Forecasting the Post-COVID Recovery of Washington DC Public Transportation Usage

Steven VanOmmeren

Boston College

Abstract

The COVID-19 lockdowns of early 2020 saw the use of public transportation plummet. Since lockdown restrictions have loosened, the use of public transportation has slowly increased and in some cases exceeds pre-pandemic levels. Residents of the DC-Maryland-Virginia (DMV) area heavily rely on Washington DC's public rail system, run by the Washington Metropolitan Area Transit Authority (WMATA). I use daily ridership data provided by WMATA to analyze rail transport usage from March 2020 through March 2024. I then make predictions of average daily ridership for April 2024 through March 2025. I use a range of statistical models and compare their predictive accuracy, finding that ridership can be reasonably predicted using simple models. I also compare my model predictions with WMATA's budget forecasts.

Introduction & Significance

Forecasting is the prediction of future events using historical information. Accurate forecasts are vitally important to ensuring that appropriate business and government decisions are made. When sufficient data are available, statistical forecasting methods can approximate the time-series behavior of relevant variables using historical data of both the variable of interest and any additional relevant variables that may increase a model's performance. In this paper, I use a variety of statistical forecasting methods, some of which make use of additional control variables, to predict WMATA ridership.

Forecasting long-term ridership patterns is critical to transportation systems. WMATA is required to predict its future costs and revenues in its annual budgets for government approval; revenue predictions require an accurate estimate of future ridership. Additionally, there has been much discussion of the lingering effects of COVID-19 on public transportation usage, and whether the current state is still trending upwards or a "new normal."

With 5 years passed since the start of COVID-19 lockdowns, we have sufficient data to analyze the post-COVID recovery of train ridership in Washington DC. I analyze the average daily train ridership per month, leaving a prediction of bus rides for future work

Methods

Data

I use two sources of data:

- Daily rail boardings in Washington DC (Source: WMATA)
  I average to the monthly level.
- Monthly average consumer gasoline prices (Source: U.S. EIA)

This yields 60 observations (5 years of monthly data).

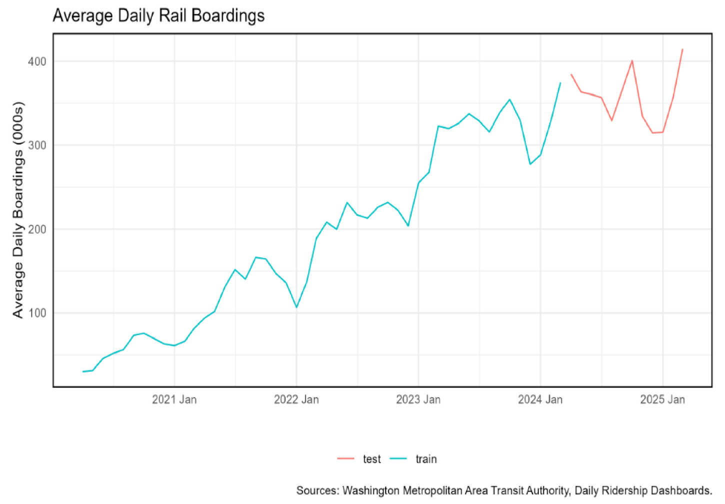
I split the data into training and test subsets, with the first 48 months used as training data.

Analysis

I test 5 predictive models of average daily train rides:

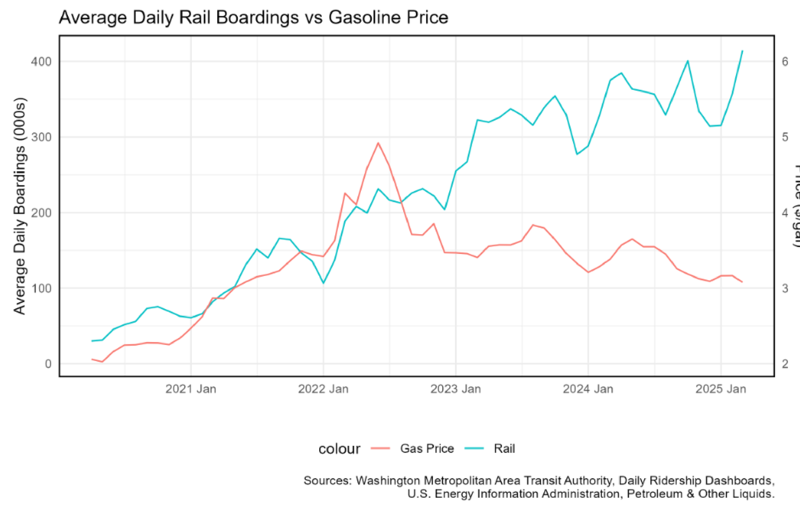
- Seasonal Naïve
- ETS: includes a damped-additive trend component and additive seasonal component.
- TS Regression: regress Rides on gas price, seasonal indicators, and a piecewise-linear trend with knots at Jan. 2021, Jan. 2022, Jan. 2023, and Jan. 2024.
- ARIMAX: includes gas price as an exogenous regressor.
- Ensemble: a simple average of the above models.

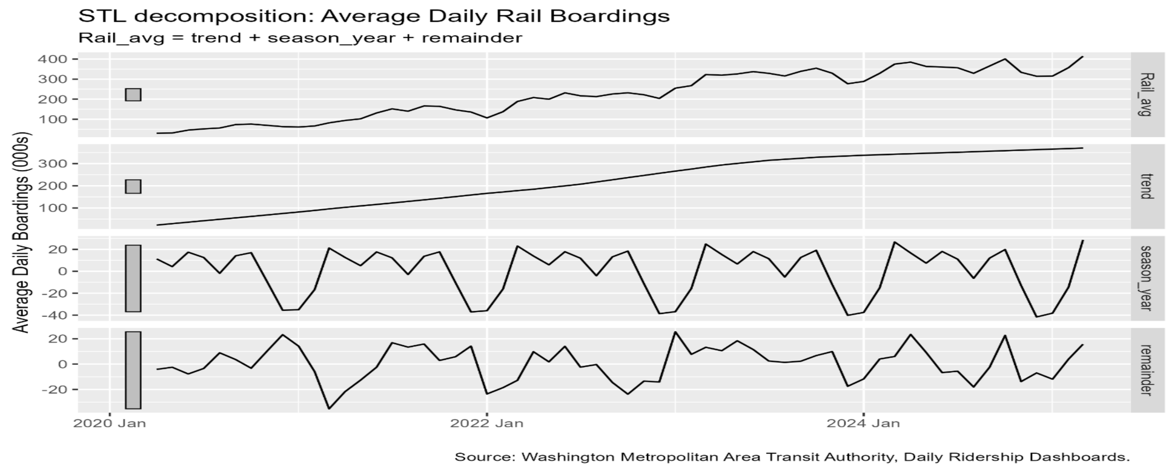
For out-of-sample predictions, I use a Seasonal ARIMA(1,1,1) model to forecast gas prices.



Variable of Interest, Training and Testing Periods

Average Ridership vs Gas Prices





STL Decomposition, Rail Ridership

Results

During the 5-year period following COVID, average daily rail use in Washington DC has consistently increased, though with noticeable seasonal changes: early Spring months appear to have higher volumes, while November through January appear to have lower volumes.

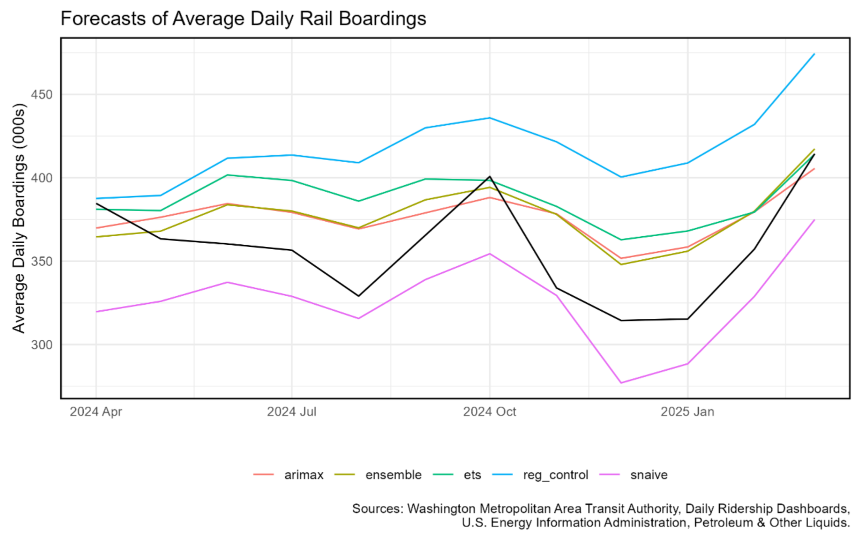
Gas prices appear to have the same trend as rail ridership until mid-2022, it then breaks from the increasing trend, decreasing to just over $3 per gallon in March 2025.

Applying STL decomposition to ridership shows the trend component has increased linearly until around 2023, where the trend increases at a slower rate.

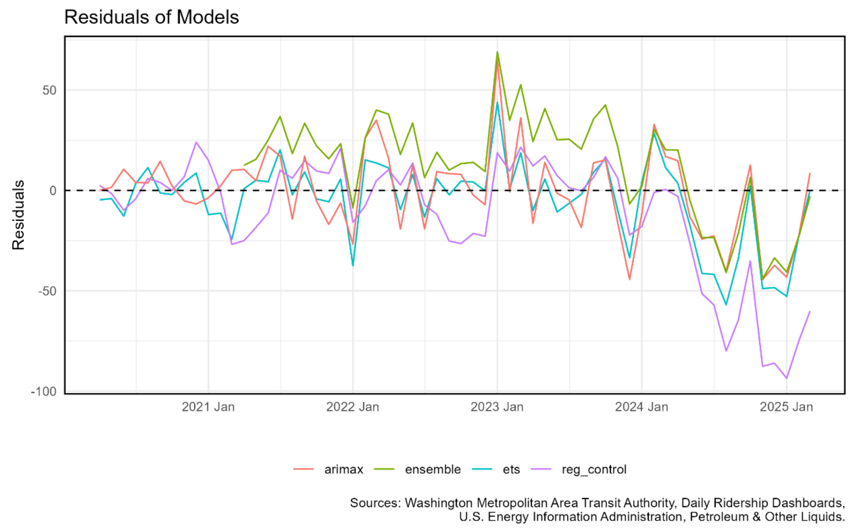
All models capture the monthly seasonal pattern in the test data (black line). ARIMAX and Ensemble models fit the data best, though they underestimate the strength of monthly seasonality in the test data. TS Regression (blue line) consistently overestimates ridership, while Seasonal Naïve (purple line) underestimates.

The in-sample residuals are all quite similar, but by April 2024 they diverge significantly from the horizontal line at y = 0. Interestingly, While the ensemble models in-sample residuals appear to be among the worst, the out-of-sample residuals are closest to 0 and visually appear the best.

On the training data, the ETS model has the lowest RMSE, MAE, and MAPE, while the Seasonal Naïve model performs worst by a large margin. But when we look at out-of-sample performance, the results are reversed. The Ensemble model has the best out-of-sample performance.



Out-of-Sample Forecasts, Rail Ridership. The black line is actual data, while colored lines are model predictions.



In- and Out-of-Sample Residuals (actual minus predicted rail ridership)

.model	.type	ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE	ACF1
ets	Training	1.12	13.98	10.28	-1.02	7.59	0.12	0.15	-0.01
reg_control	Training	0.00	14.09	11.58	-0.26	8.47	0.13	0.15	0.56
arimax	Training	3.83	18.77	14.10	2.54	8.37	0.16	0.20	-0.01
ensemble	Training	23.95	28.41	24.81	10.96	11.54	0.28	0.31	0.16
snaive	Training	89.29	93.14	89.29	42.26	42.26	1.00	1.00	0.61
ensemble	Test	-19.27	27.32	23.70	-5.80	6.95			0.15
arimax	Test	-18.72	27.78	24.75	-5.71	7.23			0.09
snaive	Test	31.35	34.72	31.35	8.63	8.63			-0.13
ets	Test	-29.75	36.70	30.77	-8.82	9.08			0.17
reg_control	Test	-59.94	65.46	59.94	-17.31	17.31			0.43

In- and Out-of-Sample Performance Metrics for my models.

Discussion

Seasonal Naïve model underpredicts by implicitly assuming no year-over-year growth, while ETS overpredicts by assuming that existing trends would continue going forward.

The Ensemble method performs well because it balances out the underpredictions and overpredictions of my models.

Compared to Mucci and Erhardt (2018), my best model had a mean average percent error of 6.95%, versus their 18% out-of-sample error. Similarly, Hoque et al. (2024) reported an average of 25% overestimation of ridership in public transportation projects. While there are differences in estimation strategies, my Ensemble model is well within the out-of-sample forecasting accuracy of the academic literature.

WMATA forecasts 113.7 million ridership for Jul. 2024 through Jun. 2025. The available data show total ridership of 127.7 million for Apr. 2024 through Mar. 2025. My ensemble model predicts ridership of 137.7 million, 7.8% higher than actual data. Interestingly, my ensemble estimates are closer to WMATA's forecasted ridership for FY2026, rather than FY2025. My results show that WMATA's forecasted decrease going into 2025 does not appear to be correct, and that post-COVID ridership recovery is still ongoing.

To improve my model, I would source additional exogenous variables that influence rail ridership. I could modify my analysis to forecast train station-level ridership, rather than system-wide volume. I would incorporate additional modes of transportation into the model to account for substitution from rail to buses, cars, bicycles, etc.

Conclusion

Forecasting public transportation usage is a difficult task post-COVID. I ran five different time series models on 2020-24 rail ridership data to forecast 2025 volumes. The most useful model was an Ensemble average of the other four models. My models have out-of-sample predictive accuracy on par with academic research and WMATA's own forecasts. My results show that Washington DC rail ridership appears to still be on a post-COVID recovery path, with transportation volume continuing to increase in the short term.

References

Brey, Jared (2023). "Predicting Future Transit Ridership Is Trickier Than Ever".

Hoque, Jawad Mahmud et al. (2024). "Are public transit investments based on accurate forecasts? An analysis of the improving trend of transit ridership forecasts in the United States".

Mucci, Richard A. and Gregory D. Erhardt (2018). "Evaluating the Ability of Transit Direct Ridership Models to Forecast Medium-Term Ridership Changes: Evidence from San Francisco".

Pinho, Paulo et al. (2024). "The application of direct ridership models in the evaluation of the expansion of the Porto Light Rail Transit".

BOSTON COLLEGE
Woods College of Advancing Studies