variable may damage the long term validity of the model if public health measures ease and unemployment rates return to pre-pandemic levels, but a substantial portion of the workforce continues to work from home. The effects of a permanent shift away from commuting has major implications for public transit recoveries, as it may result in permanent decreases in transit use from the pre-pandemic baseline.

To produce useful ex ante forecasts, conditional forecasts can be generated under different assumptions. For example, changes in public transit demand can be forecast under a “best-case” scenario with low COVID-19 case rates, decreasing unemployment, and easement of pandemic measures. They can also be forecast under a “current-state” scenario with similar trends in COVID-19 case rates, unchanged unemployment, and extended pandemic measures. Finally, forecasts can be produced under a “worst-case” scenario with assumptions of increased COVID-19 case rates, higher unemployment, and more severe pandemic measures. Such conditional forecasts can enable transit agencies to make predictions about how transit demand will be effected during the pandemic.

Conclusion

The models built in this paper demonstrate that it is indeed possible to forecast changes in public transit demand form the pre-pandemic baseline to with a very small margin of error. These forecasts perform substantially better than autoregressive and moving average generated predictions. Using the models generated from this study, it is possible for transit agencies to develop conditional ex ante forecasts about transit demand responses to potential future public health restrictions, enabling these agencies to have a highly accurate understanding of the revenue losses they will incur coming out of the pandemic.

Works Cited

Liu, L., Miller, J. H., & Scheff, J. (2020). The impacts of COVID-19 pandemic on public transit demand in the United States. *PLOS ONE, 15*(11). https://doi.org/10.1371/journal.pone.0242476

R Core Team (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org

Transit App. (2020). *About our COVID-19 data.*

Transit App. (2020). *Who's left riding public transit? A COVID data deep-dive.*

Toronto Transit Commission. (2020). *COVID-19: Restart and Recovery Update.*

Tyler, M. (2021). COVID-19 & Transportation Funding in Illinois One Year Later. *Illinois Economic Policy Institute.*