How to Increase Rail Ridership in Maryland? Direct Ridership Models (DRM) for Policy Guidance

By

Chao Liu*
Faculty Research Associate
National Center for Smart Growth Research and Education
1226D Preinkert Field House
University of Maryland, College Park,
MD 20742
P (301) 405 9519, F (301) 314 5639
Email: cliu8@umd.edu
(*Corresponding Author)

Ting Ma
Ph.D. Research Assistant
National Center for Smart Growth Research and Education
1226 Preinkert Field House
University of Maryland, College Park,
MD 20742
P (301) 405 1112, F (301) 314 5639
Email: tingma@umd.edu

Sevgi Erdogan
Faculty Research Associate
National Center for Smart Growth Research and Education
1112J Preinkert Field House
University of Maryland, College Park,
MD 20742
P (301) 405 9877, F (301) 314 5639
Email: serdogan@umd.edu

Frederick W. Ducca
Senior Research Scientist
National Center for Smart Growth Research and Education
1112N Preinkert Field House
University of Maryland, College Park,
MD 20742
P (301) 405 91945, F (301) 314 5639
Email: fducca@umd.edu

Word Count: 5,606
Number of Tables: 5, Number of Figures: 1
Total Count: = 5,606+ (6x 250) = 7,100

Date Submitted: August 1, 2013

Submitted to: Public Transportation Planning and Development (AP025)
Submitted for Presentation at the 93rd Annual Meeting of the Transportation Research Board and Publication in Transportation Research Record
How to Increase Rail Ridership in Maryland? Direct Ridership Models (DRM) for Policy Guidance

ABSTRACT

The state of Maryland aims to double its transit ridership by the end of 2020. The Maryland Statewide Transportation Model (MSTM) has been used to analyze different policy options at a system-wide level. Direct ridership models (DRM) estimate ridership as a function of station environment and transit service features rather than using mode-choice results from large-scale traditional models. They have been particularly favored for estimating the benefits of smart growth policies such as Transit Oriented Development (TOD) on transit ridership and can be used as complementary to the traditional four-step models for analyzing smart growth scenarios at a local level and can provide valuable information that a system level analysis cannot provide. In this study, we developed DRMs of rail transit stations, namely light rail, commuter rail, Baltimore metro, and Washington D.C. metro for the state of Maryland. Data for 117 rail stations were gathered from a variety of sources and categorized by transit service characteristics, station built environment features and social-demographic variables. The results suggest that impacts of built environment show differences for light rail and commuter rail. For light rail stations, employment at half-mile buffer areas, service level, feeder bus connectivity, station distance to the CBD, distance to the nearest station, and terminal stations are significant factors affecting ridership. For commuter rail stations only feeder bus connection is found to be significant. The policy implications of the results are discussed.

Key Words: Direct Ridership Models (DRMs), rail transit ridership, Maryland
INTRODUCTION

Transportation is a critical issue for Maryland, both as the foundation for the state’s economy and for meeting the travel needs of Maryland residents. Maryland’s transportation system features extensive intra-urban travel within two major metropolitan areas, Washington D.C., and Baltimore, as well as inter-urban travel that can traverse the Appalachian Mountains and the Chesapeake Bay. The State expects significant impacts on its transportation system through changes in population, economy, and environment over time (1). By 2030, the population of the state is expected to increase to 6.7 million (about a million increase) increasing the pressure on the transportation system that is already experiencing congestion. The State prepares itself to tackle these challenges with a range of policy options such as Smart Growth, travel demand management, strategies that target reducing demand for transportation, and operational strategies that target use of existing system efficiently (1).

Maryland is one of the pioneers of progressive land use policies. These policies intend to serve preventing further sprawl in the State as well as its negative impacts on the transportation system. Now known as Smart Growth policies, these land use policies were initiated in 1992 with the Economic Growth, Resource Protection and Planning Act, a policy that outlined seven (later eight) visions for the future growth in Maryland (2).

Later in 1997, the Smart Growth and Neighborhood Conservation initiative was launched to use state funds as incentives to direct growth. Since then, the State has adopted a variety of Smart Growth laws and policies namely Priority Funding Areas Act of 1997, legislations of Planning in 2006 and Sustainable Communities in 2010 and finally the Sustainable Growth & Agricultural Preservation Act of 2012 (2).

The most recent initiative added to Maryland’s smart growth agenda is Smart, Green and Growing Initiative, a multiagency, Statewide initiative that aims to achieve a more sustainable future for Maryland through community revitalization, transportation improvements, economic development, Smart Growth, and environmental restoration efforts (1). As part of this initiative and as a transportation policy, the state establishes Transit-Oriented Development (TOD) to strengthen coordination between land use and transportation planning. By encouraging development around existing and planned transit stations, the State aims to maximize the value for its investment in transit, reduce congestion, reduce greenhouse gas emissions and pollution, and provide an alternative to sprawl (1). The State aims to double its transit ridership by the end of 2020, while reducing its GHGs by 25%. Among other strategies to achieve these goals, TOD has a significant role. This is partially due to the high potential of Maryland for TOD application with over 75 rail, light rail, and metro stations, and many more planned ones in the next 20 years. However, challenges exist such as higher upfront infrastructure costs, necessity to use public land and complex community related issues (1). Therefore, identifying the factors that influence the success of the TOD program is a critical issue for Maryland and other states.

The State utilizes the Maryland Statewide Transportation Model (MSTM), a three-tier integrated land-use and transportation model, to analyze different policy options including TOD at a system-wide level (3). However, the state agencies also need more detailed, local level policy guidance to meet this goal. In this study, we developed DRMs of rail transit stations, namely light rail, commuter rail, Baltimore metro, and Washington D.C. metro for the state of Maryland. These models are developed to complement MSTM analysis results regarding rail transit ridership. Data for 117 rail stations were gathered from a variety of sources including Maryland Transit Administration (MTA), U.S. Census, and National...
Center for Smart Growth (NCSG) data inventory and categorized by transit service characteristics, station built environment features and social-demographic variables. The results suggest that impacts of built environment show differences for light rail and commuter rail. For light rail stations, employment at half-mile buffer areas, service level, feeder bus connectivity, station distance to the CBD, and terminal stations are significant factors affecting ridership. For commuter rail stations only feeder bus connection is found to be significant, which suggests that commuter rail behave differently from light rail. The policy implications of the results and challenges of incorporating DRMs with traditional demand models are discussed.

BACKGROUND

Direct ridership models (DRM) have emerged in the US as a low cost, quick alternative to traditional four-step travel demand models to forecast transit ridership at the station or corridor level (4–6). DRMs estimate transit ridership based on multiple regression analysis of built environment characteristics of station areas, transit features, such as transit service and station facilities, and socio-demographic characteristics of riders. DRMs can be used complementary to the conventional four-step travel demand models.

The four-step models typically operate on a regional scale, predicting travel behavior at an aggregate level, and are useful in guiding highway and transit network capital investment priorities but not able to estimate the travel impacts of neighborhood-level development (7, 8). This presents an issue for regional travel demand models given the relatively low transit usage in the U.S. that even minor model imprecisions can cause significant changes in the location-specific ridership estimates yielding to unreliable transit share forecasts (9). The four-step models also cannot adequately reflect the built environment’s impact on boosting transit ridership, a point that was consistently found in a large quantity of empirical studies (5, 7, 8). Using the four-step models for transit ridership forecast also faces institutional and financial barriers as developing and maintaining them require staff resources and interagency or consultant involvement (10).

DRMs respond to changes in built environment and transit service in the immediate station area (7, 9). Utilizing multivariate regression analysis based on empirical data, DRMs quantify how land use factors at the local level influence transit ridership in a direct and immediate way. Thus, it is a low-cost prediction approach compared to complex travel demand models because it only requires data input associated with immediate station areas. These areas are called catchment areas, large enough geographic buffers to capture neighborhood attributes (5). The catchment size for built environment measurement is flexible in the literature. For example, in analyzing the BART system in the San Francisco Bay Area, it was defined as the contiguous area that historically captured 90% of all access trips to and egress trips from the station. An application in Charlotte defined the catchment size as the distance to the nearest adjacent station (7, 8). Comparing different catchment sizes using a sample of about 1500 stations, Guerra et al. (2011) found that station catchment size is not a significant factor in predicting ridership (6). They suggest that the use of easiest or readily available data related to catchment area would be sufficient when estimating DRMs.

DRMs are built upon understanding of factors at the end points of possible trips that contribute to transit ridership. Literature suggests three categories of independent variables in their DRM model; built environment characteristics (of neighborhood), socioeconomic factors, and transit/station attributes (5, 6, 11).
Built Environment – Population density and employment density have been included in almost all DRM empirical studies as key built environment factors affecting ridership (7, 9). As research on built environment’s effect on travel behavior evolved, new findings suggested additional variables reflecting smart growth to be added in, such as distance to central business district (CBD) and mixed land use index (7, 12). Kuby et al. (2003) found that the number of employment, population and renters within walking distance are significantly associated with the average weekday boarding of light rail stations. However, CBD dummy variable was not significant in their study (11). Land use mix was found positively associated with transit ridership in Gutierrez et al. (2011) and Filion (2001) (6, 13). Chu (2004) found that the number of jobs and pedestrian environment positively associated with weekday total boardings (14).

Socioeconomic Features – Kuby et al. (2003) found the percentage of households who rent, instead of own, is significantly associated with light rail ridership, that one percent increase in the number of renters will generate 6.24 daily boardings (11). Gutierrez et al. (2011) found that the number of foreign population, the number of workers, and specifically, the number of jobs in commercial and educational sectors, are associated with transit monthly boardings in Madrid Metro system (6). Chu (2004) found that the share of households without car, the percentage of the work force, the share of female and the share of Hispanic population are positively associated with weekday total boardings in Jacksonville, Florida, while the median household income, the percentage of the youth (under 18) and the share of white population are negatively associated (14).

Transit Service/Station Attributes – Transit service variables include frequency of service, operating speeds, as well as the number and frequency of feeder bus lines (5, 7, 8). Cervero (2006) found in a Charlotte case study that the inclusion of service level greatly improved model accuracy. Using a DRM to predict Bus Rapid Transit ridership in Los Angeles County, Cervero et al. (2009) found that service frequency strongly influences BRT ridership estimates, as well as high intermodal connectivity (both feeder bus routes and rail-transit services). Kuby et al. (2004) found that bus connections and park and ride spaces are significantly associated with the average weekday light-rail ridership, using data from 268 light-rail station in the U.S. (11). Station attributes including bus shelter, bus benches, park and ride lot capacity, and availability of information systems, have been found significant in influencing transit ridership. For example, using DRMs to predict BRT ridership in Los Angeles County, Cervero (2009) found that park and ride lot capacity is a significant factor (positive) in increasing ridership. Kuby et al. (2003) also found that dummy variables for terminal and transfer stations are all significant (positive) (11). Gutierrez et al. (2011) found that nodal accessibility at stations, number of transit lines, and bus feeders are significant characteristics (positive) (6). Chu (2004) found that the Transit Level of Service (TLOS), an indicator developed by the Florida Department of Transportation (FDOT) to capture transit availability and mobility features (includes three separate measures: service frequency, hours of service, and spatial service coverage) within a walking distance positively affects the total weekday boardings (14).

DRMs have been demonstrated to predict ridership impacts of smart growth policies such as Transit Oriented Development (TOD). Cervero used direct ridership model to specify transit ridership bonus that TOD design elements contribute to. In a case study analyzing Hong Kong, he compared coefficients in three ridership models: one without TOD elements as a baseline model, one with additional park and ride lot dummy variable, and the other one with TOD dummy variable. The results showed that the model with park and ride lot did not increase transit ridership significantly, either for weekends or weekdays, but TOD did
increase station boarding for both weekdays and weekends (12). This empirical study illustrates that the DRMs helps to quantify TOD’s positive effects on transit ridership with more accurate estimates. DRMs also improve our understanding of conventional four-step model results and extend analysis capability by providing additional information. For example, Cervero’s Charlotte case study findings showed that the DRMs can add further detail into the conventional four-step model results. He found that higher combined population and employment density implies jobs-housing balance, and thus reduces the car ownership rate of TAZs with TOD (7).

There are also challenges in applying DRMs. Direct ridership models generally have small sample sizes since observations consist of transit stations or stops (from a high of 261 in a nationwide TCRP model to a low of 27 for a model of St. Louis MetroLink Stations) (8). Thus, degree of freedom constraints often limit the number of variables that can be included as well as their specifications (e.g., inclusion of interaction terms) (5, 7). DRMs are useful policy tools since they can express the ridership's elasticity with respect to independent variables (8). However, they do not consider some variables that four-step models utilize, typically in mode-choice models, such as comparative travel times and prices of transit versus auto (5). Thus DRMs should be seen as a sketch-planning tools that complement conventional travel demand models, well suited for producing order-of-magnitude estimates of patronage and for probing the sensitivity to key input variables in smart growth scenarios, but not as a replacement of a fully specified travel demand models (5, 7).

In this study we developed DRMs for rail transit using 117 stations in the Baltimore-Washington Metropolitan area. These models are developed considering as many factors as possible from the literature that are known to impact transit ridership. The model results are analyzed with specific emphasis on supplementing four-step model analysis results regarding rail transit ridership. The purpose is to provide policy guidance to the State of Maryland in achieving its transit ridership and GHG reduction goals by providing station level analysis results. The study is expected to help improve effectiveness of Maryland’s TOD policies and give insight on further opportunities.

The paper is organized as follows: the next section describes the data and the analysis framework followed by the modeling approach and the model estimation. The results of the analysis and policy implications are presented in section 4. Finally, conclusions are given in Section 5.

BUILDING A DRM FOR MARYLAND

This section covers building the DRM for Maryland. It includes the study area, data preparation, and modeling approach.

Study Area

The study area encompasses the State of Maryland. The State consists of 23 counties, one independent city and with estimated population of 5.8 million in the year 2012 (15). The state operates a transit system that includes urban bus and metro rail transit in Baltimore and Washington D.C. operated by Maryland Transit Administration (MTA) and the Washington Metropolitan Area Transit Authority (WMATA) respectively. The MTA service also include the Maryland Area Regional Commuter (MARC) intercity train service, light rail (LRT), an extensive commuter bus service, and 25 Locally Operated Transit Systems (LOTS) (1). We included Washington DC Metro (WMATA) service in the study area as well since it is part of the Baltimore Washington Metropolitan area and serves to Montgomery and Prince Georges.
Counties of Maryland. We identified 117 rail stations in the area where ridership data is available (Figure 1).

**FIGURE 1** Study area and rail system in Maryland.

**Data and the Modeling Approach**

In order to develop DRMs, we identified available data in three main categories: transit service, station built environment and socio-demographics (Table 1). Transit service data included daily boardings, availability of park and ride facility and feeder bus services, transit service frequency at the station, station catchment size (defined by the distance to the nearest adjacent station), whether it is a terminal station or not and the station connectivity (defined by a composite index including variables such as transit routes, coverage, speed, capacity and urban form (16)). The data regarding station built environment is generated for three buffer sizes to investigate the impact of buffer size on the ridership. Station built environment data included population and employment density, land-use mix index (defined by entropy land-use mix which is explained in the next section), street network connectivity (defined by the number of intersections around station area), regional accessibility (defined by the jobs that can be reached in 30 minutes by auto and transit), distance to central business district (CBD) (since the study area includes two CBDs, we included both Baltimore and Washington DC downtown areas) and walk score at census tract level. Finally, we used vehicle ownership and income level as socio-demographic data.
### TABLE 1 Data category and sources

<table>
<thead>
<tr>
<th>Category</th>
<th>Variables</th>
<th>Description</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transit service</td>
<td>Boardings</td>
<td>Daily boardings (MARC, LRT, Baltimore Metro, and WMATA)</td>
<td>MTA 2011, WMAT</td>
</tr>
<tr>
<td></td>
<td>Park-and-ride</td>
<td>Park-and-ride at station (0,1)</td>
<td>MTA 2011, WMATA 2011</td>
</tr>
<tr>
<td></td>
<td>Feeder bus services</td>
<td>Bus connection at station (0,1)</td>
<td>MTA 2011, WMATA 2011</td>
</tr>
<tr>
<td></td>
<td>Service level</td>
<td>Number of trains in one direction in AM peak</td>
<td>MTA 2011, WMATA 2011</td>
</tr>
<tr>
<td></td>
<td>Catchment size</td>
<td>Distance to nearest adjacent station</td>
<td>MTA 2011, WMATA 2011</td>
</tr>
<tr>
<td></td>
<td>Terminal station</td>
<td>Station is a terminal station (0,1)</td>
<td>MTA 2011, WMATA 2011</td>
</tr>
<tr>
<td></td>
<td>Station connectivity index</td>
<td>Composite index including transit routes, coverage, speed, capacity, urban form, etc</td>
<td>NCSG 2010</td>
</tr>
<tr>
<td>Station built environment</td>
<td>Density</td>
<td>Population density, employment density</td>
<td>U.S. Census 2010, LEHD 2008</td>
</tr>
<tr>
<td>(¼ -, ½-, 1-mile buffer)</td>
<td>Land-use mix index</td>
<td>Entropy land-use mixture</td>
<td>Property View 2010</td>
</tr>
<tr>
<td></td>
<td>Street network connectivity</td>
<td>Number of intersections around station</td>
<td>U.S. Census 2010</td>
</tr>
<tr>
<td></td>
<td>Regional accessibility</td>
<td>Accessible jobs within 30 minutes by auto and transit</td>
<td>State wide model, LEHD 2008</td>
</tr>
<tr>
<td></td>
<td>Distance to CBD</td>
<td>Distance to Baltimore and DC downtown</td>
<td>Census 2010</td>
</tr>
<tr>
<td></td>
<td>Walk score</td>
<td>Walk score at census tract level</td>
<td>NCSG 2010</td>
</tr>
<tr>
<td>Socio-demographics</td>
<td>Vehicle ownership</td>
<td>Mean vehicle per occupied housing unit within catchment area of station</td>
<td>U.S. Census 2010</td>
</tr>
<tr>
<td>(¼ -, ½-, 1-mile buffer)</td>
<td>Income</td>
<td>Income level within catchment area of station</td>
<td>U.S. Census 2010</td>
</tr>
</tbody>
</table>

**Model Specification**

We developed a DRM to estimate the effects of built environment variables together with transit service operational and design features, and socio-demographic characteristics on transit ridership. Station built environment variables data were gathered from multiple data sources (see Table 1) for light rail (LRT), metro (Baltimore and Washington D.C) and commuter rail (MARC) stations in the Baltimore-Washington metropolitan area and were further merged to generate composite data for 117 station proximity areas for both light rail stations and commuter rail stations in the Baltimore–Washington metropolitan areas. We
Liu, Erdogan, Ma and Ducca utilized five built environment variables in this research. Each measurement is discussed in detail next. In order to test the sensitivity of transit ridership in response to built environment in different transit catchment areas, a series of station buffer areas were created: (quarter-mile buffer, half-mile buffer and one-mile buffer). The variables used in DRM and model specification are presented next.

Direct Ridership models estimate boardings at a station for a predefined period of time (e.g., daily, weekly) as a function of three variable sets. In this study, we used average daily station boardings in 2011 as dependent variable. The model uses a combination of variables directly input and variables constructed from other data. The transit service variables tends to be directly input while data on population, employment and station characteristics are converted to constructed variables such as density and accessibility. The details of variables and model specification are presented below:

**Transit Service Variables** - MTA provided average daily boardings for: three MARC lines (the Brunswick Line, the Camden Line, and the Penn Line), Baltimore Metro lines and for Baltimore light rail. WMATA provided station boardings data for five metro lines (Green, Orange, Red, Yellow and Blue Lines). It is worth noting that multiple routes can use a single transit station. Thus, boardings were modeled separately for each line in the analysis since different rail systems may behave differently. Level of service (LOS) was captured by using number of trains in one direction during AM peak period (from 5:50 AM to 11:20 AM). Parking and connecting feeder bus service availability were the key variables used in the DRM models to estimate boardings at a given station, as they capture the access and egress to and from other modes. Dummy variables were utilized to indicate whether parking lots and feeder bus service were available at a given rail station. Terminal stations tend to have more boardings than other stations (7). It is also treated as a dummy variable in the model. To measure the nodal connectivity of the station within the network, a node-level transit connectivity index was incorporated in the analysis (16). Finally, Catchment area, a standard measurement in DRM defined by the distance to the nearest station on the same line, is operationalized.

**Built Environment Variables** - Six main variables are used to describe built environment: density, street network connectivity, land-use mix, accessibility, distance to CBD and walk score.

**Density** - Population density and employment density at the block level were utilized. Block-level population data collected from the U.S. Census 2010 Summary File 1 (SF1) were merged to station buffer areas. Employment data at block-level were obtained from the Longitudinal Employer and Household Dynamics (LEHD) 2008 (17). Densities were then calculated by dividing population and employment data by the area of buffer zones. We applied log transformation to both population density and employment density and generated interactive terms to account for non-normality. We used (population, population density) and (employment, employment density) as the interactive variables allows us to avoid from the nonlinear relationship of station boardings and density.

**Street Network Connectivity** - Street network connectivity is measured by the number of intersections (except cul-de-sacs) within a ¼ mile, ½ mile, and 1 mile buffer zone of each station. Connectivity measurement is a variable indicating the connectivity of streets. To obtain connectivity measure, a street network layer was overlaid on the buffer layer and the number of intersections within the buffer zones were calculated. Street network data were
obtained from the U.S. Census Tiger 2000 files. The connectivity of a station increases as the number of intersections within station buffer zone increases.

**Land-use mix index** - Three land-use types, namely residential, commercial and industrial are considered in this study. A land-use mix index is used to capture how evenly the square footage of commercial, residential, and industrial floor area is distributed within station buffer zones (¼ mile, ½ mile, and 1-mile buffers). The land-use mix index is calculated as follows (Eq. 1):

\[
\text{Land-use mix} = \left(\frac{-1}{\ln n}\right) \sum_{i=1}^{n} p_{i} \ln p_{i}
\]

(Eq. 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) (18). 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** - A gravity-based accessibility measure is used to define accessibility from a zone to all other zones. The gravity-based accessibility measure provides accurate estimates of the accessibility of zone \(i\) to opportunities in all other zones in the region, where fewer and/or more distant opportunities provide diminishing influences (19). Accessibility measure for zone \(i\) in a region with \(n\) TAZs \((i = 1, 2, ..., n)\), \(A_{i}\), is represented as a function of number of opportunities in zone \(j\) \((j = 1, 2, ..., n)\) and impedance function between zones \(i\) and \(j\) as follows:

\[
A_{i} = \sum_{j} O_{j} f(C_{ij})
\]

(Eq. 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 \(f(C_{ij})\) is based on the theoretical work of Wilson (1971) (20), and is expressed as \(f(C_{ij}) = \exp(-\beta C_{ij})\), where \(\beta\) is an empirically calibrated parameter. Employment data that were used to represent the opportunities in TAZ \(j\) in calculating accessibility were obtained from LEHD 2008.

**Distance to CBD** - Central city remains the main trip attractor of the Baltimore-Washington Metropolitan region. We would expect that stations that are closer to the Central Business District (CBD) would have higher ridership.

**Walk score** - There are many walkability indices cited in the literature. Based on the area scale, level of detail, and data availability for this project, we adapted the Walk Score\textsuperscript{®}
method to calculate the walkability of each census tract in the study area (21). The walk score
uses data from a number of sources: Amenity data from ES202 2008 and Road network data
from U.S. Census Tiger file 2000.

Socio-demographic Variables—Median household income and mean vehicle per
occupied housing unit within the catchment area of a given station were used. Previous
literature suggests that low-income households and households without access to vehicles
tend to rely more on transit than higher-income households and households owning one or
more vehicles.

ESTIMATING DIRECT RIDERSHIP MODEL

Ordinary least squares (OLS) regression was used to estimate direct ridership model based on
Maryland experiences data since it was shown to have a better prediction performance than
other statistical models (5). We also tested the sensitivity of transit ridership in response to
built environment in different transit catchment areas and found that different catchment
areas have little influence on a model’s predictive power. Nevertheless, we found that land
use variables performed better when using the half-mile buffer as the catchment benchmark
in the model.

Tables 2 and 3 present the descriptive statistics for the dependent and independent
variables. The comparison of descriptive statistics of socio-demographics and built
environment characteristics between MARC stations and light rails stations showed that
unlike light rail stations are located in the city or suburb areas, the commuter rail lines serve
more dispersed rural areas, likely with lower density, less walkable, lower job accessible,
highly residential, and concentrated with rural development patterns. People who reside in the
MARC station surrounding areas have higher vehicle ownership and higher income. We
learned that people who are taking commuter rail behaved differently from the people who
are taking light rail, the model was split into different modes (7). Due the small sample size
of light rail stations, we decided to combine different light rail stations together to compare
the results. Many variables have the multi-colinearity issue in the model. After testing
performance of different variable combinations in the models, we came up with the final
model specification. The variables used in the final models are discussed below.

<table>
<thead>
<tr>
<th>TABLE 2 Descriptive statistics of MARC stations (N=39)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Minimum</strong></td>
</tr>
<tr>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>Boardings</td>
</tr>
<tr>
<td>Park and Ride</td>
</tr>
<tr>
<td>Bus</td>
</tr>
<tr>
<td>Level of Service</td>
</tr>
<tr>
<td>Distance to Nearest Station</td>
</tr>
<tr>
<td>Terminal</td>
</tr>
<tr>
<td>Connectivity Index</td>
</tr>
<tr>
<td>Population Density within Half Mile Buffer</td>
</tr>
<tr>
<td>Employment Density within Half Mile Buffer</td>
</tr>
<tr>
<td>Land Use Mixed Index at Half Mile Buffer</td>
</tr>
<tr>
<td>Street connectivity at half</td>
</tr>
</tbody>
</table>
### TABLE 3 Descriptive Statistics of Light Rail, Baltimore Metro and WMATA stations

(N=73)

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boardings</td>
<td>113.0</td>
<td>13865.0</td>
<td>2938.7</td>
<td>2959.6</td>
</tr>
<tr>
<td>Park and Ride</td>
<td>0.0</td>
<td>1.0</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>Bus</td>
<td>0.0</td>
<td>1.0</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>Level of Service</td>
<td>20.0</td>
<td>60.0</td>
<td>41.9</td>
<td>20.0</td>
</tr>
<tr>
<td>Distance to Nearest Station</td>
<td>0.0</td>
<td>3.2</td>
<td>0.9</td>
<td>0.6</td>
</tr>
<tr>
<td>Terminal</td>
<td>0.0</td>
<td>1.0</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>Connectivity Index</td>
<td>0.0</td>
<td>4.3</td>
<td>0.6</td>
<td>1.1</td>
</tr>
<tr>
<td>Population Density within Half Mile Buffer</td>
<td>0.0</td>
<td>18754.1</td>
<td>6105.0</td>
<td>4940.0</td>
</tr>
<tr>
<td>Employment Density within Half Mile Buffer</td>
<td>258.6</td>
<td>137032.0</td>
<td>22626.4</td>
<td>35352.6</td>
</tr>
<tr>
<td>Land Use Mixed Index at Half Mile Buffer</td>
<td>0.0</td>
<td>1.0</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>Street connectivity at half mile buffer</td>
<td>4.0</td>
<td>347.0</td>
<td>113.5</td>
<td>104.2</td>
</tr>
<tr>
<td>Accessibility of Auto Half Mile Buffer</td>
<td>539932.2</td>
<td>1951165.5</td>
<td>1074789.</td>
<td>310931.4</td>
</tr>
<tr>
<td>Accessibility of Transit within Half Mile Buffer</td>
<td>155979.4</td>
<td>1722464.6</td>
<td>777796.9</td>
<td>359076.2</td>
</tr>
<tr>
<td>Distance to CBD</td>
<td>22668.8</td>
<td>238366.1</td>
<td>134976.3</td>
<td>72062.9</td>
</tr>
<tr>
<td>Walk Score</td>
<td>2.3</td>
<td>237.3</td>
<td>54.8</td>
<td>66.3</td>
</tr>
<tr>
<td>Number of Vehicle Owned by Household</td>
<td>0.6</td>
<td>2.2</td>
<td>1.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Household Income</td>
<td>25783.00</td>
<td>1110961.00</td>
<td>100976.13</td>
<td>168758.96</td>
</tr>
</tbody>
</table>

4
Results

5

Tables 4 and 5 present the results of the final regression models of commuter rail stations and light rail stations. The results from commuter rail stations suggest that the bus connection is statistically significant in the model. Also, the adjusted R² shows that roughly 44 percent of the variation in the boarding of commuter rail stations is explained by all the variables in combination.
Non-commuter rail stations were grouped into three samples to show the differences: light rail alone, light rail and the Baltimore Metro line, and the combination of light rail, Baltimore Metro line, and WMATA stations. It is interesting to see that two variables are significantly associated with station boardings across all three samples: bus connection and employment. Both variables have expected signs. The results suggest that bus connection and higher employment will increase the station boardings for all rail lines. Population is not significantly associated with the station boardings. This suggests that employment plays a more important role in promoting transit uses via affecting built environment around station area, even though the magnitude of the coefficient of the employment variable is modest. The rail station catchment area shows significantly and expected relationship in the LRT model only. The higher level of service is statistically significant when the Baltimore Metro and WMATA are included with the LRT but not for LRT alone. The adjusted R² shows that roughly 56.6 percent of the variation in the boardings of light rail stations is explained by all the variables, 72.3 percent of the variation in the boardings of light and Baltimore Metro stations is explained by all the variables, and 81.2 percent of the variation in the boardings of light rail, Baltimore Metro and WMATA stations is explained by all the variables.

Parking at stations bears further discussion. The initial intent was to determine the significance of parking as a factor in transit station boardings. While one only needs to view parking at the stations to understand it is a significant factor we were not able to determine the statistical significance. Parking was treated as a dummy variable, and every commuter rail station has parking, eliminating the possibility of statistical tests.

**TABLE 4 Regression model predicting MARC station boardings (N = 39)**

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-6.91</td>
</tr>
<tr>
<td>Park and Ride</td>
<td>NA</td>
</tr>
<tr>
<td>Bus</td>
<td>2.99</td>
</tr>
<tr>
<td>Ln(Distance to CBD)</td>
<td>0.792</td>
</tr>
<tr>
<td>Ln(Population density within half mile buffer)</td>
<td>0.015</td>
</tr>
<tr>
<td>Ln(Employment density within half mile buffer)</td>
<td>0.066</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.44</td>
</tr>
</tbody>
</table>
### TABLE 5 Regression model predicting LRT, Baltimore Metro and WMATA station boardings

<table>
<thead>
<tr>
<th>Variables</th>
<th>LRT Coefficient</th>
<th>Significant</th>
<th>LRT, and Baltimore Metro Coefficient</th>
<th>Significant</th>
<th>LRT, Baltimore Metro, and WMATA Coefficient</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-6.54</td>
<td></td>
<td>4.783</td>
<td></td>
<td>4.666</td>
<td></td>
</tr>
<tr>
<td>Terminal</td>
<td>0.615</td>
<td>0.48</td>
<td>**</td>
<td></td>
<td>0.528</td>
<td>***</td>
</tr>
<tr>
<td>Park and Ride</td>
<td>0.103</td>
<td></td>
<td>0.195</td>
<td></td>
<td>1.147</td>
<td></td>
</tr>
<tr>
<td>Bus</td>
<td>0.515</td>
<td>**</td>
<td>0.520</td>
<td>**</td>
<td>0.537</td>
<td>**</td>
</tr>
<tr>
<td>Ln(Distance to the nearest station)</td>
<td>0.049</td>
<td>**</td>
<td>0.034</td>
<td></td>
<td>0.042</td>
<td>**</td>
</tr>
<tr>
<td>Ln(Distance to CBD)</td>
<td>0.901</td>
<td></td>
<td>-0.158</td>
<td></td>
<td>-0.244</td>
<td>**</td>
</tr>
<tr>
<td>Ln (Population density within half mile buffer)</td>
<td>NA</td>
<td></td>
<td>NA</td>
<td></td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Level of service</td>
<td>NA</td>
<td></td>
<td>0.980</td>
<td>***</td>
<td>0.132</td>
<td>***</td>
</tr>
<tr>
<td>Number of employment within half mile buffer</td>
<td>0.00001</td>
<td>***</td>
<td>0.00001</td>
<td>***</td>
<td>0.000009</td>
<td>***</td>
</tr>
<tr>
<td>#Pop* Population density within half mile buffer</td>
<td>0.000000002</td>
<td></td>
<td>0.0000000001</td>
<td></td>
<td>0.000000001</td>
<td></td>
</tr>
</tbody>
</table>

N=32                                        | R²=0.566        |             |                                      |             | N= 46                                       | R²=0.723    |

N= 72                                        | R²= 0.812       |             |                                      |             |                                             |             |

Notes: “***” indicates coefficient is statistically significant at the 0.01 level. “**” indicates coefficient are statistically significant at the 0.05 level.
Policy implications

The DRM model provides useful insights into how land use and transportation attributes interact to influence transit boardings at the station level. The results suggest that these influences have significant impacts on transit ridership but that the impacts differ depending on whether light rail or commuter rail is being analyzed. For light rail stations, employment within a half mile buffer area of the station, transit service level, feeder bus connectivity, station location in the CBD, the distance to the nearest neighboring station, and whether or not the station is a terminal station all affect transit boardings. For commuter rail, (MARC stations in this study), the feeder bus connection is found to be significant. This suggests that commuter rail behaves differently from light rail. For light rail, employment is the most significant predictor of station boardings. Increasing the employment in a transit station area 1% can result in a 1.6% increase in boardings.

The DRM results have implications for areas wishing to increase transit ridership. They show that increasing employment can impact transit ridership and increase transit boardings at individual transit stations. Areas desiring to increase transit ridership should consider zoning regulations and site design requirements which allow for denser development around transit stations. However, increasing densities must be combined with transit service levels, parking and feeder bus service to take full advantage of transit lines.

The DRM cannot substitute for traditional mode choice models, which are essential for determining system level characteristics of transit and for developing forecasts on ridership on individual transit lines. At the same time, the DRM can complement traditional mode choice models by estimating the effects of changes in station boardings, and thus transit ridership, resulting from changes in urban form.

CONCLUSIONS

The DRM model can be useful to provide some insights on how different land use and transit attributes can affect boardings. The results suggest that impacts of built environment show differences for light rail and commuter rail. For light rail stations, employment at half-mile buffer areas, service level, feeder bus connectivity, station located in CBD, and terminal stations are significant factors affecting ridership. Among these variables, employment is the most significant predictor of station boardings (p<0.001) and the results are consistent across all the models. Increasing the employment by 1% can result in 1.6% increase in boardings. For MARC stations (commuter rail), only feeder bus connection is found to be significant, which suggests that commuter rail behave differently from light rail. Adding more bus access to the commuter rail station will be a good policy for improving the commuter rail ridership. Parking was not included in the MARC model since all the MARC stations have parking available. Therefore, we need additional variables to test the significance of parking of MARC stations.

In summary, DRM can provide rough estimates of station boardings without relying on complicated transportation demand model and extensive data collection process. It can capture the key features of built environment characteristics, transit service attributes and their relationships with station boardings, which can provide timely policy guidance on how to improve the transit usage. It is also worth noting the limitations of the current DRM model. First, the DRM was developed for state of Maryland only. Future research should be carried out for other regions to get a general pattern. Second, other important factors should be incorporated in
the model, such as safety and transit reliability attribute. Third, more detailed parking information, such as parking fee should also be included in the model.

ACKNOWLEDGEMENT

The authors are thankful to Maryland Department of Transportation (MDOT) for the continued support in the development of models used in this study. The opinions and viewpoints expressed are entirely those of the authors, and do not necessarily represent policies and programs of any agency.

REFERENCES

3. The National Center for Smart Growth Research and Education at the University of Maryland. Maryland Scenarios Project Draft Final Year 3 Report. 2012.