# Predicting long-term effects of COVID-19 on U.S. Mass-Transit

Noah Duggan Erickson

6 December 2023

**Abstract**

This paper used data provided by the National Transit Database, LYNX GIS, CivicImpactJHIJ, and ESRI DM to attempt investigation of the lasting effects of the COVID-19 pandemic's surge in "working from home" on US mass-transit networks. It attempts to investigate the effects both from a ridership numbers perspective and from a geographical perspective. A combined SSA+ARMA model is used to estimate trends of transit ridership pre-pandemic and post-pandemic to conclude that ridership will return to pre-pandemic levels in mid-2026 or late-2027. However, there is insufficient post-pandemic data to reasonably construct a multi-dimensional model for geographic distributions that reflects this convergence.

# 1 Introduction & Literature Review

It's no secret that the COVID-19 pandemic has brought on a massive shift in how work gets done. Specifically, the adoption of work from home (WFH) increasing ten-fold from pre-pandemic levels by May 2020 (Barrero et al. 36). Furthermore, this shift in work arrangements shows no sign of returning to the pre-pandemic level, with employers planning to keep WFH for around 21.3% ($SE = 0.2$) of full paid work days post-COVID, at odds with the 45.5% ($SE = 0.3$) desired by employees (Barrero et al. 48). However, employer sentiment is not uniform - jobs in areas with high population density plan for a greater proportion of days to remain WFH[1] ($t \approx 12.2, p \approx 1.58 \times 10^{-34}$) (Barrero et al. 44, Author's calculations). This has implications not only for those workers, but also the 62% of employed adults who state that most of their job responsibilities are not compatible with WFH (Parker et al. 8) through a decrease in traffic volume reducing travel times to the Central Business District (CBD) by 20-40% for only a 10% reduction in volume (Rappaport et al. 17) brought on by the increased reallocation of WFH employees to less populated areas (Delventhal and Parkhomenko 29).

# 2 Data

## 2.1 Sources

Data for transit ridership was acquired from the UPT sheet of the "Complete Monthly Ridership (with adjustments and estimates)" provided by the National Transit Database, a division of USDOT. Additionally, GIS data for the UZA polygons came from LYNX GIS, county-level COVID data from CivicImpactJHIJ, and county boundaries from ESRI DM.

[1] $H_0 : \beta_1 = 0, H_1 : \beta_1 > 0, \hat{\beta_1} = 2.2(0.18), N = 19447$

## 2.2 Pre-processing

Prior to modelling, there were a large number of pre-processing steps done on the data.

First, ridership data was imported to Python using the Pandas library, grouped by UZA number, and manipulated into formats that could be read by $R$ and ArcGIS.

Then, various geoprocessing tools were used in ArcGIS to copy the aforementioned data into a project, group COVID cases by county and month (collapsing from daily), then intersect and dissolve the UZAs on counties (since the majority of UZAs span multiple counties, the COVID case count for a UZA was taken as the sum of the counties it spans). Additionally, the Python ridership data was imported into ArcGIS, and was then transposed and joined to the UZA layer. These two UZA datasets were then joined together, cleaned, and pre-COVID null case values were filled with zeroes.

# 3 Modelling

## 3.1 Ridership Forecast Modelling

### 3.1.1 SSA+ARMA

An SSA+ARMA model was developed in $R$ to forecast ridership values. This model begins by estimating and removing the trend and seasonal components of the series through SSA with automated grouping using the W-correlation matrix, then attempts to enhance the forecasts by applying an ARMA model to the residuals of the SSA. The model is therefore parameterized by the maximum number of eigenvalues to consider in the grouping step of the SSA (`ssa.resolution`), and the information criterion and maximum values of $p$ and $q$ for the ARMA sub-model.

### 3.1.2 SARIMA

Modelling was also attempted with a SARIMA model, but this was discarded in favor of the SSA+ARMA model due to the variance of the residuals spurred by the complex nature of the seasonal components.

### 3.1.3 Sigmoid Convergence

In order to reasonably model the convergence of the post-COVID datasets into the true-trended pre-COVID datasets, the ”sigmoidal convergence” model was developed in $R$. Details of this model are discussed in Appendix A.

### 3.1.4 Forest-based

A third attempt at modelling the data was attempted using the *Archetypal Analysis*-based (Cutler and Breiman) Forest-based regression that is built into ArcGIS, using COVID cases as an explanatory variable. However, this was also discarded due to poor detection of post-COVID resurgence in ridership, instead stabilizing around the most recent datapoint.

## 3.2 Geographical Ridership Distributions

ArcGIS was used to build pre-COVID and post-COVID space-time cubes for analysis. Due to RAM limitations, the UZA polygons were converted to points for analysis.

Using the built-in tools, a trend analysis, clustering (using a pseudo-F-statistic), and an emerging hot spot analysis using $k = 8$ nearest neighbors as the neighborhood relationship conceptualization method and ”neighborhood time step” as the global window. were performed both on the space-time cube with all data and a pre-COVID subset cube. Additionally, a manual analysis comparing the pre-COVID and post-COVID trends was performed.

Figure 1: Bus Forecasting Model

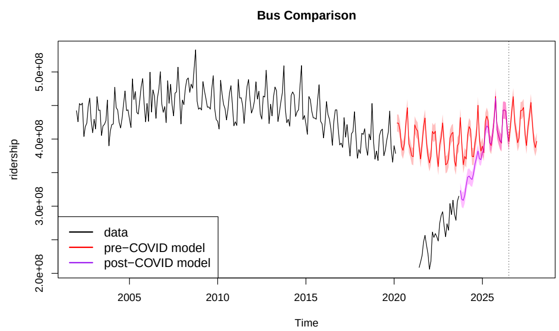
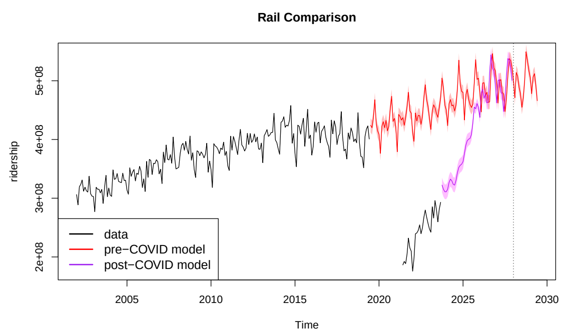



Figure 2: Rail Forecasting Model




# 4 Results

## 4.1 Ridership Forecasting

The results of the bus and rail SSA+ARMA models are shown in figures 1 and 2 respectively.

Across all models, the AICc information criterion was used with maximum $p$ and $q$ values of 5. All post-COVID models used an SSA resolution of 5 eigenvalues, whereas the bus and rail pre-COVID models used resolutions of 23 and 24 eigenvalues, respectively.

The ARMA processes that were generated on the SSA residuals are shown in Table 1.

Table 1: SSA residuals ARMA models

| Model | | ARMA | L-B p-value |
|---|---|---|---|
| Pre-COVID | Bus | AR(1) | 0.2481 |
| Pre-COVID | Rail | NA | 0.09423 |
| Post-COVID | Bus | AR(1) | 0.5352 |
| Post-COVID | Rail | AR(2) | 0.2649 |

### 4.2 Geographical Distributions

Despite the narrow global windows, in both pre-COVID and full-scale data emerging hot-spot analyses, the only pattern that was detected were "intensifying hot spots" in and around New York. Furthermore, both clusterings resulted in three clusters being generated, which are largely separated by order of magnitude. However, the results of the trend analyses are non-trivial, and thus have been included in Appendix B alongside the trend comparison map.

## 5 Analysis and Discussion

### 5.1 Ridership & Returning to "Normal"

General comments about the models:

- Bus is sinusoidal?
- Rail to continue growing?

As shown in Fig. 1 bus ridership is predicted to return to pre-pandemic levels in June 2026, while train ridership is predicted to recover in December 2027, shown in Fig. 2. However, there is an unexplained long-term sinusoidal component to the bus model, despite the rail model trending generally linearly, which is interesting in its own right.

With the peak of the long-term sinusoidal component of the bus model's trend falling around 2008, one might think that this has some sort of economic association, but this explanation fails to consider the relative "smoothness" of this component, which remains unexplained. Furthermore, the trend of the rail data appears to start "leveling off" starting around 2013, where at the same time there is a substantial increase in the noise variance. While the trend difference is unexplained, the noise variance may be related to increased reporting of smaller agencies.

### 5.2 Geographical Shifts

The maps in appendix B show no clear pattern to the pre-COVID trends, but the high prevalence of positive trends post-COVID indicates that it is still too early to identify subtle shifts in population.

## 6 Conclusions and Future Work

While it is still too early to gauge many of the long-term effects of COVID-19 on US public transit, there are still a variety of interesting side effects of attempting to do so anyways, such as general forecasting of trends "if the pandemic never happened", and the interestingly uninteresting geographical distribution of pre-covid ridership trends. Running similar analyses when more data is available may return better results, as well as using a more powerful model that is able to better handle the unprecidented nature of the COVID-19 pandemic.

# A Sigmoid Convergence

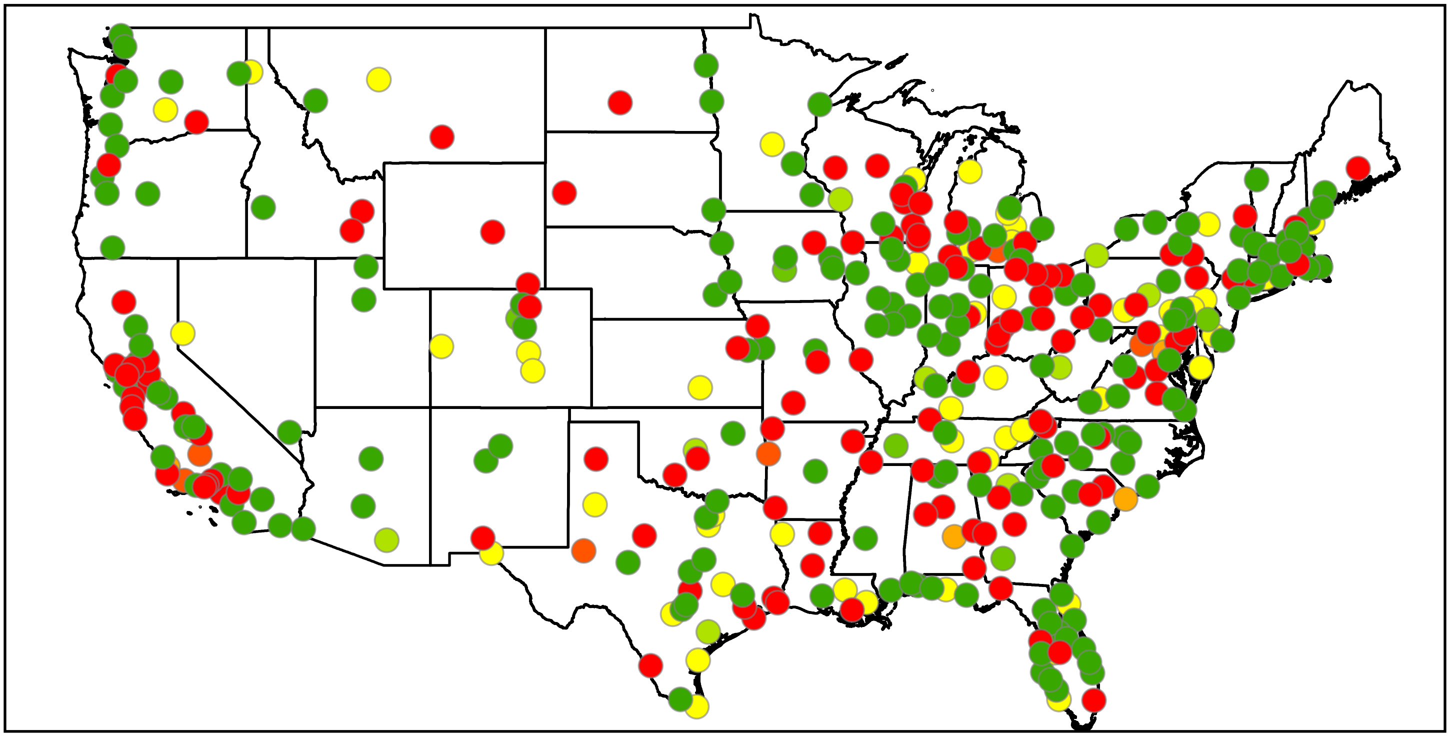
The ”sigmoid convergence” model used in this paper reflects that of a weighted average, using a sigmoid function to gradually shift influence of the model from the post-pandemic trends to the (assumed) true pre-pandemic model. The sigmoid function used to model this shift is shown in Equation 1, which was determined experimentally where $\delta$ is the time distance between the end of the false model and the point of convergence.

$$\hat{\sigma}(x) = \frac{1}{1 + e^{-(1/3(x - \frac{\delta}{2}))}} \tag{1}$$

Thus, the model in the time range indicated by $\delta$ that follows from this approach is shown in Eqn. 2 for ”true” model $M$ and ”false” model $W$.

$$\hat{X}_t = \hat{\sigma}(t)\hat{M}_t + (1 - \hat{\sigma}(t))\hat{W}_t \tag{2}$$

# Appendix B: Trend Comparisons

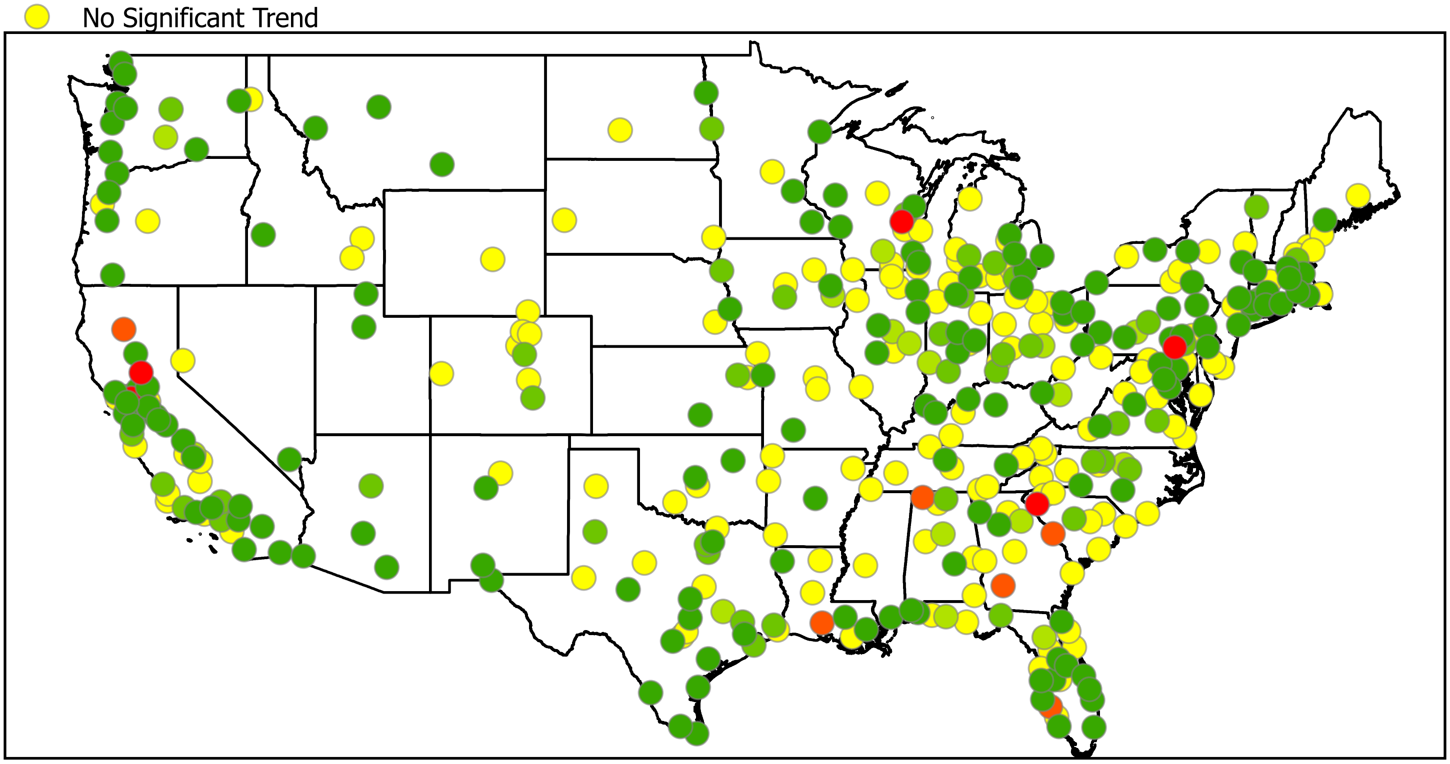


Legend

- Up Trend - 99% Confidence
- Up Trend - 95% Confidence
- Up Trend - 90% Confidence
- No Significant Trend
- Down Trend - 90% Confidence
- Down Trend - 95% Confidence
- Down Trend - 99% Confidence

PRE

POST



Works Cited

Barrero, Jose Maria, et al. “Why Working from Home Will Stick.” *National Bureau of Economic Research*, Apr. 2021, https://doi.org/10.3386/w28731. Accessed 12 Dec. 2023.

Cutler, Adele, and Leo Breiman. “Archetypal Analysis.” *Technometrics*, vol. 36, no. 4, Nov. 1994, pp. 338–47, https://doi.org/10.1080/00401706.1994.10485840.

Delventhal, Matthew, and Andrii Parkhomenko. *Spatial Implications of Telecommuting in the United States*. National Center for Sustainable Transportation, Feb. 2023, doi.org/10.7922/G2GB22D2. Accessed 12 Dec. 2023.

Parker, Kim, et al. *How the Coronavirus Outbreak Has – and Hasn’t – Changed the Way Americans Work* . Pew Research Center, 9 Dec. 2020.

Rappaport, Jordan. “Hybrid Working, Commuting Time, and the Coming Long-Term Boom in Home Construction.” *The Federal Reserve Bank of Kansas City Economic Review*, vol. 107, no. 4, Sept. 2022, https://doi.org/10.18651/er/v107n4rappaport. EBSCO. Accessed 12 Dec. 2023.