The National Center for Smart Growth Research and Education
University of Maryland, College Park
in partnership with
The Maryland Department of Transportation

Time-of-Day Direct Ridership Model for Maryland Rail Transit

Final Report

July 21st, 2017

Prepared by:

Chao Liu, Ph.D., Faculty Research Associate

Hiroyuki Iseki, Ph.D., Research Faculty, Assistant Professor

Sicheng Wang, Research Assistant

Table of Contents

List of Tables

List of Figures

Executive Summary

In cooperation with the Maryland Transit Administration (MTA), NCSG has developed a direct ridership model (DRM) for Maryland's rail transit systems. This DRM is a regression-based model to estimate rail transit ridership *at the station level by time of day (AM peak, PM peak, and off-peak periods)* based on a set of selected location-specific factors related to transit services, land use and built environment, and demographics both at the place of residence and place of work within the walksheds of rail stations. The walksheds are defined by three walking distances: quarter-mile, half-mile, and one-mile from each station of four rail systems—Baltimore Light Rail, Baltimore Metro subway, MARC commuter rail service, and the Metrorail of WMATA (Washington Metropolitan Area Transit Authority). More specifically, walksheds are created using the most updated pedestrian network that can better represent characteristics of pedestrian-oriented walking areas.

The model development has involved many analyses including pre-regression descriptive analysis, regression model development, and post-regression diagnostics. The results suggest that transit service related variables are the strongest predictors of ridership and the results are consistent across models of all time periods. Parking capacity has achieved higher coefficient in the AM model, suggesting that it might be the case that more riders used park-and-ride services in the AM periods. Feeder bus service is significant and positive in both PM and off-peak models.

In the variables of land use and the built environment, employment and household are the two key predictors. As expected, the number of households is significant and positive in the AM model but is not significant in the PM and off-peak models. Employment is significant in both PM and off-peak models, but is not significant in AM model. Employment categorized as midday and weekend jobs is significant and positive in the off-peak model for non-MARC stations and the coefficient is even higher than the total employment numbers in the PM model.

It is also interesting to see how the dummy variables capture the variations of the ridership prediction by system and by locations.

Finally, model limitations and future model implementations are discussed.

1. Introduction – Study Objective

A transportation system is essential to the lives of residents and workers, by facilitating the movement of people and goods, it connects them to places where people conduct economic, social, and other activities. Transportation systems require continuous development, improvements, and maintenance to ensure the efficient growth of economic opportunities. At the same time, a system's adverse environmental impacts must be prevented or mitigated. Within the State of Maryland, the Maryland Transit Administration (MTA) serves communities by encouraging the use of public transportation, stimulating local and regional economies, and minimizing negative impacts on the environment.

Maryland's 1992 Economic Growth, Resource Protection, and Planning Act provides a vision for future growth in the State (MDP, 2013). The act is part of progressive smart growth and land use policies that actively address the adverse impacts of sprawling development. The State also started the Smart, Green and Growing Initiative that requires coordination among multiple agencies to achieve a more sustainable future through transportation improvements, economic development, community revitalization, and environmental restoration efforts (MDOT, 2009). As part of this initiative, Transit Oriented Development (TOD) has gained attention as a way to improve transportation and land use coordination, encourage future development around existing and planned transit stations, and obtain the maximum value from transit investment. Ideally, TOD should generate the further benefits of less traffic congestion, lower levels of vehicle emissions, and enhanced travel choices.

Transit Oriented Development, first devised in the early 1990s by architect and urban planner, Peter Calthrope, is a planning strategy that coordinates development with public transit investment. TOD is commonly adopted in regional transit plans to achieve economic growth, sustainable land use patterns, and pedestrian-friendly communities (Cervero, 1989; Calthrope, 1993). It seeks to create a synergy between land use and transit with characteristics such as moderate to high density, a mix of land uses, good street connectivity, a built environment that supports varied travel modes and reduced parking (Cervero et al 2004).

TOD is also considered an important component of sustainable regional and local planning. Similar to Maryland's initiative, the U.S. Department of Housing and Urban Development (HUD), in coordination with the U.S. Department of Transportation (DOT) and the U.S. Environmental Protection Agency (EPA), created the Sustainable Communities Regional

Planning grant program to implement the Federal Livability Principles, under which metropolitan planning organizations can pursue transportation and infrastructure investments that are integrated with housing, land use, economic and workforce development projects.

A successful TOD implementation requires coordination among governments and public agencies at the state, regional, and local levels, as well as the private sector, through the transportation and land use policies, the planning process, public private partnerships, infrastructure investments, and other efforts. An important step in successfully implementing TOD that increases ridership is to identify the factors that influence regional and local rail ridership levels.

This report develops a time-of-day direct ridership model (DRM) for Maryland's rail systems. The model allows users to estimate rail transit ridership at the station level by time of day based on a selected set of location-specific factors related to transit services, land use and built environment, and demographics by place of residence and work.

The next section describes the study area, data, and data sources used for this study. The third section briefly describes the analytical method used to conduct direct ridership modeling for Maryland's rail systems. The fourth section presents model development results and discusses findings. The report concludes with the summary of findings.

2. Study Area, Data, and Data Sources

2.1. Rail Stations in the Model Development

This study covers the State of Maryland, which consists of 23 counties and one independent city. In 2012, the state's estimated population was 5.8 million (U.S. Census, 2013). Within the state, there are four passenger rail systems operated by two different entities. The Maryland Transit Administration (MTA) manages and operates the MARC train, the Baltimore Metro, and Baltimore Light Rail Link. MARC is an intercity rail service connecting Baltimore and Washington D.C. with surrounding counties: Anne Arundel, Baltimore, Frederick Hartford, Howard, Montgomery, and Prince George's, as well as West Virginia (Figure 1).

The Washington Metropolitan Area Transit Authority (WMATA) operates Metrorail within Maryland providing passenger rail service to Montgomery and Prince George's Counties within “the Baltimore-Washington Metropolitan area.” Table 1 summarizes the number of stations by rail system in Maryland.

Figure 1 Study Area and Rail Systems

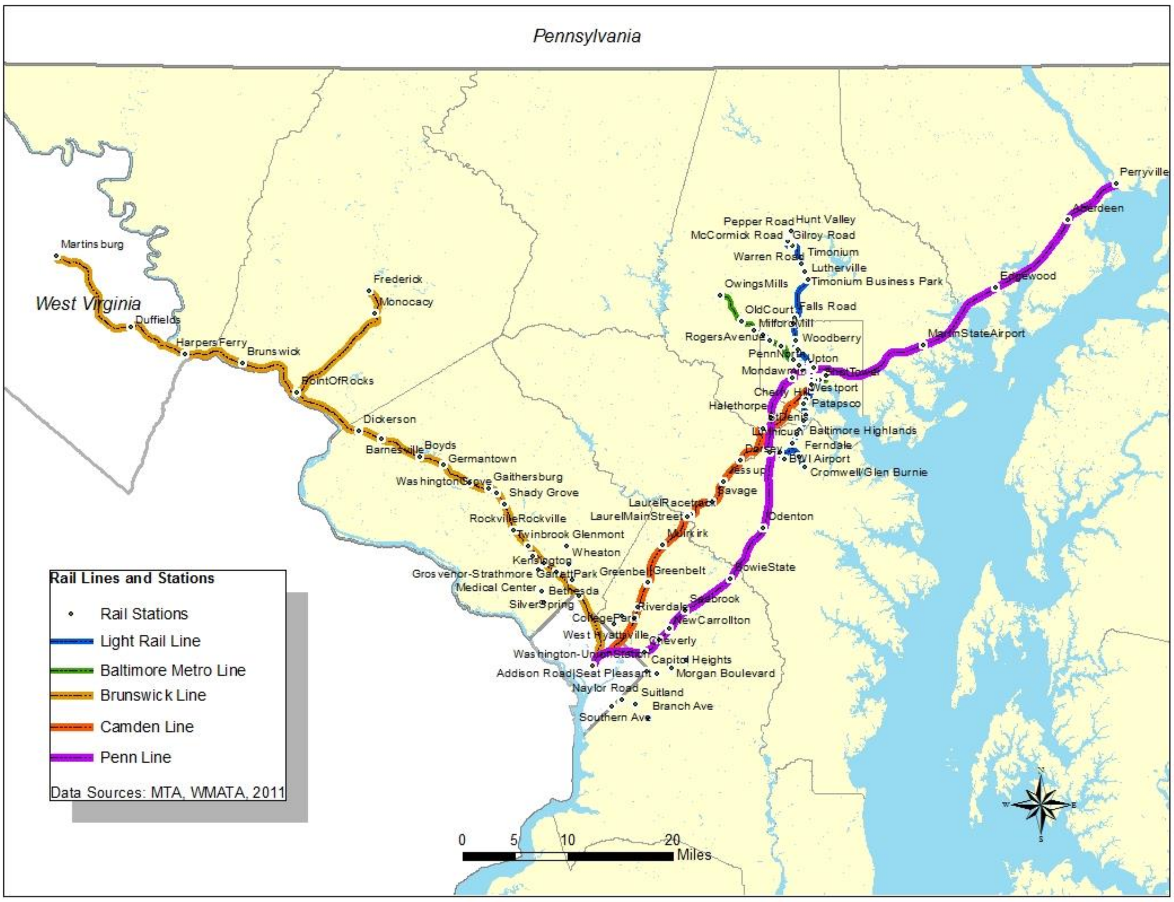


Table 1 Rail Stations by System

System	Number of Stations
MTA MARC	39
MTA Baltimore Metro	14
MTA LRT	33
WMATA Metrorail	26
Total	112

2.2. Data and Primary Data Sources

NCSG gathered and organized information on land use and other location-specific characteristics of the rail stations and assembled the data set for the DRM development, combining data from multiple sources. Table 2 summarizes the variables and data sources. Three different categories of variables were collected: transit operations and parking, land use and built environment, and socio-demographic.

Table 2 Data and Data Sources for the DRM Development

	Descriptions	Source (year)
Dependent Variables	Ridership boardings AM Peak, PM Peak, Off-peak	MTA and WMATA, 2015-2016
Independent Variables	**Descriptions**	**Source (year)**
Transit Operational and Parking	Number of trains by time of day, number of bus lines, number of parking spaces	MTA and WMATA, 2015-2016
Built Enviroment and Land Use (0.25, 0.5 and 1-mile buffer areas)	households in 0.5 miles, jobs in 0.5 miles, job accessibility, street connectivity,mixed land use, distance to downtown	U.S. Census 2010, LEHD 2014, MTA, MSTM
Socio-Demographics (0.25, 0.5 and 1-mile buffer areas)	density of households; median income, household types, poverty level, employment status, housing types (owner occupied vs. renter occupied), and median gross rent; Jobs by industry, jobs by income, jobs by education attainment, midday and weekend jobs	ACS 2010-2014 5-year estimates LEHD 2014

Definition of Station Area

The most common way of defining walksheds has been to establish a buffer from a station using a straight line distance (NCSG, 2012; MDP, 2016). This approach fails to account for pedestrian barriers, such as highways and rivers, as well as street variations that can influence connectivity in a study area. To overcome these limitations, we constructed walksheds using Network Analyst in ArcGIS, based on the Open Street Map network. Limited access highways were removed to better represent pedestrian-oriented walking distance to rail stations. Walksheds with different distance thresholds (quarter-mile, half-mile, and one-mile) were created to test how sensitive rail ridership is to variations in land use and the built environment among different walksheds and how sensitive ridership response is to such variations.

Rail Ridership by Station by Time of Day

The regression analysis for model development uses rail ridership by station by time of day as a dependent variable, so it was critical to obtain station-level rail ridership by time of day, rather than daily ridership used previously. The day was divided into four time periods at the data collection-processing stage: AM peak (6:30 a.m. to 8:29 a.m.), Midday (8:30 a.m. to 3 p.m),. PM peak (3 p.m. to 6:29 p.m.), and other (before 6:29 a.m. and after 6:30 p.m. within the hours of operation).

NCSG obtained WMATA's Metrorail station ridership by time of day on five metro lines (Green, Orange, Red, Yellow and Blue Lines). While MTA had reasonable data for the Baltimore Metro stations, they could not easily supply station-level ridership by time of day for MARC and Baltimore LRT.

To make up for this, we computed MARC average time-of-day ridership by station based on the average weekday ridership of each MARC station over 12 months in FY2015-16. Light rail ridership data by station and by time of day were not available at all. To remedy this absence, NCSG spent substantial time exploring a variety of approaches to collect ridership data, including the MTA's most recent passenger survey data.[1] After several different approaches, NCSG developed a way to reasonably estimate LRT ridership by station by time of day, using five different data sets:

- National Transit Database Survey FY2015-16
- MTA Light Rail Operation Data FY 2015-16
- MTA Light Rail Average Weekday Boarding Counts FY 2015-16
- WBA On Board Survey
- Baltimore Metro Ridership data.[2]

(See Appendix A for more details on this approach.)

It is worth noting that several stations were identified as multimodal stations served by more than one rail system (Table 3). For example, the Lexington Market station in downtown Baltimore, is served by both LRT and Baltimore Metro. Similarly, the New Carrollton station is served by both MARC and WMATA Metrorail. This could lead to a modeling challenge because stations may have very different levels of ridership despite sharing the same location-specific variables.

---

[1] This data was collected and processed by MTA's consultant, WBA Research. Although WBA research initially indicated that they could get a reasonable estimate of LRT ridership by station by time of day, they did not provide the data after all, forcing the NCSG to explore another approach.

[2] The obtained estimates were considered reasonable by MTA.

Table 3 Multimodal Stations

Systems	Stations
MARC & WMATA Metrorail	College Park New Carrollton Silver Spring Rockville Greenbelt
MARC & Baltimore LRT	Penn Station BWI Airport Camden Yard
Baltimore LRT & Baltimore Metro	Lexington Market

Independent Variables

To develop DRM, independent variables were identified and grouped into three categories: transit service, land use and built environment, and socio-demographics. Transit service data included the availability of park-and-ride facilities, feeder bus services, train service frequency, and whether it is a terminal station, which tend to have more boardings (Cervero, 2016). Land use and built environment data included population and employment (numbers and density), employment by industry sectors, skill levels, wages, land use mixed index, street connectivity, regional accessibility, distance to downtown, and walk scores. Socio-demographic variables included the number of high schools, income levels, vehicle ownership, employment status, age, housing types, rent levels, and poverty levels.

Initially, more than 130 independent variables were obtained from a variety of sources. After several iterations of selecting variables that are both statistically significant and have practical use for MTA, about 30 were selected for final model testing. (A detailed list is in Table 10, Appendix B.)

Transit Service Variables

Level of service (LOS) was measured by the number of trains of both directions in each time period. The number of trains at Baltimore Metro and WMATA stations was provided by MTA and WMATA, respectively. The number of trains at LRT and MARC stations was not readily avaiable and were calculated by NCSG based on timetable schedules[3]. Parking and feeder bus service were the variables used to estimate boardings at a given station, because

[3] http://mta.maryland.gov/light-rail; http://mta.maryland.gov/marc-train

they capture entries and extis from other modes. Information on parking and bus connection for Baltimore Metro, light rail, and MARC were obtained from MTA's TOD data portal and TOD profile tool[4]. Parking and bus connection data at WMATA stations were provided by WMATA. Both parking and bus connection varibles were also treated as a dummy variable in the model.

Land Use and Built Environment Variables

Five variables are used to describe land use and built environment: households and employment, street network connectivity, mixed land use, accessibility, and distance to downtown.

The number of households was obtained at the census block level from the U.S. Census 2010 Summary File 1 (SF1). Employment numbers were obtained at the census block level from the Longitudinal Employer and Household Dynamics (LEHD) 2012 (LEHD, 2014). To better capture the impacts of employment type on ridership, we used the North American Industry Classification System (NAICS) codes to classify types of jobs by time of day (i.e. midday and weekend jobs). The household and employment variables were applied to station walksheds by using the areal allocation approaches.

Street Network Connectivity – street network connectivity is measured by the number of intersections (except cul-de-sacs) within station walksheds. Open street map networks were used to calculate this measurement. A station's connectivity increases as the number of intersections within the station walkshed increases.

Land-use mix index – This study considers three land use types—residential, commercial and industrial. A land-use mix index is used to capture how evenly land use floor area is distributed within station walksheds. (Details on the calculation of the land-use mix index are in Appendix E.)

Accessibility – A gravity-based accessibility measure is used to define accesibility from one zone to all other zones. This measure provides accurate estimates of the accessibility of zone *i*

---

[4] https://data.maryland.gov/Transportation/MTA-Transit-Oriented-Development-TOD-Data/cqt2-ypem/data. We found some inconsistencies between data from the dashboard and data from the MDP TOD profile tool. They were corrected in the final datasets. MDP conducted an analysis on TOD prioritization and the underlying data were also included in the initial dataset for DRM development. However, these data were computed as composite indices and are only available at the half-mile buffers.

to opportunities in all other zones *j* in the region, where fewer and/or more distant opportunities provide diminishing influences (Geurs and Wee, 2004). (Details on the calculation process are in Appendix E.)

Distance to CBD – The central city remains the main trip attractor in the Baltimore-Washington Metropolitan region. We would expect that stations closer to the central business districts (CBD) would have higher ridership. Initially, distance to the center of Baltimore and distance to the center of Washington D.C. were calculated. Between the two varaibles, the shorter one was selected as a measurement for the distance to CBD.

Socio-demographic Variables

Several socio-demographic variables capturing various charateristics of station areas were collected. A recent study conducted by the Baltimore Education Research Consortium (BERC) found that about 60 percent of high school students rely on public transportation to commute to school. These students represent a large public transportation user group, especially during the peak time on school days[5]. As suggested by MTA staff, the number of high schools within station walksheds was collected. Other socio-demographic variables include income, vehicle ownership, employment status, poverty, population in the labor force, population of young and elderly, and housing types were collected from the American Community Survey 2010-2014 five-year estimates.

3. Regression Analysis

Based on the complete data set, NCSG has developed the DRM for three time periods: AM peak, PM peak, and off-peak periods. Initially, ridership data were collected and processed for four time periods. However, due to the limited off-peak service of MARC lines, many MARC stations with suspended and reduced service have no ridership. To maintain statistical validity, ridership in the midday and other time periods were combined.

DRM is a regression-based model that relates transit ridership to factors that affect ridership. In the case of rail ridership modeling, the influential factors typically include types and attributes of land use, characteristics of the built environment and urban design,

[5] BERC. 2017. Getting to High School in Baltimore: Student Commuting and Public Transportation. http://baltimore-berc.org/wp-content/uploads/2017/01/GettingtoHighSchoolinBaltimoreJanuary2017.pdf

demographics of residents and workers near rail stations, as well as transit service levels and facilities (Kuby, Barranda, and Upchurch 2004; Cervero, Murakami, and Miller 2009; Gutiérrez, Cardozo, and García-Palomares, 2011). Applying DRMs to transit ridership modeling has become more popular in recent practice (Cervero 2007; Fehr & Peers 2013).[6]

In this report's analysis, Ordinary least squares (OLS) regression was used to estimate a DRM. Equation (1) shows the linear relationship between station boardings Y_i and all the independent variables. Transit services, land use characteristics, and socio-demographic variables are represented by X_1, X_2, and X_3, respectively. α is the constant term. β_1, β_2, and β_3 are the coefficients estimated from the linear regression, and ε_i represents the unobserved random error.

$$Y_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon_i \qquad \text{Equation (1)}$$

All continuous variables were transformed using the logarithmic function. In this way, an estimated coefficient can be interpreted as *elasticity*—a percentage change of ridership in response to a percentage change of a particular independent variable, holding all other independent variables constant.

Several model specifications were tested using different combinations of the independent variables. NCSG also tested the sensitivity of transit ridership in response to the built environment in three different catchment sizes, and found no significant difference in model's predictive power. Nevertheless, the results suggest that most of the variables performed best when using the half-mile buffer in the model.

Variables with multicollinearity issues were eliminated by inspection of the Variance Inflation Factor (VIF) and the correlation matrix. All insignificant independent variables (*p-value*>0.1) were systematically eliminated from the final model. Variables with a *p-value* slightly higher than 0.1 were kept when they are considered important explanatory variables intended to test for policy. Several tests were conducted to ensure that the OLS regression assumptions are satisfied. Scatter plots were created to check the relationships between the rail boardings and individual independent variables. Heteroscedasticity was examined by generating plots of residuals versus predicted values.

---

[6] There are a few cases of DRM applied to bus ridership modeling, including a study by Cervero, Murakami, and Miller (2009).

NCSG also investigated appropriate approaches to categorize stations based on transit service, land use characteristics, and socio-demographics. One approach we took was cluster analysis, a statistical approach to identify groups (clusters) of observations in a way that the observations are more similar in the one group than those in a different group. Although several iterations of analysis were conducted based on different combinations of variables and different numbers of clusters, the cluster analysis did not help categorize the stations by types that could incorporated in the DRM to make it have more prediction power.

Another approach is to combine the rail system groups (i.e., Maryland's four systems) with geographic locations (e.g., in the Baltimore region or in the DC region, Montgomery County or Prince George's Country, or within or outside the City of Baltimore). This would create a more intuitive station categorization, by taking into account the interrelationship of rail ridership and each of the independent variables.

4. Results and Discussion

4.1. Descriptive Statistics

A series of analyses were conducted for the DRM development. Descriptive analyses were conducted first to provide an overall examination of the data, spatial patterns of ridership and key independent variables. These analyses also looked at variations of all the variables. Table 4 is an overview of rail ridership by time of day and by rail system, as well as the total ridership for all four rail systems combined. While data were collected for four time periods during one day, for modeling, the midday and other time periods were combined to create the "off-peak" period.

On average, more than 10,000 passengers per day use rail transit in Maryland. Among the four rail systems, more than 50 percent ride WMATA Metrorail, about 34 percent ride Baltimore Metro, six percent ride Baltimore light rail, and five percent ride MARC. Ridership in the AM peak is generally higher than other time periods for MARC and WMATA, while ridership is relatively evenly distributed throughout the day for Baltimore light rail and Metrorail.

Table 4 Ridership by Time of Day and System

Variables	No. of Samples	Mean	St. Dev.	Min.	Max.
All Systems					
Boardings in AM Peak Period	110	1048.82	1528.10	7	9411
Boardings in PM Peak Period	103	518.95	750.44	1	3714
Boardings in Off-peak Period (Midday & Other)	105	584.86	736.23	1	3822
Baltimore Light Rail					
Boardings in AM Peak Period	33	201.55	192.86	35	980
Boardings in PM Peak Period	33	236.08	192.92	35	936
Boardings in Off-peak Period (Midday & Other)	33	232.80	217.83	16	1086
MARC					
Boardings in AM Peak Period	37	405.97	542.73	7	2316
Boardings in PM Peak Period	30	85.63	147.45	1	625
Boardings in Off-peak Period (Midday & Other)	33	86.57	158.63	1	863
Baltimore Metro					
Boardings in AM Peak Period	14	1029.71	512.37	484	2533
Boardings in PM Peak Period	14	1201.00	909.58	326	3103
Boardings in Off-peak Period (Midday & Other)	14	1185.86	577.46	398	2252
WMATA					
Boardings in AM Peak Period	26	3049.29	1962.48	941	9411
Boardings in PM Peak Period	26	1010.69	975.54	191	3714
Boardings in Off-peak Period (Midday & Other)	25	1370.76	839.92	340	3822

Note: The number of samples in each time period is limited to those that have more than one train in each time period.

Among all the systems, WMATA stations, on average, have the most frequent service (Table 5). For all three MARC train lines, only the Penn line runs in the off-peak period on weekdays.

Table 5 Number of Trains by Time of Day and System

Variables	No. of Samples	Mean	St. Dev.	Min.	Max.
All Systems					
No. of Trains in AM Peak Period	110	27.82	18.71	0	80
No. of Trains in PM Peak Period	103	43.53	32.67	2	140
No. of Trains in Off-peak Period (Midday & Other)	105	75.59	61.47	1	191
Baltimore Light Rail					
No. of Trains in AM Peak Period	33	33.09	10.94	12	49
No. of Trains in PM Peak Period	33	37.73	11.68	15	55
No. of Trains in Off-peak Period (Midday & Other)	33	123.58	41.02	52	191
MARC					
No. of Trains in AM Peak Period	37	8.81	5.81	2	22
No. of Trains in PM Peak Period	30	8.93	5.23	2	19
No. of Trains in Off-peak Period (Midday & Other)	33	6.36	6.71	1	19
Baltimore Metro					
No. of Trains in AM Peak Period	14	28.14	1.17	27	30
No. of Trains in PM Peak Period	14	49.21	1.05	48	51
No. of Trains in Off-peak Period (Midday & Other)	14	157.71	2.09	155	162
WMATA					
No. of Trains in AM Peak Period	26	50.15	14.14	40	80
No. of Trains in PM Peak Period	26	87.77	24.75	70	140
No. of Trains in Off-peak Period (Midday & Other)	25	57.64	6.42	53	65

The number of feeder buses that serve each station also varies among rail systems (Table 6). WMATA has the highest number of bus lines. Only limited feeder bus service is available at light rail and MARC stations. This is usually more so for suburban stations than for downtown stations, reflecting less use for rail ridership at suburban stations and the likely dependence on auto-access to those stations.

Table 6 Number of Bus Lines by System

Variables	No. of Samples	Mean	St. Dev.	Min.	Max.
All Systems					
No. of Bus Lines	112	5.92	8.07	0	45
Baltimore Light Rail					
No. of Bus Lines	112	3.82	5.32	0	23
MARC					
No. of Bus Lines	112	3.33	7.90	0	45
Baltimore Metro					
No. of Bus Lines	112	6.00	6.66	0	25
WMATA					
No. of Bus Lines	112	12.42	8.66	3	45

Table 7 presents the descriptive statistics for all the key independent variables. Many of the independent variables show large variations, suggesting that both land use and built environment variables, and socio-demographic variables vary significantly among all the stations.

Table 7 Descriptive Statistics of Selected Independent Variables

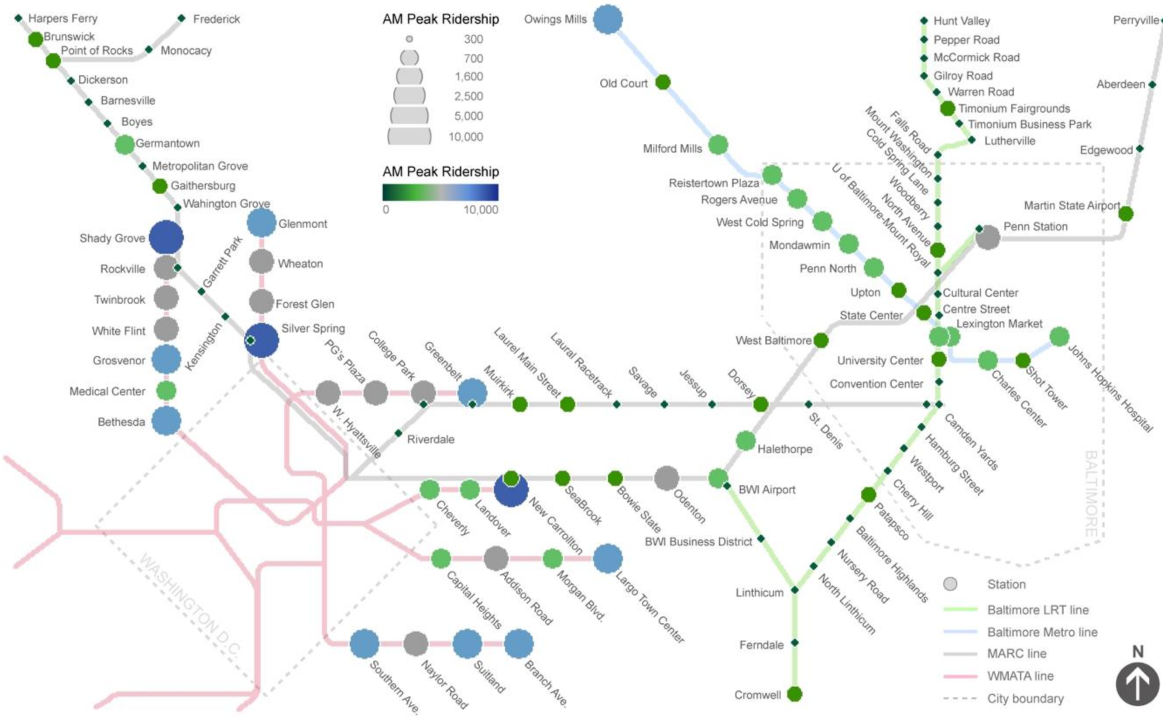
Variables	No. of Samples	Mean	St. Dev.	Min.	Max.
No. of Parking Spaces	112	441.455	812.768	0	5227.95
Auto accessibility	112	968554.80	406232.20	50749	1951165
Transit accessibility	112	682800.20	400953.80	94	1722465
Population density	112	5198.76	4762.32	0	18754
Number of households	112	2046.46	2175.99	0	9472
Number of households with single parents	112	161.16	193.22	0	1105
Number of households with no vehicle	112	469.81	747.33	0	3026
Median household income	112	67574.22	26830.02	0	140340
Number of jobs in all sectors	112	22531.19	43651.85	0.023	210001
Job density	112	5900.67	11408.17	0.006	54938
Number of midday and weekend jobs	112	855.64	1453.97	0	6938
Number of jobs in the public administration sector	112	1320.54	3128.57	0	16465
Total number of employment	112	10783.76	20859.46	0	107570
Employment density	112	18258.86	30958.72	3	137032
Number of intersections	112	96.70	94.90	0	347
Mixed land use index	112	0.50	0.29	0	1
Distance to downtown (shorter ones are selected between distance to Baltimore and distance to DC)	112	3.15	2.95	0.0002	15
Number of high schools	112	0.20	0.50	0	3

4.2.Spatial Distribution of Key Variables

A series of maps (Figures 2-7 in main document and Figures 8-13 in Appendix C) present the spatial distribution of ridership, household, employment, feeder bus connections and parking facilities. These maps can facilitate a better understanding of the relationships between ridership and potential determinants.

AM ridership on both the WMATA and Baltimore Metro systems is relatively evenly distributed, except that their terminal stations have larger ridership numbers in AM peak. Ridership of the light rail system is higher in downtown Baltimore. Penn Station on the MARC line has a larger ridership to bring Baltimore residents to downtown D.C.

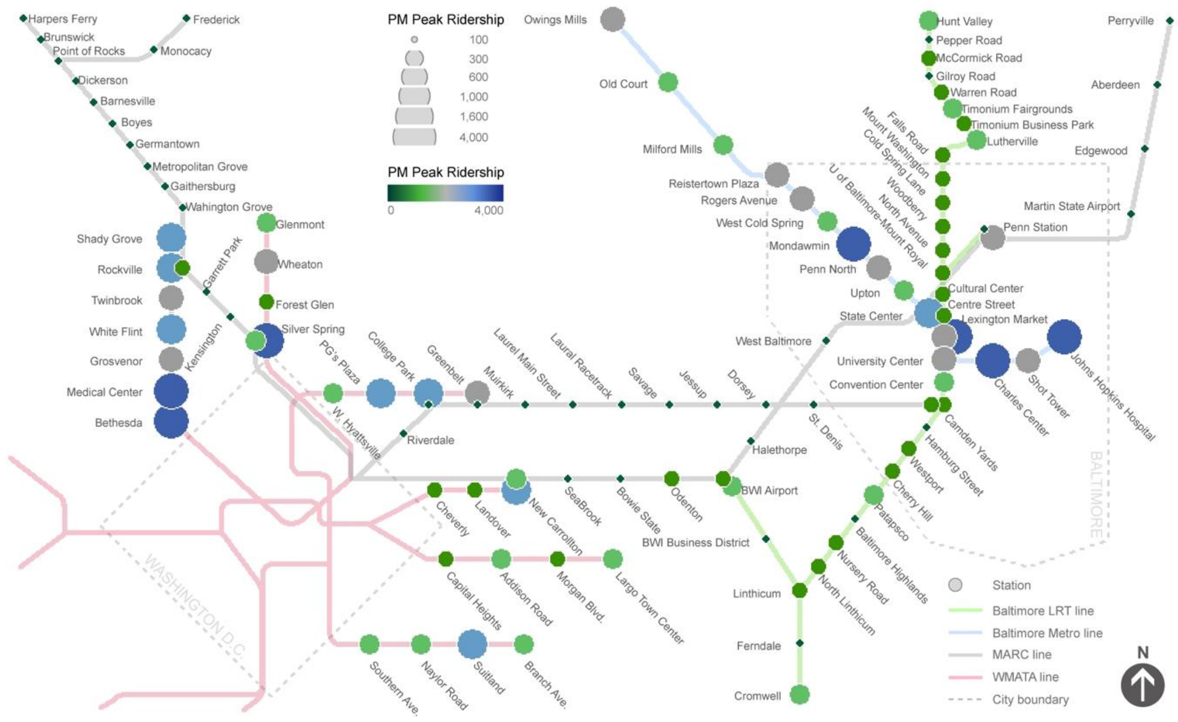
Figure 2 Ridership by Time of day: AM Peak



Note: For presentation purposes, the dimensions of this and subsequent maps are set differently from the actual maps.

PM ridership is relatively higher in downtown Baltimore for both Baltimore Metro and the light rail system. Stations with the highest ridership and jobs in station areas are Johns Hopkins Hospital, Charles Center, Lexington Market, State Center, and Penn North on the Baltimore Metro, and University Center on the light rail system. In the D.C. suburbs, many stations on the west side of WMATA's Red Line also have higher ridership due to job concentration in those areas.

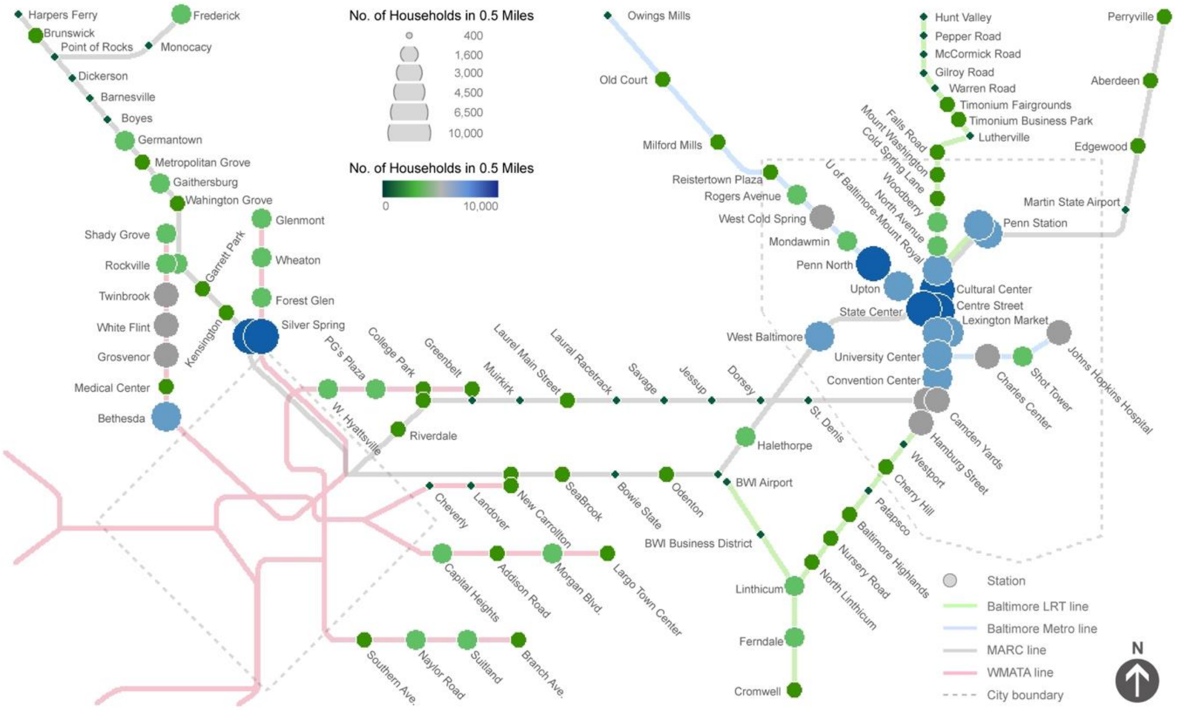
Figure 3 Ridership by Time of day: PM Peak



As major trip generators, the locations of households and employment play important roles in affecting ridership demand. Much of the region's employment is concentrated in the central business districts of Baltimore and Washington D.C., and is expected to play a dominant role in affecting ridership in the PM peak. On the other hand, a high number of households are found in both downtown Baltimore and the D.C. suburbs. While households in D.C. suburbs are expected to generate morning commuting trips mainly toward downtown D.C., those in downtown Baltimore may not immediately increase rail ridership as commuters may not necessarily find rail the most convenient mode for short commuting trips within downtown or for reverse commuting in the AM peak period.

The number of households near stations is very high in the center of Baltimore, and becomes very low toward the northern part of the light rail line (Table 4). This means that these light rail stations with relatively fewer households need to attract riders from the outside the station area, and will depend on feeder bus service and park-and-ride lots. The West Baltimore MARC station is certainly an outlier for its high number of households around a MARC station. Within WMATA Metrorail system, stations on the Red Line's west side generally have a higher number of households than those on the Green, Yellow, Orange, and Blue lines on the east side of D.C.

Figure 4 Number of Households



As expected, the number of jobs is higher in downtown Baltimore, and at several WMATA Metrorail stations on the Red Line's west side. Jobs are also concentrated in the east side of downtown Baltimore, including the Johns Hopkins Hospital station, Charles Center station, and Shot Tower station. Along light rail line, more jobs are located in the core of downtown Baltimore, including Camden Yards station, Convention Center station, University Center, Centre Street, Mount Royal, and Penn Stations. Along WMATA's Red Line, employment concentrates at several stations, including Silver Spring, Bethesda, Medical Center, and White Flint.

Figure 5 Number of Total Jobs

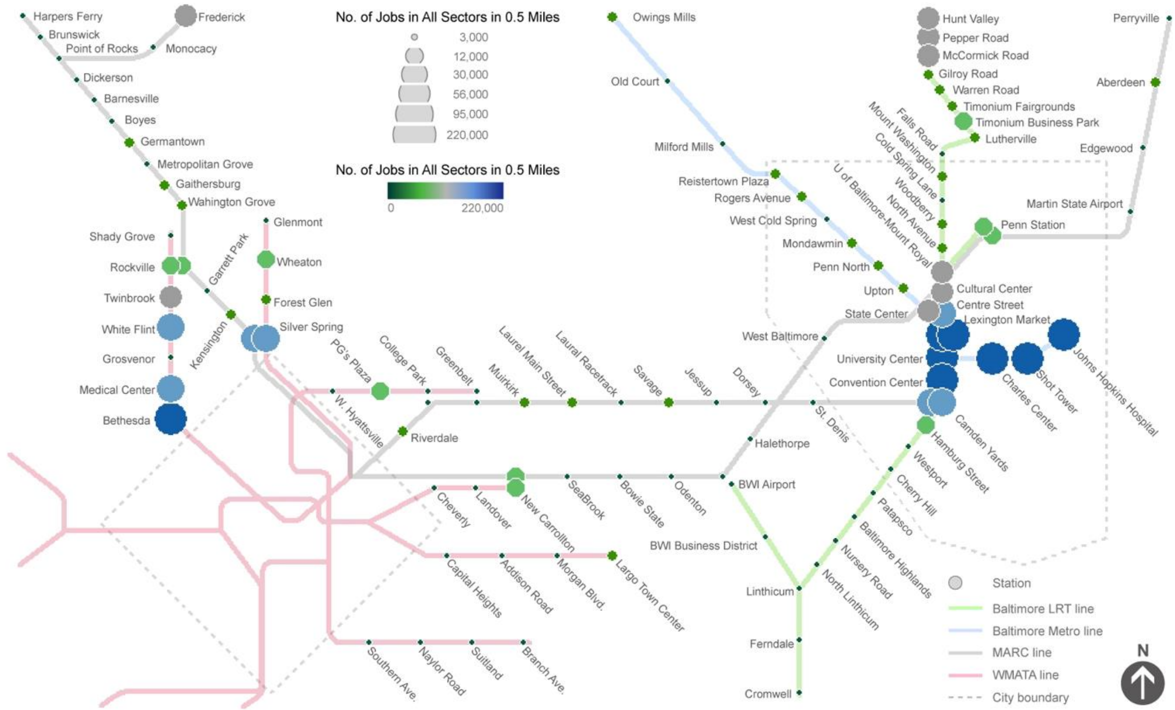


Figure 6 shows the number of train runs in the AM peak period.[7] In the AM peak period, four stations on WMATA's Red Line have the highest level of train service—Silver Spring, Bethesda, Medical Center, and Grosvenor. On Baltimore's light rail, the highest level of service is in downtown Baltimore, higher than anywhere on the two other MTA rail systems. The light rail system has three different service lines: (i) Hunt Valley-BWI (Blue), (ii) Hunt Valley-Cromwell (Yellow), and (iii) Penn Station-Camden Yards (Red). The schedule of these three lines serving downtown Baltimore explains the higher transit service between Penn Station and Camden Yards. The inner parts of the rest of light rail line have a service level equivalent to Baltimore Metro and the northern ends of WMATA's Red and the Blue Lines. The north and south ends of MTA light rail have lower levels of service, equivalent to part of the MARC Penn Line train service. As expected, the rest of MARC stations have the lowest levels of train service.

[7] Data on the number of train in the PM peak and off-peak periods are provided in Figures 10 and 11 in Appendix C.

Figure 6 Number of Trains: AM Peak

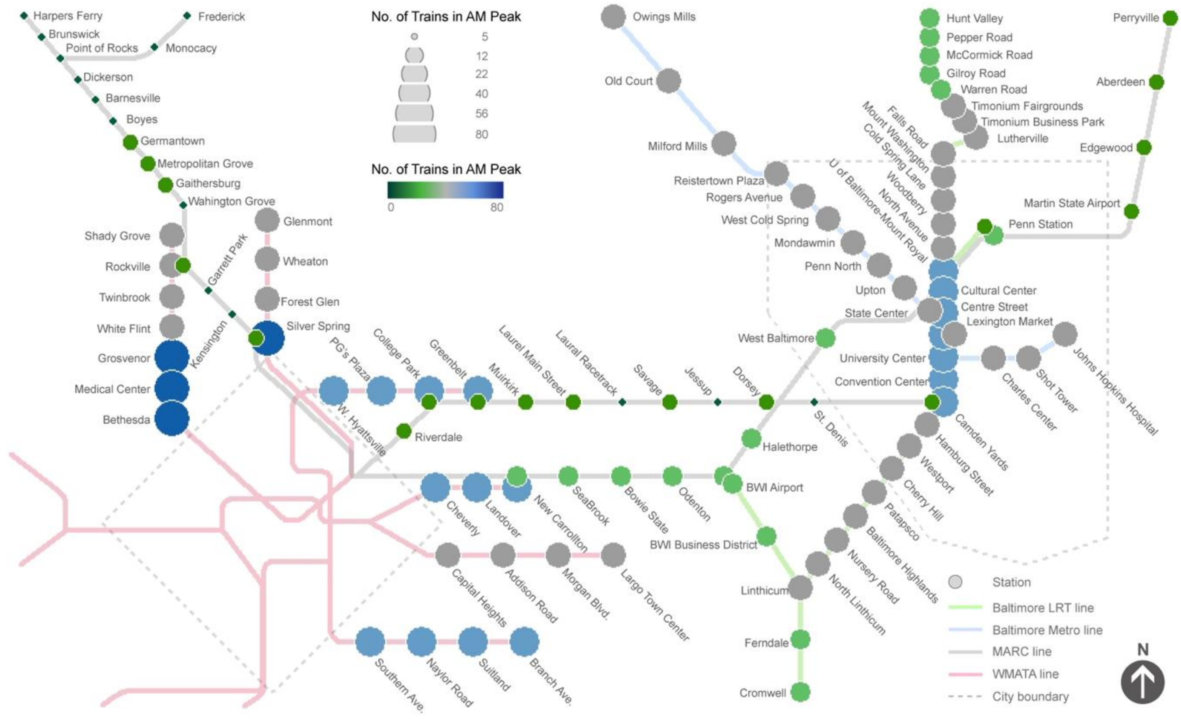
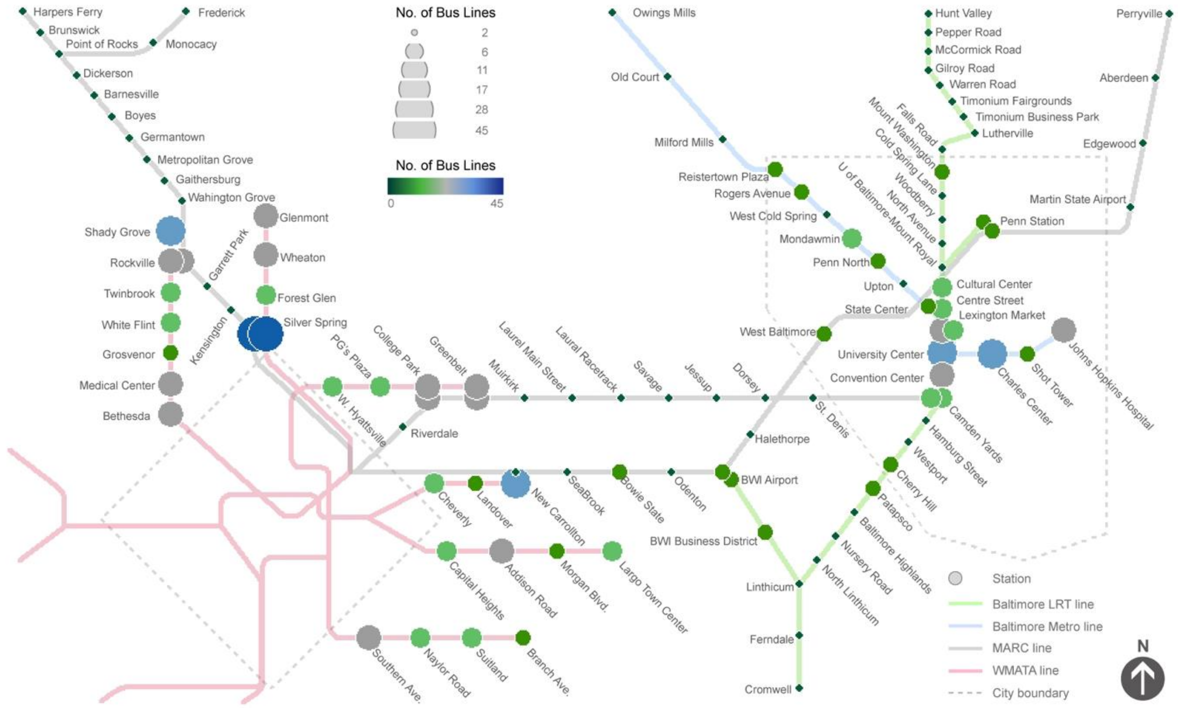


Figure 7 shows the number of bus lines that serve stations on the four rail systems. Many bus lines come together at rail stations located in the center of Baltimore—particularly at the University Center and Convention Center stations. Outside downtown Baltimore, several stations on the light rail and Metro lines have more bus lines than other stations, making them function as important transfer points. MARC stations generally have very low levels of feeder bus service throughout the system except at multimodal stations. WMATA stations generally have a higher level of feeder bus service in the D.C. region with the highest level found at the Silver Spring Red Line station, which is a major transfer hub in Montgomery County. The New Carrollton and Shady Grove stations also have higher service, extending transit service to areas outside the immediate WMATA Metrorail service area.

Figure 7 Number of Bus Lines



4.3.Model Development Results

Table 8 presents the results of the three models in the AM peak, PM peak, and Off-peak periods.[8] These models are considered the best because: (1) their higher R-squared, (2) the more parsimony, (3) estimated coefficients not easily influenced by the addition of other variables, and (4) reasonable intuition. First, the higher R-squared indicates that the model has a higher explanatory/predictive power. That is, the model with the higher R-squared can estimate and predict ridership with smaller errors. Second, although adding more variables will increase the R-squared, a model with so many independent variables does not necessary lead to better estimates of estimated coefficients, given the number of observations around 100 plus. Therefore, it is better to keep the model parsimonious. Third, drastic changes in estimated coefficients generated by adding another variable may indicate a collinearity problem, in which two or more variables are so highly correlated that it leads to biased estimation. Finally, we kept independent variables within reasonable intuition, rather than necessarily seeking the higher R-squared. This is especially the case with the station

[8] In addition to these three main models, several alternative models are shown in Appendix D.

grouping; the use of an advanced method to identify clusters based on several key attributes, such as the number of trains, the number of bus lines, and the number of households, led to the grouping of stations, some of which showed signs of coefficients that are opposite to our expectations and difficult to explain. Therefore, we generally kept the grouping by the rail system and geographic locations (e.g., within the City of Baltimore). We will discuss each of the three models below.

AM Model

The model specification for the AM peak period includes two transit service variables (number of trains and parking capacity), as well as the number of households, and three dummy variables. The R-squared is 0.861, indicating approximately 86 percent of the variance of ridership is explained by variances in the independent variables.

The number of bus lines is not significant enough to be included in the final model. As seen in Figure 7, the level of feeder bus service substantially varies among all rail systems and between downtown and suburb areas. On average, the number of buses for MARC stations is the lowest among all systems while WMATA stations have highest bus connections. For all WMATA stations in Maryland, each station has about 12 bus lines. There are also large variations of bus connections for stations in and outside downtown Baltimore for both light rail stations and Metro station. More investigation on bus connections to rail systems is needed. More particularly, we would like to investigate how rail riders use bus services to connect to rail and whether the usage varies by systems and locations.

Among all the independent variables, the number of trains has the highest explanatory power with an estimated coefficient of 0.995, which is interpreted as an increase in ridership of 9.95 percent in response to a 10 percent increase in the number of train runs in the AM period.

The number of parking spaces is significant and positive, with a coefficient of 0.47. The correlation coefficients of household and parking capacity of MARC and Metro stations are 0.405, and 0.532, respectively, which indicates that most riders using MARC and Metro trains rely heavily on park-and-ride lots, which helps explain the positive estimated coefficient of parking capacity.

As expected, the number of households also has a significant and positive coefficient (0.105), indicating that AM ridership is expected to increase by 1.05 percent in response to a 10 percent increase in the number households. The current estimated coefficient is relatively lower, compared with the results of the coefficient of households (0.197) in the previous WMATA ridership study. This suggests that the magnitude of households' effect on rail ridership is lower for the four rail systems in Maryland than for the WMATA Metrorail system.

Lastly, the inclusion of a dummy variable of light rail stations significantly improves the model fit of AM model. The coefficient is -1.635, which means a reduction of ridership by 80.5 percent for light rail stations, keeping all other conditions constant.

Table 8 Regression Model Results by Time of Day

Dependent Variable: Ln(boardings)	**AM Peak Period**			**PM Peak Period**			**Off-peak Period**		
Independent Variables	**Coef.**	**Std. Err.**	**P>|z|**	**Coef.**	**Std. Err.**	**P>|z|**	**Coef.**	**Std. Err.**	**P>|z|**
Ln(No. of trains in AM peak period)	0.995	0.087	0.000	-	-	-	-	-	-
Ln(No. of trains in PM peak period)	-	-	-	1.309	0.124	0.000	-	-	-
Ln(No. of trains in off-peak period)	-	-	-	-	-	-	0.976	0.091	0.000
Ln(No. of households in 0.5 miles)	0.105	0.036	0.004	-	-	-	-	-	-
Households in 0.5 miles (0-1)	0.998	0.650	0.128	-	-	-	-	-	-
Ln(No. of jobs in all sectors in 0.5 miles)	-	-	-	0.126	0.037	0.001	-	-	-
Ln(No. of midday & weekend jobs in 0.5 miles)	-	-	-	-	-	-	0.055	0.054	0.320
Interaction term of Midday & weekend jobs in 0.5 miles and non-MARC (0-1)	-	-	-	-	-	-	0.204	0.064	0.002
Ln(transit accessibility in 0.5 miles)	-	-	-	0.139	0.066	0.037	-	-	-
Ln(No. of buslines)	-	-	-	0.435	0.091	0.000	0.228	0.101	0.027
Bus Lines (0-1)	-	-	-	-0.423	0.208	0.045	-0.176	0.224	0.433
Ln(No. of parking spaces)	0.470	0.064	0.000	-	-	-	0.167	0.081	0.042
Parking Capacity (0-1)	2.255	0.392	0.000	-	-	-	0.878	0.488	0.075
LRT stations (0-1)	-1.635	0.152	0.000	-	-	-	-	-	-
LRT stations within the City of Baltimore (0-1)	-	-	-	-0.857	0.251	0.001	-	-	-
Metro stations (0-1)	-	-	-	0.502	0.268	0.064	1.318	0.244	0.000
WMATA stations (0-1)	-	-	-	-0.892	0.324	0.007	2.020	0.285	0.000
Brunswick & Penn lines stations of MARC (0-1)	-	-	-	-	-	-	0.952	0.255	0.000
Constant	0.178	0.389	0.648	-2.325	0.851	0.851	-0.848	0.542	0.121
Number of Samples	108			103			105		
R-squared*	0.861			0.871			0.875		

Notes: "***" indicates coefficient is statistically significant at the 0.01 level; "**" indicates coefficients are statistically significant at the 0.05 level.

PM Model

The PM model presented in Table 8 includes three transportation service variables (number of trains, number of bus lines, and parking capacity), one job related variables (number of midday and weekend jobs), and three dummy variables. The R-squared is 0.875, indicating approximately 88 percent of ridership variance is explained by the variances of the independent variables. The effect of the number of trains is more substantial in the PM period at 1.309 than in the AM peak (0.995) and off-peak (0.976) periods. This indicates that, for example, a 10 percent increase in the number of trains, on average, leads to a ridership increase of 13 percent. The positive coefficient of the number of bus lines (0.435) shows that the more bus lines, the higher the PM peak ridership. This relationship could be investigated further if the data on bus trips to and from each station were available.

The positive estimated coefficient for the total number of jobs (0.125) is expected. It is also expected this coefficient is higher than the estimated coefficient of the number of households in the AM model (0.105). Most DRMs by time of day show a more substantial effect of jobs than households on rail ridership, partly reflecting that more than households, jobs tend to concentrate in central business districts and rail station areas.

At the same time, this estimated coefficient is lower than those found in our previous study of the WMATA Metrorail system (0.358), likely indicating the overall lower numbers of jobs found near the stations in this study, compared to WMATA stations in downtown D.C.

Transit accessibility also shows a positive estimated coefficient; the higher the transit accessibility from a station, the higher the boarding counts in the PM peak period. While the total number of jobs substantially contributes to the level of transit accessibility, this variable also takes into account the synergy between pairs of stations through the rail network.

The direct interpretation of the estimated coefficients of the three dummy variables are: an increase in ridership by 65 percent for Baltimore Metro stations and a reduction in ridership by 58 percent and 59 percent for stations on the Baltimore light rail and on WMATA Metrorail respectively, compared to MARC stations and LRT stations outside the City, keeping other conditions same (*ceteris paribus*). Although these interpretations alone may be counterintuitive, some of these dummy variables were included to reduce the magnitude of overestimation, given the values for the main independent variables. In other words, given the values of the main independent variables, the model tends to overestimate the boarding counts at LRT stations within the City and at WMATA stations, which are compensated for by the inclusion of the two corresponding dummy variables.

Off-peak Model

The off-peak model presented in Table 8 includes two transit service variables (number of trains and number of bus lines), two job related variables (total number of jobs and transit accessibility), and three dummy variables. The R-squared is 0.871, indicating the 87 percent of the variance in ridership is explained by the variances of these independent variables. The effect of the number of trains is close to the unit elasticity (0.976), indicating a change in the level of ridership that is approximately proportional to a change in the number of trains; a 10 percent increase in the number of trains leads to a 9.8 percent increase in ridership. The effect of the bus service level is lower in the off-peak model (0.228) than in the PM peak model (0.435). The effect of parking capacity on the off-peak ridership is also positive but relatively low (0.167).

The estimated coefficient for the midday and weekend job at non-MARC stations is 0.204, which indicates that more of these jobs lead to higher ridership at non-MARC stations in the off-peak period.

Moreover, the effect of midday and weekend jobs needs to be combined with the effects of the two dummy variables for Baltimore Metro and WMATA Metrorail. On average, Baltimore Metro stations have higher ridership than Baltimore light rail stations and MARC stations on the Camden and Frederick lines by 273 percent in the off-peak period. Ridership at WMATA stations is 653 percent higher. MARC stations on the Brunswick and Penn lines also have, on average, ridership 159 percent higher than Baltimore light rail stations and MARC stations in the off-peak period.

Limitations

It is important to keep in mind that a DRM is based on regression analysis that estimates the average value of the dependent variable when the values of the independent variables are fixed.[9] Regression analysis also measures the magnitude and direction of change in the average value of the dependent variable when the value of one of the independent variables varies, keeping the values of the other independent variables constant. Therefore, the coefficients obtained in DRMs have a range of estimates, within which the effect of each independent variable falls with a certain level of likelihood (0.90 or 0.95). In addition, the estimated or predicted ridership based on DRMs will also have a range, as well as a certain level of error.

In the models of three time periods, error terms in DRMs tend to be larger for MARC stations and for Baltimore Light Rail stations, and smaller for Baltimore Metro stations and WMATA stations. This is due to large variances in both the dependent variable (boarding counts) and independent variables used in the models for the first two rail systems. Despite a significant amount of time and effort, it was difficult to find any independent variables to explain the relatively large variances in the dependent variable for MARC stations and for Baltimore light rail stations.

5. Concluding Remarks

The DRM model can provide useful insights on how ridership changes in response to changes in transit service, land use and built environment, and socio-demographics. In cooperation with MTA, NCSG has developed a DRM to estimate rail transit ridership at the station level for three time periods: AM peak, PM peak, and off-peak periods. These modeling efforts are built on prior DRM work with several improvements:

- it modeled detailed time of day ridership by station data

[9] In more technical terms, regression analysis estimates the *conditional expectation* of the dependent variable given the values of independent variables.

- it developed walksheds in three different distances (quarter-mile, half-mile, and one-mile) based on the most current pedestrian-oriented street network and it replaced a circular radius
- it included more recent and detailed independent variables in the model development analysis.

After data collection and processing, the model development stage involved pre-regression descriptive analysis, regression model development, and post-regression diagnostics, as well as the examination of potential station categorization.

The results suggest that transit service-related variables are the strongest predictors of ridership in all time periods. In the models of AM, PM and off-peak periods, the effect of the number of trains in PM period is more substantial than other time periods. Parking capacity has the higher coefficient in the AM model than in the off-peak model, and is insignificant in the PM peak model. This suggests that AM ridership is more dependent on parking capacity, attracting rail riders from locations outside the station's immediate area or along feeder bus lines. Feeder bus service was found significant and positive in both the PM and off-peak models, and was, surprisingly, statistically insignificant in the AM model. More detailed information on egress and access modes to rail transit are needed.

In measuring the effects of land use and built environment, employment and household are the two key determinants. As expected, the number of households is significant and positive in the AM model but not significant in the PM and off-peak models. Employment is significant in both PM and off-peak models but not significant in the AM model. But midday and weekend employment is significant and positive in the off-peak model for non-MARC stations, with an estimated coefficient even higher than total employment number in the PM model.

Transit accessibility shows a significant and positive coefficient but only in the PM model, suggesting that areas with higher job accessibility by transit can lead to higher ridership. At the same time, including this variable in the PM model with the total number of jobs, implies a more nuanced interpretation, in a narrower sense to indicate the general accessibility through the transit network *net* of the effect of jobs in the immediate vicinity.

Several dummy variables were used to capture the variations in ridership among different groups of stations determined by system and geographic location. In the AM model, adding the light rail station dummy variable improves model fit. The results suggest that light rail stations have lower ridership than other systems by 80 percent in the AM model, keeping other independent variables constant. In the PM peak model, the results suggest that an increase in ridership by 65 percent for Baltimore Metro stations and a reduction in ridership by 58 percent and 59 percent for Baltimore light rail within the city and WMATA stations, respectively, compared to MARC stations and light rail stations outside of the city. In the off-peak model, both Baltimore Metro stations and WMATA stations have ridership higher than other systems. Among MARC stations, those on the Brunswick line and Penn lines have higher ridership.

In summary, DRM can provide reasonable estimates of station-level boardings without relying on a complicated transportation demand model, such as a four-step model. DRM can capture the relationships between station-level boardings and important attributes of transit service, land use, and built environment characteristics, which can provide a basis for further analyses of operation, planning, and policy measures to increase transit ridership in a timely and cost-effective way.

It is also worth noting the limitations of the current DRM. First, regression analysis measures the magnitude and direction of change in the typical (or average) value of the dependent variable when the value of one of the independent variables varies, keeping the values of the other independent variables constant. Second, in the DRM analysis, the estimated coefficients are constant for all the stations and the analysis can’t capture variations in the relationship between ridership and explanatory variables among stations. Third, errors of predicted ridership can vary over a large range.

Potential Applications

This study’s DRM can be applied to long- and short-term ridership projections based on the changes of explanatory variables included in the model. Changes in explanatory variables could be introduced in many different ways. The levels of transit service—both rail and bus—and the capacity of park-and-ride lots are variables directly under the control of transit agencies. Land use variables, such as the numbers of households and jobs within the station walksheds, are influenced by many more factors in both the public and private sectors, including local zoning, the regional economy, real estate, labor market, and levels of public and private investment.

Higher numbers of potential transit users can be attracted at both trip origins and destinations if a station is more accessible to pedestrians. For example, a new pedestrian path over a rail track can connect a station and several apartment complexes. Likewise, the number of households can substantially increase within the walkshed leading to improved street connectivity that could lead to a ridership increase. In addition, station accessibility could be improved by better bike paths, convenient and secured bike parking stations, or a bikesharing system.

The DRM can be used to estimate ridership changes based not only on a change in one variable but also on changes in a combination of multiple variables. The model can test these changes by time of day. Ridership can also be estimated using daily ridership for a typical weekday. It should be noted that a range of estimates and the magnitude of errors tend to be larger when changing the values of multiple variables than changing the value of one variable.

Finally, the DRM developed in this study can provide better estimates with relatively small incremental changes in independent variables rather than drastic changes because of the log-linear function of the estimated models.

Acknowledgements

The authors thank Maryland Department of Transportation (MDOT) for their continued support of this study to develop the station-level DRMs. The opinions and viewpoints expressed are entirely those of the authors, and do not necessarily represent the policies and programs of any agency.

References

Cervero, Robert. *America's Suburban Centers: The Land Use-Transportation Link*. Crows Nest: Unwin Hyman, 1989.

Cervero, Robert. *Transit-oriented Development in the United States: Experiences, Challenges, and Prospects*. Vol. 102. Washington D.C.: Transportation Research Board of the National Academies, 2004.

Cervero, Robert. "Alternative approaches to modeling the travel-demand impacts of smart growth." *Journal of the American Planning Association* 72, no. 3 (2006): 285-295.

Cervero, Robert. "Transit-oriented development's ridership bonus: a product of self-selection and public policies." *Environment and Planning A* 39, no. 9 (2007): 2068-2085.

Cervero, Robert, Jin Murakami, and Mark A. Miller. "Direct ridership model of bus rapid transit in Los Angeles County." UC Berkeley Center for Future Urban Transport: A Volvo Center of Excellence (2009).

Calthorpe, Peter. *The Next American Metropolis: Ecology, Community, and the American Dream*. New York: Princeton Architectural Press, 1993.

Duduta, Nicolae. "Direct Ridership Models of Bus Rapid Transit and Metro Systems in Mexico City, Mexico." *Transportation Research Record: Journal of the Transportation Research Board* 2394 (2013): 93-99.

Fehr & Peers. "DRM (Direct Ridership Models)." Accessed Jul. 17, 2013. http://www.fehrandpeers.com/drm-direct-ridership-models/.

Frank, Lawrence D., Martin A. Andresen, and Thomas L. Schmid. "Obesity Relationships with Community Design, Physical Activity, and Time Spent in Cars." *American Journal of Preventive Medicine* 27, no. 2 (2004): 87-96.

Geurs, Karst T., and Bert Van Wee. "Accessibility Evaluation of Land-use and Transport Strategies: Review and Research Directions." *Journal of Transport Geography* 12, no. 2 (2004): 127-140.

Gutiérrez, Javier, Osvaldo Daniel Cardozo, and Juan Carlos García-Palomares. "Transit Ridership Forecasting at Station Level: An Approach Based on Distance-decay Weighted Regression." *Journal of Transport Geography* 19, no. 6 (2011): 1081-1092.

Kuby, Michael, Anthony Barranda, and Christopher Upchurch. "Factors Influencing Light-rail Station Boardings in the United States." *Transportation Research Part A: Policy and Practice* 38, no. 3 (2004): 223-247.

Longitudinal Employer-Household Dynamics. http://lehd.ces.census.gov/. 2014.

Maryland Department of Planning. "Smart Growth Planning Topics." Accessed Jul. 30, 2013. http://www.mdp.state.md.us/OurWork/smartGrowth.shtml.

Maryland Department of Transportation. *2009 Maryland Transportation Plan*. Hanover: Maryland Department of Transportation, 2009.

Mishra, Sabyasachee, Timothy F. Welch, and Manoj K. Jha. "Performance Indicators for Public Transit Connectivity in Multi-modal Transportation Networks." *Transportation Research Part A: Policy and Practice* 46, no. 7 (2012): 1066-1085.

The National Center for Smart Growth Research and Education at the University of Maryland. Maryland Scenarios Project Final Report. College Park: The National Center for Smart Growth Research and Education at the University of Maryland, 2012.

United States Census Bureau. "Longitudinal Employer-Household Dynamics." Accessed Jul. 30, 2013. http://lehd.ces.census.gov/.

Walk Score. “Most Walkable Cities in the United States, Canada, and Australia on Walk Score.” Accessed Jul. 30, 2013. http://www.walkscore.com/cities-and-neighborhoods/.

Appendix A. Technical Note on Baltimore Light Rail Ridership Estimation

Overview

To develop a station level Direct Ridership Model (DRM) for all of Maryland's rail stations, ridership by station is a critical variable; it serves as the dependent variable for model development. It is also important to collect information on ridership by time of day to reflect the interrelationships of ridership and land use that vary by time of day by station.

Currently, MTA's Light Rail system is collecting fares using daily passes with magnetic strips, which makes it challenging to gather light rail ridership by station and by time of day. The only reliable data source is the National Transit Database Survey FY 2015-2016 provided by MTA. All the survey data are in hard copies and will need substantial efforts on data entry, organization, manipulation, and validation.

Train run survey samples were collected through random sampling techniques and includes: train run dates and times, directions, boardings, alightings, start and end stations, and full-lines or shuttle lines. This document describes data sources and key steps implemented to get light rail ridership estimation by station and by time of day. Estimated ridership is further validated by other ridership data and by consulting MTA staff members.

Data Sources

- National Transit Database Survey FY2015-2016 (in hard copy

A collection of hardcopy records of samples collected by randomly selected train trips every third weekday of FY2015-16. It surveys the selected trip and the next three trips on the block for the sample pool. So if trip 3 is randomly selected, trips 3, 4, 5, and 6 will be surveyed. Every Saturday, Sunday, and holiday, 2 trips were randomly selected. The selected trip and the next trip on the block go into the pool. So if trip 1 and 13 are randomly selected, trips 1, 2, 13, and 14 will be surveyed. On the through line, randomly selected cars were checked (Car 1 or 2, or Car 1, 2, or 3 if a 3 car train). Penn-Camden shuttles are one car.

The total number of samples is 1,171. Each sample contains information on survey date (holidays or not[10]), day, line type (full line or shuttle line), departure time, direction (northbound or southbound), block number as well as car number, boardings and alightings by station, load number (actual number of riders on the car), miles travelled, and accumulated passenger miles.

- MTA Light Rail Operation Data FY2015-16

The MTA data includes information on operation miles, time, and trips for train and for car by each time period of weekdays and weekend (i.e., AM peak, Midday, PM peak, Other). The average number of cars per train of full-line train runs was derived from this dataset.

- MTA Light Rail Service Schedule/Timetable

Data was obtained from
https://mta.maryland.gov/schedules/display.php?route=200_light_rail_weekday_northbound.xls

Particularly the number of train runs by station by time of day of weekday and weekends and by full-line and shuttle line.

- MTA Light Rail Average Weekday Boarding Counts FY2015-16

MTA data and includes information on average weekday boarding by station and by month.

- WBA On Board Survey

Research company WBA conducted an on board survey under a contract with MTA. Based on the survey results, WBA extracted ridership by three time periods: AM peak, PM peak, and other, which was used as the total control to validate time of day variation for estimated ridership.

- Baltimore Metro Ridership data

Baltimore Metro ridership by time of day was also used as the total control to adjust the time of day variation for estimated ridership.

[10] Orioles game dates were used to see if the train trip samples were affected by special events. Only two games were scheduled on MTA light rail sample weekdays.

Methodology

The following equations are used to denote estimated ridership in the analysis process:

$$X_{ikd} = (f_{ikd} / m_{ikd}) * R_i * Q_{ikd}$$

$$Y_{ikd} = (s_{ikd} / n_{ikd}) * T_i * P_{ikd}$$

$$Z_{ik} = X_{ik1} + X_{ik2} + Y_{ik1} + Y_{ik2}$$

$$Z_k = \Sigma Z_{ik} \quad (\text{for } i = 1 \sim 4)$$

$$Z_i = \Sigma Z_{ik} \quad (\text{for } k = 1 \sim 33)$$

where

i=1, 2, 3, 4 time of day (am peak, midday, pm peak, other)

k=1, 2,…, 33 station

d=1, 2 direction (northbound, southbound)

F, indicating full-line samples

S, indicating shuttle line samples

f, indicating the total sample boarding count of the full line

s, indicating the total sample boarding count of the shuttle line

m, indicating the total sample number of train runs on the full line

n, indicating the total sample number of the train runs on the shuttle line

R, indicating the average car number per train on the full line

T, indicating the average car number per train on the shuttle line

Q, indicating the number of train runs on the full line within a weekday

P, indicating the number of train runs on the shuttle line within a weekday

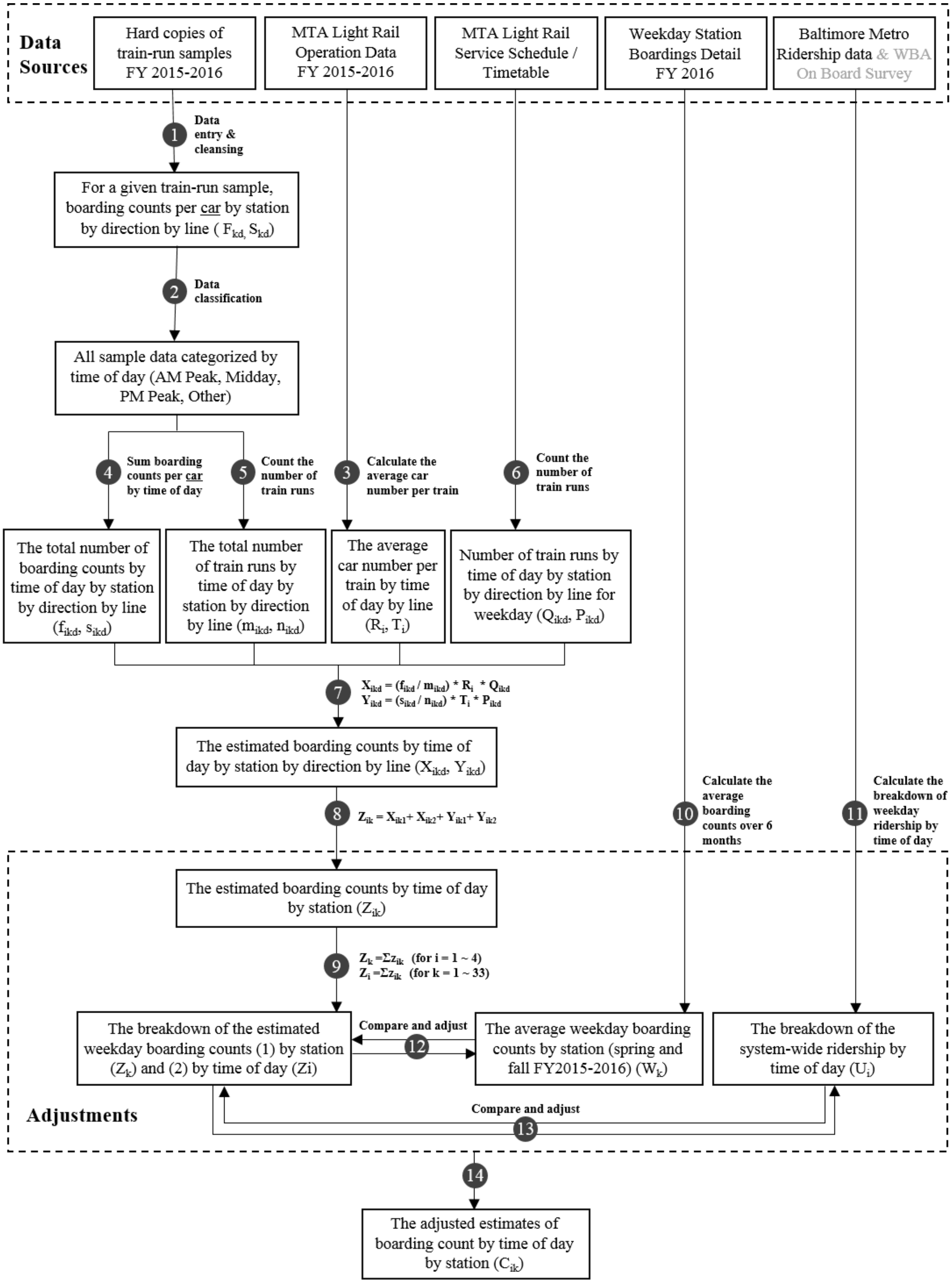
X, indicating the estimated boarding number on the full line

Y, indicating the estimated boarding number on the shuttle line

Z, indicating the estimated total boarding count combining the full line and short line

C, indicating the adjusted estimates of boarding count

Figure 8 Flow chart for estimating Baltimore Light Rail ridership



- **Step 1:** The total number of samples is 1,171. Due to time and personnel constraints, only train run samples of spring (March, April, and May) and fall (September, October, and November) were manually entered. The total number of trip samples in spring and fall is 344. Information entered includes: boarding by station, survey date and time, direction, full-line or shuttle line. The sample data were furthered organized into four datasheets: full line northbound boarding data, full line southbound boarding data, short line northbound boarding data, and short line southbound boarding data. Therefore, **boarding counts per car by station by direction by line for each of train-run samples (F_{kd}, S_{kd})** were obtained.
- **Step 2:** Based on Step 1, all sample data were categorized by time of day for a weekday, i.e., AM peak (6:30 a.m. to 8:29 a.m.), midday (9:30 a.m. to 2:59 p.m.), PM peak (3:00 p.m. to 6:29 p.m.), and other (5:00 a.m. to 6:29 a.m. and 6:30 p.m. to 12:00 a.m.), then **boarding counts per car by time of day by station by direction by line (F_{ikd}, S_{ikd})** were classified and grouped.
- **Step 3:** Train runs of full line and shuttle line are operated using different numbers of cars. Most of the full line train runs use one, two or three cars, while all shuttle line trains are operated using one-car trains. Based on *MTA Light Rail Operation Data FY 2015-2016*, **the average car number per train by time of day for the full line (R_i) and shuttle line (T_i)** was calculated. Here T_i always equals 1.
- Step 4: Based on Step 2, the total number of boarding counts by time of day by station by direction for the full line (f_{ikd}) and the shuttle line (s_{ikd}) was calculated through aggregation.
- Step 5: Based on Step 2, the total number of train runs by time of day by station by direction for the full line (m_{ikd}) and the shuttle line (n_{ikd}) was also counted.
- Step 6: The number of train runs by time of day by station by direction for the full line (Q_{ikd}) and the shuttle line (P_{ikd}) for weekday was obtained from the *MTA Light Rail Service Schedule / Timetable*.

- Step 7: Based on Steps 3, 4, 5, and 6, the estimated boarding counts by time of day by station by direction for the full line (X_{ikd}) and the shuttle line (Y_{ikd}) were calculated using the following equations:

$$X_{ikd} = (f_{ikd} / m_{ikd}) * R_i * Q_{ikd}$$
$$Y_{ikd} = (s_{ikd} / n_{ikd}) * T_i * P_{ikd}$$

- Step 8: Based on Step 7, the estimated boarding counts by time of day by station (Z_{ik}) were calculated using the following equation:

$$Z_{ik} = X_{ik1}+ X_{ik2}+ Y_{ik1}+ Y_{ik2}$$

- Step 9: Based on Step 8, the breakdown of the estimated weekday boarding counts by station (Z_k) and by time of day (Z_i) was calculated using the following equations:

$$Z_k = \Sigma z_{ik} \text{ (for i = 1 ~ 4)}$$
$$Z_i = \Sigma z_{ik} \text{ (for k = 1 ~ 33)}$$

- Step 10: Using the number of samples from *Weekday Station Boardings Detail FY 2016*, the average weekday boarding counts by station over spring (March, April, and May) and fall (September, October, and November) in FY16 (W_k) were calculated. The outcome of this calculation was used as one of the control totals to compare and adjust ridership by station.

- **Step 11:** Using the number of samples from *Baltimore Metro Ridership data*, **the breakdown of the system-wide ridership by time of day (U_i)** was calculated. The outcome of this calculation was used as a control total to compare and adjust ridership by time of day.

- Step 12: Based on Steps 9 and 10, the estimated boarding counts by station ($_k$) were compared with the average weekday boarding counts by station (W_k) and were adjusted accordingly.

- Step 13: Based on Steps 9 and 11, the estimated boarding counts by time of day (Z_i) were compared with the breakdown of the system-wide ridership by time of day (U_i) and were adjusted accordingly.

- Step 14: Based on Steps 12 and 13, the adjusted estimates of boarding count by time of day by station (C_{ik}) were obtained.

Appendix B. Zero Ridership Station after Adjustments

The presence of stations with no ridership poses a mathematical problem in transformation of data using the logarithmic function. Therefore, after discussion with MTA, we first combined the midday and other time periods into the "off-peak" time period, and second, allocated MARC trains in a way that we can reduce the number of zero-ridership stations in all of the three time periods. Table 9 shows stations with zero ridership after these adjustments; these stations are not included in the model development in each of the three time periods.

Table 9 MARC Stations with Zero Ridership by Time of Day

No.	MARC station	AM	PM Adj.	Midday Adj.	Other Adj.
1	Jessup	○	○	○	○
2	Harpers Ferry		○	○	○
3	Boyds		○	○	
4	Dickerson		○	○	
5	Barnesville		○	○	
6	Brunswick		○	○	
7	Garrett Park		○		○
8	Frederick		○		○
9	Monocacy		○		○
10	Washington Grove			○	
11	Point of Rocks			○	
12	Riverdale			○	
13	Laurel Racetrack	○			○
14	Edgewood				○
15	St Denis				○
16	Perryville				○
17	Dorsey				○
18	Baltimore-Camden				○

Table 100 Variables Collected and Considered but Not Included in the Final Models

Variables
Trains per Hour in AM Peak Period
Trains per Hour in PM Peak Period
Trains per Hour in Off-peak Period
Number of jobs for workers with Educational Attainment: Less than high school
Number of jobs for workers with Educational Attainment: High school or equivalent, no college
Number of jobs for workers with Educational Attainment: Some college or Associate degree
Number of jobs for workers with Bachelor's degree or advanced degree
Number of jobs with earnings $1250/month or less
Number of jobs with earnings $1251/month to $3333/month
Number of jobs with earnings greater than $3333/month
Number of jobs in NAICS sector 11 (Agriculture, Forestry, Fishing and Hunting)
Number of jobs in NAICS sector 21 (Mining, Quarrying, and Oil and Gas Extraction)
Number of jobs in NAICS sector 22 (Utilities)
Number of jobs in NAICS sector 23 (Construction)
Number of jobs in NAICS sector 31-33 (Manufacturing)
Number of jobs in NAICS sector 42 (Wholesale Trade)
Number of jobs in NAICS sector 44-45 (Retail Trade)
Number of jobs in NAICS sector 48-49 (Transportation and Warehousing)
Number of jobs in NAICS sector 51 (Information)
Number of jobs in NAICS sector 52 (Finance and Insurance)
Number of jobs in NAICS sector 53 (Real Estate and Rental and Leasing)
Number of jobs in NAICS sector 54 (Professional, Scientific, and Technical Services)
Number of jobs in NAICS sector 55 (Management of Companies and Enterprises)
Number of jobs in NAICS sector 56 (Administrative and Support and Waste Management and Remediation Services)
Number of jobs in NAICS sector 61 (Educational Services)
Number of jobs in NAICS sector 62 (Health Care and Social Assistance)
Number of jobs in NAICS sector 71 (Arts, Entertainment, and Recreation)
Number of jobs in NAICS sector 72 (Accommodation and Food Services)
Number of jobs in NAICS sector 81 (Other Services [except Public Administration])
Number of population
Number of population age 10-17
Number of population age 18-21
Number of population age over 65
Number of vacant households
Number of households below poverty level
Number of population in labor force
Number of population employed
Number of population unemployed
Number of renter occupied housing units
Number of owner occupied housing units
Market gross rent

Appendix C. Additional Maps

Figure 9 Ridership by Time of day: Off-peak

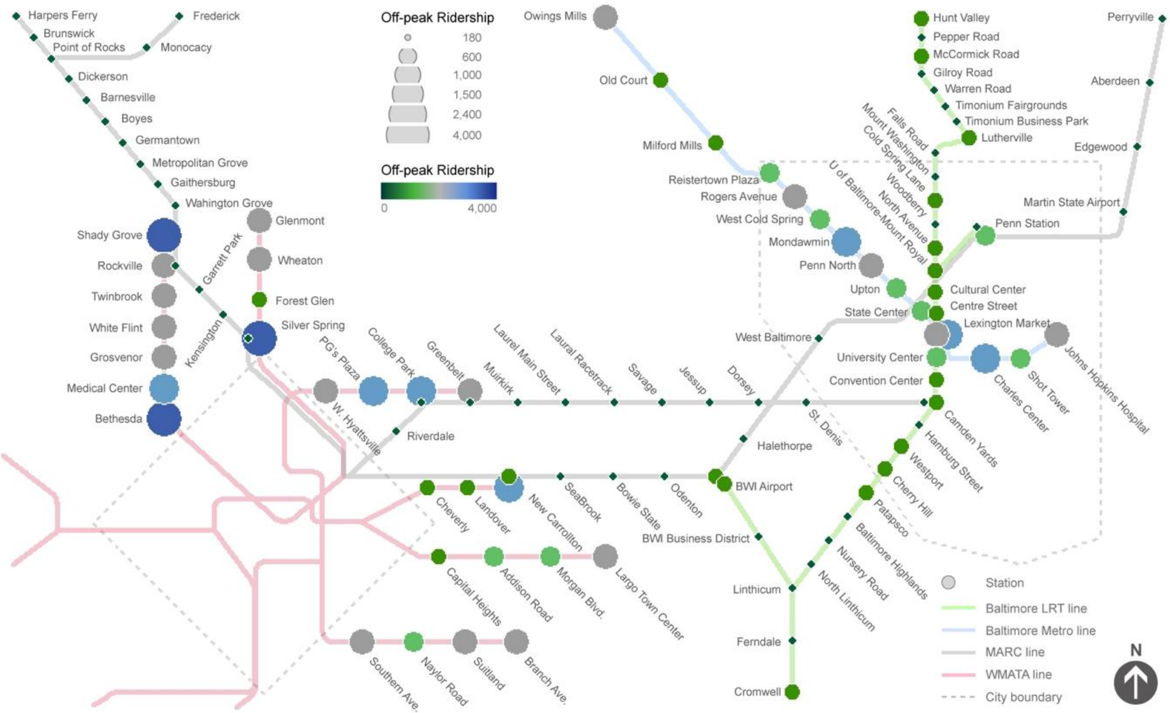


Figure 10 Number of Trains: PM Peak

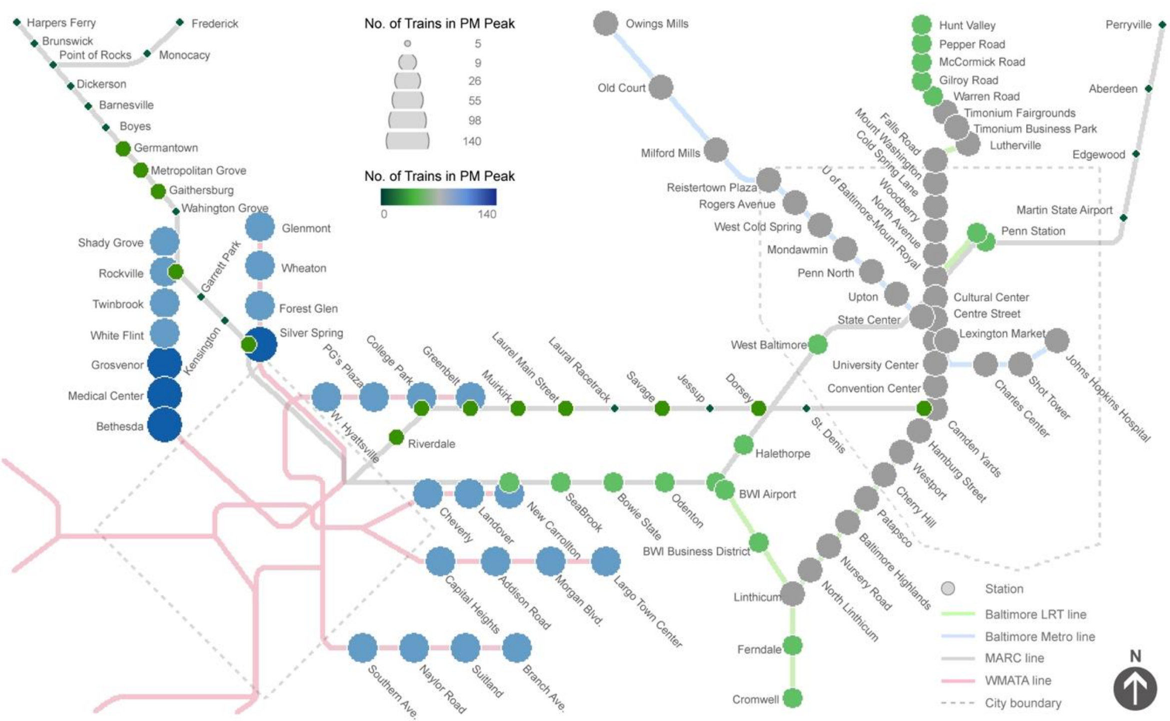


Figure 11 Number of Trains: Off-peak

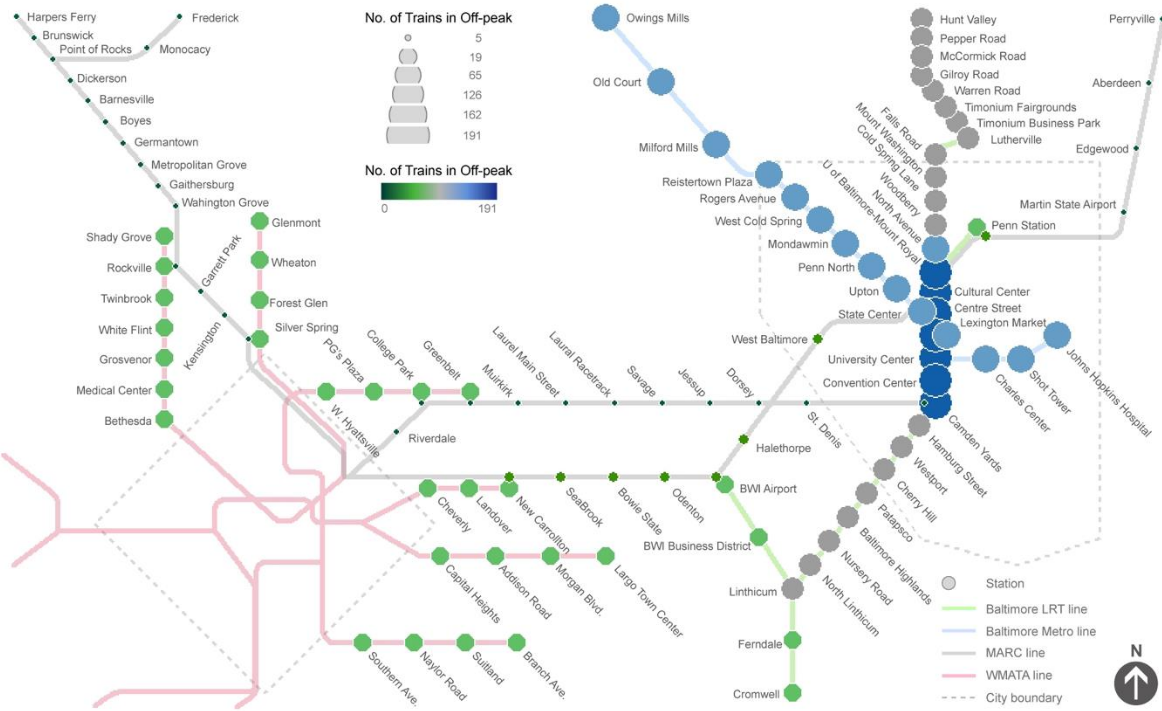


Figure 12 Parking Capacity

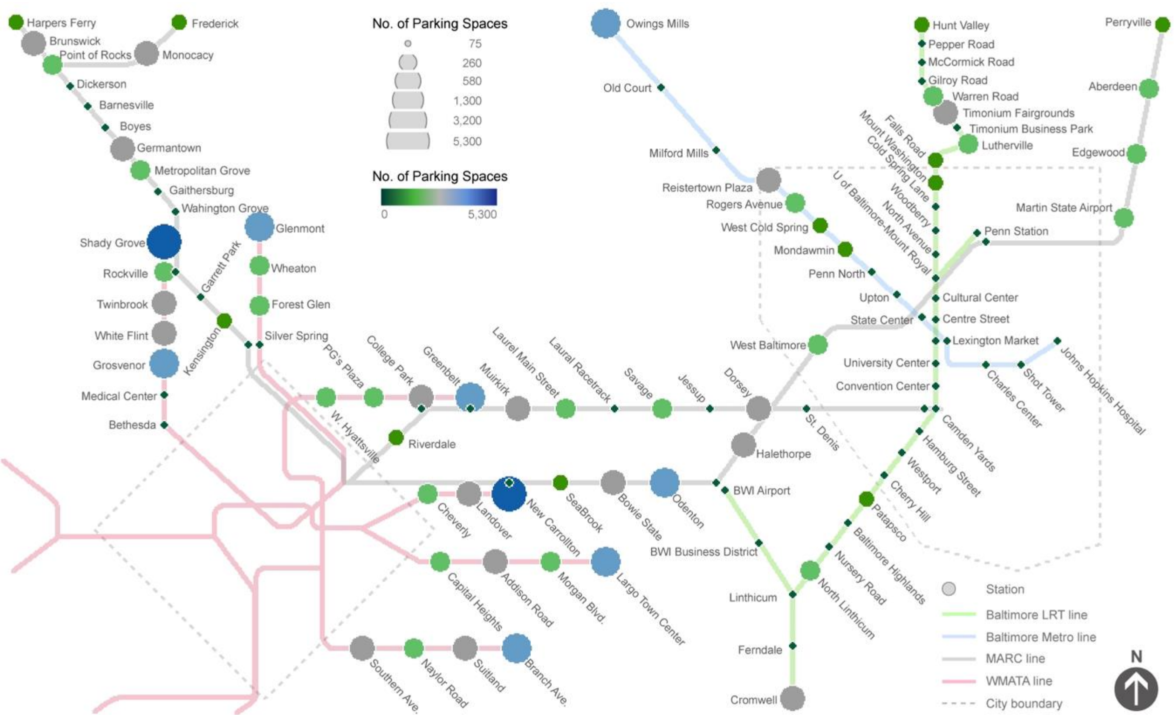
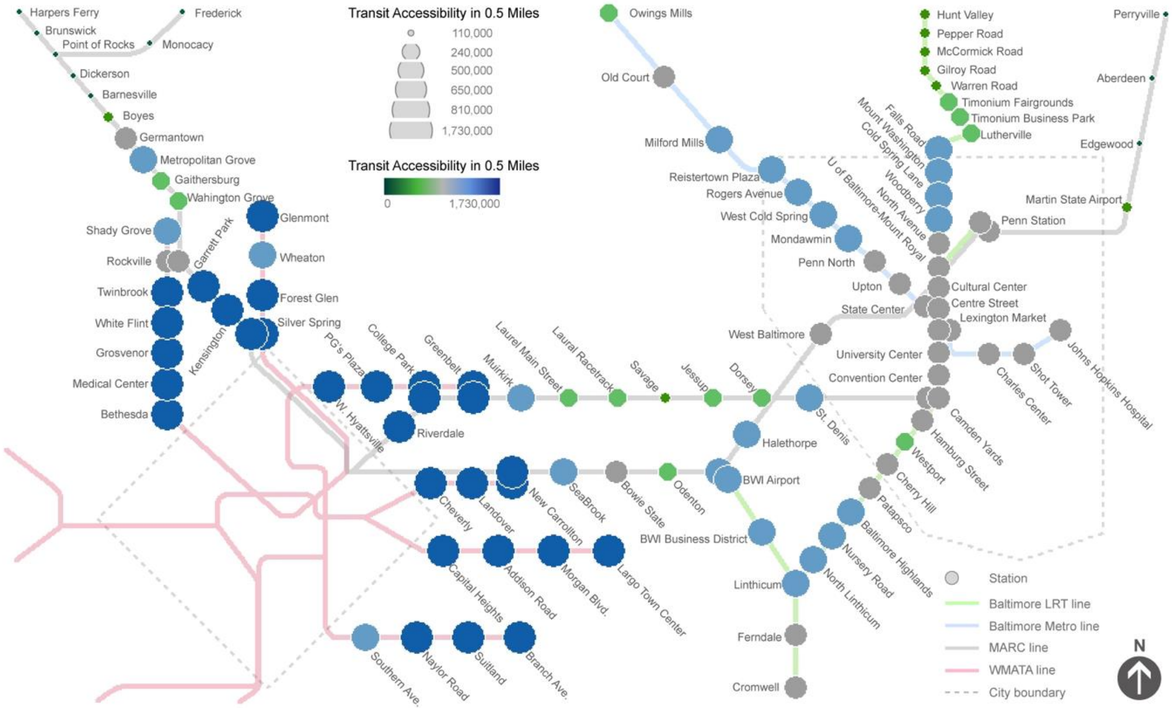


Figure 13 Transit Accessibility



Appendix D. Alternative Models

Dependent Variable: Ln(boardings)	**AM Peak Period**			**PM Peak Period**			**Off-peak Period**		
Independent Variables	**Coef.**	**Std. Err.**	**P>|z|**	**Coef.**	**Std. Err.**	**P>|z|**	**Coef.**	**Std. Err.**	**P>|z|**
Ln(No. of trains in AM peak period)	0.445	0.122	0.000	-	-	-	-	-	-
Ln(No. of trains in PM peak period)	-	-	-	1.310	0.124	0.000	-	-	-
Ln(No. of trains in off-peak period)	-	-	-	-	-	-	1.019	0.089	0.000
Ln(No. of households in 0.5 miles)	0.169	0.048	0.001	-	-	-	-	-	-
Households in 0.5 miles (0-1)	0.993	0.686	0.155	-	-	-	-	-	-
Ln(No. of jobs in all sectors in 0.5 miles)	-	-	-	-	-	-	-	-	-
Ln(No. of jobs in Information sector in 0.5 miles)	-	-	-	0.097	0.029	0.001	-	-	-
Ln(No. of jobs in Public Administration sector in 0.5 miles) ┤	-	-	-	-	-	-	0.087	0.032	0.008
Ln(No. of midday & weekend jobs in 0.5 miles)	-	-	-	-	-	-	-0.027	0.066	0.685
Interaction term of Midday & weekend jobs in 0.5 miles and non-MARC (0-1)	-	-	-	-	-	-	0.258	0.064	0.000
Ln(transit accessibility in 0.5 miles)	-	-	-	0.110	0.071	0.125	-	-	-
Ln(No. of buslines)	-	-	-	0.459	0.090	0.000	0.166	0.099	0.100
Bus Lines (0-1)	-	-	-	-0.386	0.210	0.069	-0.237	0.221	0.285
Ln(No. of parking spaces)	0.702	0.093	0.000	-	-	-	0.216	0.079	0.008
Parking Capacity (0-1)	3.306	0.541	0.000	-	-	-	1.051	0.475	0.030
Interaction term of Ln(No. of parking spaces) and WMATA (0-1)	-0.154	0.079	0.054	-	-	-	-	-	-
LRT stations (0-1)	-	-	-	-	-	-	-	-	-
LRT stations within the City of Baltimore (0-1)	-	-	-	-0.858	0.253	0.001	-	-	-
Metro stations (0-1)	-	-	-	0.508	0.266	0.059	1.268	0.236	0.000
WMATA stations (0-1)	1.688	0.498	0.001	-0.848	0.323	0.010	2.058	0.275	0.000
Brunswick & Penn lines stations of MARC (0-1)	-	-	-	-	-	-	0.823	0.249	0.001
Constant	-0.505	0.608	0.408	-1.273	0.877	0.150	-1.190	0.538	0.029
Number of Samples		110			103			105	
R-squared*		0.731			0.874			0.888	

Note: ┤ indicates that in the off-peak model, the coefficient of jobs in Public Administration section is positive and significant. This may suggests that, in the off-peak period, many riders use rail services to go to government and public agency buildings, rather than commuting.

Appendix E. Land Use Mix Index and Accessibility Calculation

Land use mix index – This study considers three land-use types—residential, commercial and industrial. A land-use mix index is used to capture how evenly the land use square footage and floor area are distributed within station zones (quarter-mile, half-mile, and one-mile).The land use mix index is calculated as follows:

$$\text{Land use mix} = \quad ((-1)/\ln n) * \sum_{i=1}^{n} p_i \ln p_i \qquad (1)$$

where p_i is the percentage of land use type i of the total land area and n is the total number of different land use types. The land use mix ranges from 0 (homogeneous land use, such as in rural areas or suburban subdivisions) to 1 (most mixed, such as diverse city centers)(Frank, Martin, and Schmid, 2002.). Land use data were originally acquired from the 2010 Maryland Property View data set, which are point-based data that include X,Y coordinates of properties, land acres, and land use types including residential, commercial, and office of each property.

Accessibility – The accessibility measure for zone *i* in a region with *n* TAZs *(i* = 1, 2, ...,*n), A_i*,is represented as a function of the number of opportunities in zone *j (* j = 1, 2, ..., n*)* and impedance function between zones *i* and *j* as follows:

$$Ai = \Sigma jOjf(Cij) \qquad (2)$$

where

A_i accessibility for TAZ *i*;

O_j number of relevant opportunities in TAZ *j*;

C_{ij} travel time or monetary cost for a trip from TAZ *i* to TAZ *j*;

$f(C_{ij})$ is the impedance function measuring the spatial separation between TAZ *i* and TAZ *j*;

The impedance function, $f(C_{ij})$, is an indicator of the difficulty of travel between TAZ i and TAZ j. A commonly used mathematical formula of the impedance function f(Cij) is based on the theoretical work of Wilson (1971), and is expressed as $f(C_{ij}) = \exp(-\beta C_{ij})$, where β is an empirically calibrated parameter. Employment data that were used to represent the opportunities in TAZj in calculating accessibility were obtained from LEHD 2014.