The Demand for Rail Transit: an Analysis of Boston's MBTA

Lucas Conwell

April 27, 2014

lucas.conwell@tufts.edu
423-227-0963

Submitted to:

EC 127: Urban Economics

attention Dr. Anna Hardman

This paper estimates a derived demand model at the station level for the MBTA heavy and light rail system, making use of demographic, transit service-related, and built environment variables to run a simultaneous equations regression. I confirm the hypotheses that average speed of the rail line, median income, percentage of residents under age 35, and WalkScore positively impact transit ridership at the station level, but find no evidence that headway is negatively correlated with demand. Ideally, my results should provide Massachusetts policymakers and practitioners in the field of transportation planning with a clearer picture of the driving forces behind rail transit demand

Introduction

As cities grow, diseconomies of agglomeration such as traffic congestion and pollution begin to counterbalance the benefits of urban agglomeration, decreasing firm productivity and consumer utility. Enacting policies or programs that counteract these negative externalities could thus yield large payoffs, and one of the most widespread tools used to reduce congestion and air pollution, or at least improve mobility, is public transportation. In order to inform this policy conversation, this paper attempts to isolate the main determinants of public transportation demand.

The effectiveness of public transportation in reducing highway congestion has been widely studied with mixed results (Anderson 2013; Duranton and Turner 2009); whether or not public transportation is truly an effective policy tool in this sense, it will indisputably continue to play a large role in the development of urban areas, and so much can be gained from a deeper knowledge of the factors that influence the success, as measured by ridership, of existing and new transit lines. As McFadden (1974) explains, a reliable cost-benefit analysis, some form of which accompanies every new transit project, must take into account the predicted ridership, which can best be obtained from a derived demand model. This type of analysis expresses some demand variable, i.e. ridership, as a function of characteristics of people, the built environment, the transit service, and so on (Cervero 1995; McFadden 1974).

My paper seeks add to the existing literature by applying this framework to the MBTA heavy and light rail system in Boston, testing more specifically the hypotheses that the average speed of the rail line, median income, percentage of residents under age 35, and mixed use and pedestrian-friendly development, as measured by the Walk Score framework, positively impact transit ridership at the station level. I then hypothesize that headway, on the other hand, is negatively correlated with demand. The use of Walk Score data in this type of regression is, to my knowledge, unprecedented, and few other studies incorporate both built environment and transit quality variables. Ideally, my results will provide Massachusetts policymakers and practitioners in the field of transportation planning with a clearer picture of the driving forces behind rail transit demand, allowing them to more accurately predict ridership on proposed extensions and new routes while also identifying underperforming stations in the current system.

Literature Review

The intuitive nature of the direct demand model can be deceiving; indeed, there exist a multitude of more specific models which build on this basic framework. I identify three categories of direct demand research papers: first, those which analyze ridership between pairs of origins and destinations, second, papers which examine mode choice decisions at the level of the individual, and third, studies focusing on the determinants of individual stations' ridership.

The first class generally aggregates data for origin-destination pairs of geographic regions to develop a regression model. Thompson, Brown, and Bhattacharya (2012) use census and travel survey data on bus trips in Broward County, FL to study travel demand between pairs of census tracts. By including measures of travel time to represent the price of travel by bus, the authors find that traditional measures of walkability, employment density, and mixed-use development are statistically insignificant in explaining origin-destination ridership. Frank and Pivo (1994) use a similar framework, making use of data from the Puget Sound region of Washington State. Their findings echo those of Thompson, et al. in that their measure of land use mix is significant and positive only in the walking regression, suggesting that mixed-use development plays only a minor role in determining transit and auto mode shares. I test these counterintuitive findings directly in my model by including each station's WalkScore, which incorporates both mixed-use development and the pedestrian amenities of a neighborhood.

Thompson et al. and Frank and Pivo differ, however, on their assessment of the impact of employment density: while Thompson et al. find that even employment density is insignificant

(Thompson, Brown, and Bhattacharya 2012), Frank and Pivo's analysis does reveal a significant employment density coefficient (Frank and Pivo 1994). Naturally, my model does include a proxy for employment density to control for its impact.

While Frank and Pivo do not consider the time "price" of transit travel, Thompson et al. include a measure of its impact and find in vehicle time is negatively correlated with ridership and highly significant. Accordingly, they conclude that reducing travel time by linking far-flung employment centers should be a transit system's first priority and that creating walkable, mixed-use environments may not be as important to driving transit ridership as is often thought. Thompson et al. do mention themselves that one should avoid extrapolating their findings to larger transit systems that also operate rail transit (Thompson, Brown, and Bhattacharya 2012); my paper nonetheless includes both the aforementioned WalkScore and a measure of travel time to downtown Boston as well as the headway of transit service and can thus shed light on the relative importance of the built environment and of travel time.

A second, larger body of literature analyzes public transportation demand using individual trips as the unit of analysis. These papers often make use of binary logit models to predict the probability of choosing transit over driving for a given trip. McFadden (1974) pioneered this form of direct demand model, using data from travel surveys in the San Francisco area and binary dependent variables indicating mode choice. His regression of the mode choice variable on measures of bus and car travel and wait time as well as income indicates, among other things, that travelers seem to value time spent in buses and cars equally and that wait time at a bus stop is significant while access time is not (McFadden 1974). McFadden's results underscore the importance of travel time (a vital component of the overall "cost" of a trip) for each mode in explaining mode choice and thus provide a rationale for including the speed of travel to downtown and the headway of rail transit service, a measure of wait time, in my own model.

McFadden did not consider the impact of the built environment on mode choice, but its impact plays a major role in later research. Cervero and Kockelman (1997) focus on the influence of the density, diversity (as in mixed-use development), and pedestrian-friendly design in generating individual transit trips, using data from field and travel surveys in fifty San Francisco neighborhoods. They find, however, that pedestrian-friendly design and mixed-use development are only borderline significant – but do have positive coefficiants – in explaining non-work trips. Density, surprisingly, is insignificant (Cervero and Kockelman 1997).

Cervero and Kockelman are not alone in discounting the influence of mixed-use development and pedestrian-friendly design. Boarnet and Crane (2001) find further evidence in support of Cervero's earlier work, making use of ordered probit estimation to conclude that land use variables—density and mixed-use development—are insignificant in explaining non-work auto trips, but that these might influence mode choice and thus travel demand by influencing the relative "prices" of automobile and transit modes (Boarnet and Crane 2001). Kitamura (1997) and Cervero (2005) further discount the influence of traditional built environment variables, finding that individuals' preferences for transit over auto (or vice versa) play the largest role in determining transit ridership, thus suggesting a self-selection problem that could invalidate the vast majority of transit-demand literature (Kitamura 1997; Cervero 2005).

Not all studies in this second, individual level class dismiss the role of the built environment. Cervero (2002), using a binomial logit model and individual-level data from a travel survey in the Metro Washington, D.C. Region, concludes that high density and mixed land uses do positively impact transit ridership; his analysis makes use of not only an entropy index to measure land use mix but also of gross density (employment and population) and miles of sidewalk. These variables all have positive effects and are significant. His variables measuring the additional time cost of transit (or lack thereof), however, are insignificant (Cervero 2002). Similarly, Cervero (1995) estimates a direct demand model using nationwide, individual level American Housing Survey data, and every single one of his built environment variables is significant. By including population density, each station's WalkScore, and a proxy for employment density, my paper tests these two competing conclusions.

Finally, a third class of models uses route segment or station-level data to estimate public transportation demand within a given transit system. Peng (1997) published the first of these papers, making use of a simultaneous equations model to account for the likelihood that transit demand and supply are simultaneously determined and adding variables such as headway, population in the "pedestrian catchment area" surrounding the station, number of poor households in the PCA, parking spaces, and nearby competing routes to analyze the ridership in segments of bus routes in Portland, OR. He does find that previous year ridership plays a large role in determining the supply of transit service, confirming the need for a simultaneous model and also concludes that residential as well as employment densities, competing route service, and park and ride lots all significantly influence demand for transit—these effects are all positive, with the exception of competing route service (Peng 1997).

Peng's model, however, is quite susceptible to omitted variables bias; not only does he include few demographic variables such as age or percentage foreign born, but he also includes no measure of the built environment beyond population and employment density. My analysis will include most of Peng's variables, including park and ride lots, and will use the Walk Score framework to account for the built environment influences ignored by Peng.

Estupinan and Rodriguez (2008) build on Peng's model by studying the bus rapid transit system in Bogota, Colombia at the station level. They point out that the built environment's characteristics—including pedestrian amenities, density, and mixed uses (or lack thereof)--may well also be endogenous to transit ridership but that this concern likely does not apply to Bogota since the BRT system is so new. Estupinan's factor analysis and simultaneous equations model yield the result that density, pedestrian friendly design, and mixed-use development all contribute to transit ridership while income negatively affects transit ridership; oddly, factors representing connectivity are insignificant (Estupinan and Rodriguez 2008). My model tests each of these factors in turn, but I am unable to make the claim that the Boston region's built environment is exogenous to transit use.

While Estupinan (2008) includes a vast array of variables, both his and Peng's instruments are of questionable validity. Peng (1997), for example, uses employment density as an instrument for the supply variable in his demand equation, thus suggesting, implausibly at best, that employment density has no effect on ridership (demand) except indirectly through its influence on supply. Likewise, Estupinan introduces the presence of feeder buses, the presence of transit alternatives, population density, and income around a station as instruments to estimate the coefficient on supply in the demand equation. These all seem to be characteristics which might influence the consumer's utility maximizing choice of transit mode, and would thus appropriately belong in the demand equation and could not be employed as instruments. Accordingly, this paper attempts to rectify this problem, adopting the same form of simultaneous equations model while selecting instruments that could more plausibly be exogenous to transit supply.

Station-level models are not limited to simultaneous equations: Choi et al. (2012) and Zhao et al. (2014) make use of multiplicative and poisson models to study ridership between pairs of stations in the Seoul, Korea and Nanjing, China metro systems, respectively. Choi finds that both residential and office floor area are significant and positively affect ridership but that travel time of auto and metro play the greatest role in explaining station to station ridership. Zhao's regression produces similar results, but also finds bicycle park and ride facilities, feeder buses, and a CBD dummy variable to be significant. The CBD dummy variable can serve as a proxy for other factors, such as congestion, which especially plague the CBD but are difficult to measure and include.

However, both studies consider very few demographic or land use variables (Choi et al. 2012;

Zhao et al. 2014). In fact, neither study considers any demographic variables, and one can hardly argue that income, race, or vehicles per capita are constant across neighborhoods or irrelevant. As for the built environment, Choi et al. (2012) do include measures of employment, pedestrian friendly intersections, population, and floor area devoted to commercial uses, but these variables are quite rough proxies for the qualities that are normally assumed to influence transit use, namely true mixed-use development—not simply the presence of large factories or warehouses, which would add to the commercial floor area—and walkable design. Moreover, neither paper addresses the possible endogeneity of supply or the built environment. Here again, in my model, the WalkScore variable and the simultaneous equations attempt to correct for the aforementioned biases.

Theoretical Foundation

My paper will fall into this third category: I will estimate a direct demand model for MBTA rail transit at the station level, using MBTA heavy and light rail stations as my unit of analysis. Direct, or derived, demand models, at their core, originate from the basic microeconomic theory of consumer utility maximization. Using an adapted version of a theoretical framework employed by Boarnet and Crane (2001), I start with the assertion that individuals face the usual problem of constrained utility maximization,

$\text{Max}_{\{T,W,A,X\}}\ U(T,W,A,X)$ s.t. $I = p_T T + p_W W + p_A A + p_X X$ (Boarnet and Crane 2001)

T is the number of trips by public transit, W is the number of trips on foot, A is the number of trips by car, and X represents all other goods and activities. Both prices and income take into account not only the monetary cost of buying a good or engaging in an activity but also the time cost. Solving this problem then yields an individual's demand function for transit trips, as modeled by Boarnet and Crane (2001):

T^* = f(prices, “income”, socio-demographic variables and preferences) (Boarnet and Crane 2001)

Aggregating the individual transit demand curves for all individuals within a given radius (pedestrian catchment area) of a transit station yields the total demand for transit service at that station:

$Q_d = \Sigma T_i^* =$ F(prices, "income", socio-demographic variables and preferences)

However, one cannot directly measure any of the variables in the above function, so proxy variables must suffice. Relative prices of different modes, then, can enter into the model as trip speed, waiting time, and even density and walkability, since these all influence the time cost of residents' trips. Competing routes' presence can be viewed as reducing the price of substitutes, and nominal income can proxy both the opportunity cost of time and the total income in the above model, while feeder bus routes and park and ride lots serve as complements, and an increase in their "convenience" - a decrease in "price" - should increase demand for transit service at a given station. Age and income proxy for socio-demographic variation, while employment and density jointly control for the varying sizes of "markets" served by each station. In the case of transit, one must amend the model to include the supply of transit service (S):

Q_d = Ridership = F(S, prices, income, socio-demographic characteristics)

This extension is necessary because greater supply reduces crowding and wait times, directly increasing the attractiveness of transit—this type of relationship is relatively unique to the "market" for transit. Supply certainly affects the time price of transit travel, but the overall price is not determined jointly with quantity, as in a normal market.

Because the quantity of transit service "supplied" plays this unique role, one must also consider the role of ridership (demand) in determining supply. Therefore, after estimating the demand equation by itself, I expand the model into a two equation system.

S = G(Ridership, other factors)

Ridership = F(S, prices, income, socio-demographic characteristics)

Whether this extension is necessary, however, depends on whether ridership indeed directly influences the relative level of supply at one station. Certainly, since all MBTA trains run the entire length of their lines, ridership differences between stations on each line will not cause the supply to vary between stations on the same line. However, one could plausibly argue that increases in aggregate ridership on each line might lead the MBTA to slightly increase service frequency, or in the long run, buy larger train cars. Not only the work of Boarnet and Crane (2001) but also that of Estupinan and Rodriguez (2008) and Peng (1997) influenced the development of this model.

Data and Variables

Table 1. Summary of data

Variable	Description	Mean	Std. Dev.	Min	Max
rid	Station entries on typical weekday	4492.513	5153.573	48	28156
supp	Total weekday capacity through station	310937.8	250824	47112	894024
bus	# of bus routes serving station	2.760684	3.845395	0	16
comp	Other rail stops w/i 0.5 mi	0.794872	0.942653	0	3
downt	dummy for downtown location	0.136752	0.345063	0	1
dist	distance to downtown transfer station	3.980376	2.55925	0	12.019
speed	average speed of trip to downtown	0.237646	0.082795	0.125	0.472727
heavyr	dummy for heavy rail lines	0.410256	0.493996	0	1
branch	dummy for stations on branch	0.581197	0.495485	0	1
par	# park and ride spots at station	119.1282	413.8052	0	2733
headw	rush hour headway at station	5.697479	1.428784	1.555	9
popdens	Population density in PCA	18338.46	11860.23	1937.9	60205.3
emp	employed workers in census tract	8514.829	13620.63	285	72110
ws	WalkScore	81.20513	14.01326	37	100
pcfborn	% foreign born in census tract	24.54761	10.86379	10.4	68.2
noveh	% households with no vehicle in tract	37.09658	21.34965	2.7	84.9
pcyoung	% under age 35 in PCA	58.77009	14.9294	28.7	94.7
medinc	Median income in PCA	62269.31	31114.37	11860	161959

N=117

Sources: "MBTA Subway Map: Interactive Street Map"; "Ridership and Service Statistics" 2010; "System Map"; American Community Survey 2008-2012, Social Explorer; American Community Survey 2006-2010, Census Transportation Planning Products; Google; Walk Score.

First, I collect data on variables ("Ridership and Service Statistics" 2010) related to the quality and quantity of MBTA service at 117 of the 121 rail transit stations on the Red, Blue, Orange, Green, and Ashmont-Mattapan lines, excluding the four transfer stations, Downtown Crossing, Park Street, State Street, and Government Center, because ridership at these four is likely driven by different factors than at outlying stations. Naturally, *ridership* measures the total station entries (i.e. trips in both directions originating at that station) on a typical weekday in 2009, while *supply* indicates the total passenger capacity of all trains on a line in both directions on a typical weekday, calculated using the number of one-way trips, the number of cars per train on each line, and the passenger capacity of each line's cars. One would expect supply to be positively correlated with ridership (Peng 1997).

The Ridership and Service Statistics document also provided insight into the time cost of MBTA travel ("Ridership and Service Statistics" 2010). *Headway*, then, measures the average time between trains during weekday rush hour service in 2009. To model in-vehicle component of travel time, I calculated the *average speed* in miles per minute of this trip to downtown, again using 2009 data. Across the board, measures of wait time and in-vehicle time are negative and significant (Peng 1997; Thompson, Brown and Bhattacharya 2012; McFadden 1974; Boarnet and Crane 2001; Choi et al. 2012). Here, then, headway should be negatively correlated with ridership and average speed should be positively correlated. Unlike Choi et al. (2012), however, my measure of travel time is only a proxy for the travel time that travelers at each station experience; of course, many travelers boarding at each station are not traveling to downtown. It does seem reasonable to believe that, given the extremely monocentric configuration of the MBTA rail system and the reasonably CBD-oriented structure of Boston, that the travel time to downtown can reasonably represent the travel times experienced by the majority of travelers.

Finally, the MBTA also provided data on the number of park-and-ride spaces in each station's lot ("Ridership and Service Statistics" 2010). As logic would suggest, Peng (1997) finds that (auto) park and ride lots positively and significantly influence ridership. One should note, however, that park and ride lots are much less common along the heavy and light rail lines in the Boston region than along the commuter rail system.

The Ridership and Service Statistics document omits several important service-related variables, however. Using the MBTA website ("MBTA Subway Map: Interactive Street Map."), I obtained the number of bus routes serving each rail station in 2014 as well as a variable measuring the number of MBTA rail transit routes with stations that lie within a half mile radius of a given station ("MBTA Subway Map: Interactive Street Map;" Google; American Community Survey 2008-2012, Social Explorer). Feeder buses bring passengers from outlying areas to rail stations and thus increase ridership at the stations they serve, all else equal (Zhao et al. 2014; Estupinan and Rodriguez 2008). A look at the system map reveals that the vast majority of MBTA bus routes converge on rail stations from neighborhoods in between or beyond the reach of rail lines and thus function as complementary feeder routes ("System Map"). On the other hand, rail stations on other lines located within half a mile of a rail station are substitutes. Peng (1997) finds, not surprisingly, that ridership on competing routes has a negative effect on a route's ridership.

Demographic data also plays an important role in my analysis. The Census Transportation Planning Products website provided data on the number of workers employed in the census tract in which each station is located, a very rough proxy for employment density (Google; American Community Survey 2006-2010, Census Transportation Planning Products). Generally, one would expect that greater employment density around a station would increase demand (ridership) since more workers would have the option of walking to the station, and indeed a substantial body of literature confirms this assumption (Frank and Pivo 1994; Peng 1997; Chu 2004). Surprisingly, some studies find that employment density is in fact insignificant (Thompson, Brown, and Bhattacharya 2012; Zhao et al. 2014). To measure further demographic variables, I created a rough pedestrian catchment area for each station which encompasses all block groups that lie, in part or in whole, within a 0.2 mile radius of the station on the Red, Orange, and Blue lines, and within 0.1 mile of each station on the Green line, since many stops on the Green line are located extremely close together. In my final analysis, I include, calculated for each pedestrian catchment area, the population density, the median income, and the percentage of residents under age 35 (Google; American Community Survey 2008-2012, Social Explorer).

Not surprisingly, many analyses reveal that population density (or a proxy thereof) has a positive effect on ridership (Cervero 2002; Peng 1997; Estupinan 2008; Chu 2004; Choi et al. 2012; Zhao et al. 2014; Cervero 1995); after all, higher population densities mean that more people live within a given walking distance of each station. Then again, Boarnet and Crane (2001) as well as Cervero and Kockelman (1997) find that population density is insignificant.

Income, on the other hand, negatively influences transit use in most studies (Estupinan 2008, Peng 1997; Cervero and Kockelman 1997; Boarnet and Crane 2001; Chu 2004; Kitamura et al. 1997) since lower-income riders are less likely to be able to afford the cost of fuel for extensive driving.

Age, surprisingly, has rarely been a focus and enters into most studies only as a control. Age is insignificant in Kitamura et al (1997) and Frank and Pivo (1994) but has a positive effect on transit ridership in Boarnet and Crane (2001)'s San Diego data and in Chu (2004). On the other hand, Boarnet and Crane (2001)'s Orange County data as well as Cervero (2007) find that age negatively affects transit demand. Thus, the evidence is quite mixed; in this paper, I posit that age will negatively affect transit use, following recent studies and news stories' hypotheses that young professionals today, much more so than their parents, prefer transit to driving (Homan 2014).

Surprisingly, not a single study that I examined explicitly included the percentage of foreign born residents; however, I will include this variable (Google; American Community Survey 2008-2012, Social Explorer) as it seems reasonable to suppose that foreign born residents will be more likely to use transit since they often come from countries in Europe, Asia, or even Latin America where transit's mode share is much higher.

Including the number of vehicles per capita or per household is quite standard; naturally, households without cars will be more likely to use transit. This variable (Google; American Community Survey 2008-2012, Social Explorer) could also proxy for households' preferences for transit use over automobile or vice versa. While the vehicles variable is insignificant in Thompson, Brown, and Bhattacharya (2012) and Boarnet and Crane (2001), most other studies find that the number of vehicles is negatively related to transit use, or equivalently, that the number of carless households is positively related to demand (Cervero 1996; Kitamura et al. 1997; Cervero 2002; Frank and Pivo 1994; Chu 2004; Cervero 2007; Cervero and Kockelman 1997).

In addition to population density and employment, a variable recording each station's Walk Score accounts for variation in the built environment (Walk Score). The Walk Score of a station serves as a joint proxy for the pedestrian friendliness of the street grid and the level of mixed-use development —as the website explains,

> For each address, Walk Score analyzes hundreds of walking routes to nearby amenities. Points are awarded based on the distance to amenities in each category. Walk Score also measures pedestrian friendliness by analyzing population density and road metrics such as block length and intersection density ("Walk Score Methodology").

Frank and Pivo (1994) and Thompson et al. (2012) find that the coefficients on mixed land use and pedestrian friendly design are insignificant, as do Kitamura et al. (1997) and Cervero (2007). On the other hand, Cervero (2002), Cervero (1995), Cervero and Kockelman (1997), and Estupinan (2008) find that both mixed use development and pedestrian friendly design positively affect transit use by allowing errands to be accomplished on foot on the way to transit stations and by encouraging this same pedestrian travel that is vital to robust transit use.

In a similar vein, a downtown dummy variable ("System Map") proxies for, as described by Zhao et al. (2014), "other [omitted] factors that may affect CBD ridership, such as the parking costs and congestion and the greater densities very close to the station" and the "aggregation effect of monocentric urban form" (Zhao et al. 2014, p. 144). Accordingly, this dummy takes the value 1 for stations located in the central business district of Boston, loosely defined as the Downtown/Financial District, West End, Back Bay, Chinatown, Theater District, and Kenmore areas. Most past literature finds the CBD variable, for obvious reasons, to be positively correlated with ridership (Zhao et al. 2014; Cervero 1996; Cervero 2002; Choi et al. 2012); Thompson et al., however, find that stops in the CBDs of Hollywood and Ft. Lauderdale receive less ridership, all else equal, likely because employment in Broward County, FL follows a dispersed, decentralized pattern (Thompson, Brown, and Bhattacharya 2012).

Finally, I include three additional variables in supply that serve as instruments to estimate the structural demand equation. It should be noted that these variables are based on logic rather than on the results of previous research; indeed, instruments used in similar papers do not seem to be valid for the MBTA rail system. *Distance to downtown* measures the distance along the rail line's route from each station to the closest of the aforementioned four downtown transfer stations ("Ridership and Service Statistics" 2010). This variable proxies for the additional costs of operating longer lines, under the assumption that stations further out will see lower levels of service on average since longer lines often receive less frequent service. *Branch* is a dummy variable taking the value 1 for stations located past (moving outward from downtown) a junction where a line branches into two or more branches ("System Map"). Here, the joint "trunk" line portion where branches converge limits capacity on the branches below the branches' individual capacity. Finally, *Heavy Rail* is a dummy variable taking the value 1 for the Red, Orange, and Blue lines and 0 for the Green line; since heavy rail vehicles have a higher capacity than the light rail vehicles used on the Green line and run in longer trains, the heavy rail lines should see a higher level of service supplied, all else equal ("System Map").

My analysis will add a new angle to existing station-level transit demand literature. To my knowledge, no study in the vein of Estupinan and Rodriguez (2008) or Zhao, et al. (2014) has been conducted for the MBTA rail system, and so I seek to discover whether the results of these past studies also hold for the Boston region. Few past studies have included both comprehensive built environment variables and variables representing the speed and quality of the supply of transit service—Estupinan and Rodriguez (2008) is one of the only other studies to focus on both of these areas. My analysis, then, extends their model to study a rail transit system in the United States, a typology for which there exist surprisingly few transit demand studies. I also introduce the use of Walk Scores to represent pedestrian friendliness and mixed-use development; this computer algorithm did not exist when much of the literature was written and may in fact more accurately capture these qualities than field surveys (Estupinan and Rodriguez 2008) or simple counts of the commercial floor area (Choi, et al. 2012) might.

I hypothesize that income around a station in Boston is in fact positively related to use of rail transit, rather than negatively, as in many other American cities (Thompson, et al. 2012; Glaeser, Kahn, and Rappaport 2008; Estupinan and Rodriguez 2008; Peng 1997; Cervero and Kockelman 1997; Boarnet and Crane 2001; Chu 2004; Kitamura et al. 1997). Despite the above mentioned past findings to the contrary, my hypothesis is that a station's WalkScore will positively impact ridership. Few studies have considered the impact of age on transit use, and here I hypothesize that, in Boston, younger neighborhoods will in fact see higher ridership. My final two hypotheses—that headway is negatively related to ridership and that the speed of the trip between a given station and downtown is positively related to ridership—are quite conventional, but the significance of these variables would provide confirmation of the findings of previous studies, including Peng (1997), Choi, et al. (2012), and Zhao, et al. (2014) and of commonly held beliefs.

Results

First, I estimate the simple demand equation using OLS. A Breusch-Pagan test for heteroskedasticity shows no evidence thereof (p-value=1.000).

$$Rid_i = \beta_0 + \beta_1\, supp_i + \beta_2\, bus_i + \beta_3\, comp_i + \beta_4\, downt_i + \beta_5\, speed_i + \beta_6\, par_i + \beta_7\, headw_i + \beta_8\, popdens_i + \beta_9\, emp_i + \beta_{10}\, ws_i + \beta_{11}\, pcfborn_i + \beta_{12}\, noveh_i + \beta_{13}\, pcyoung_i + \beta_{14}\, medinc_i + u_i$$

Table 2. OLS

Model 1: Demand Equation, OLS

dependent variable: ridership

Variable	Description	Coefficient	Standard Error
supp	Weekday station entries	0.008852***	0.0015
bus	# feeder bus routes	337.67***	77.42
comp	Other rail stops w/i 0.5 mi	-603.7	376.48
speed	average speed of trip to downtown	11226.15**	5057.79
par	# park and ride spots	0.0193	0.646
headw	rush hour headway	-42.73	216.49
popdens	Population density in PCA	0.008009	0.0293
emp	employed workers in census tract	0.04089	0.0251
ws	WalkScore	48.336*	26.224
downt	dummy for downtown location	3513.8***	1160.55
pcfborn	% foreign born in census tract	-20.328	26.145
noveh	% households with no vehicle in tract	22.493	22.493
pcyoung	% under age 35 in PCA	53.145**	25.098
medinc	Median income in PCA	0.0259**	0.0111
constant		-11111.81***	3138.761
R^2		0.787	
Adj. R^2		0.758	
Prob > F		0	
N		117	

* indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level

Sources: see Table 1

Average speed of the trip to downtown, as expected, positively affects ridership. The speed variable is significant at the 5 percent level (p-value = .029); an increase of .5 mi/min (30mi/hr) in average trip speed to downtown is predicted to increase station ridership by roughly 5,613 passengers per weekday. Evidently, the time price of travel plays a large role in passengers' decision to ride MBTA rail transit. However, when discussing the average trip speed for urban rail transit, 30mi/hr is an enormous increase, so ridership increases from plausibly attainable speed increases would likely be smaller.

More surprisingly, the OLS results also confirm my hypothesis that ridership increases with median income. Median income is significant at the 5 percent level, yet the effect is quantitatively much smaller: a $10,000 increase in median income in the station's pedestrian catchment area increases predicted ridership by only 259 passengers per day. This result, however small, contradicts almost all previous literature on the subject (Glaeser, Kahn, and Rappaport 2008; Estupinan and Rodriguez 2008; Peng 1997; Cervero and Kockelman 1997; Boarnet and Crane 2001; Chu 2004; Kitamura et al. 1997; Thompson, Brown, and Bhattacharya 2012)—Thompson, et al., for example, found the opposite effect in South Florida—yet substantiates the intuitive observation that Boston residents, much more often than those in the vast majority of metro areas, ride transit by choice rather than because they have no other option. The positive effect of income might also demonstrate that the MBTA system better serves higher-income riders or that lower-income riders have a stronger preference for car travel.

Like income, the percentage of young people in the neighborhood has a positive effect on ridership—a five percent increase in the percentage of residents under 35 increases predicted ridership by 265.75 passengers/day; once again, this effect is quantitatively small yet reveals an interesting trend. Ordinarily, one might assume that younger neighborhoods might house more families who would be less likely to use transit; though this measure of the "age" of a neighborhood is a rough one, these results seem to confirm that either the millennial generation or families tend to have a relatively greater preference for transit.

Pedestrian friendliness and mixed land uses do appear to play a role in determining the popularity of rail transit; a 10 point increase in a station's WalkScore increases rail ridership by roughly

480 persons per day, but this effect is significant only at the 10 percent level (p-value=.068). Since the mean station ridership is 4,492.5, this effect is not extremely substantial either. Nevertheless, the walkability of a neighborhood does seem to play a role in incentivizing transit use.

I find no substantial role for headway; it is extremely insignificant in the OLS regression (p-value=.844), but its coefficient is negative as expected, so the evidence contradicts my hypothesis. Of the remaining built environment variables, only the downtown dummy variable is significant at the five percent level, with a positive effect on ridership. Employment in the census tract of a station is almost significant at the 10 percent level (p-value=.106), but surprisingly, population density is completely insignificant (p-value=.785); both variables do have the expected positive signs. To some extent, this lack of significance can probably be explained by multicollinearity. *Ws*, *downt, emp,* and *popdens* are jointly quite significant (p-value=.0001), so the built environment does seem to play a major role in determining a station's ridership.

Most other control variables do not seem to influence ridership appreciably. Only the number of bus routes serving a station is significant (p-value=.000); one additional bus route serving a station is predicted to increase ridership by 338 passengers per day. The significance of this variable reveals the importance of bus-rail transit transfers in the MBTA system—when seeking to expand rail ridership, the MBTA might do well to concurrently expand its bus service, a complement for riding the subway or light rail. The remaining two demographic variables, the percentage foreign born and the percentage of households with no vehicles in the PCA, are both insignificant at any conventional level of significance. One might note that the four purely demographic variables—*pcfborn, noveh, pcyoung,* and *medinc*—are jointly very significant (p-value=.0239). The number of competing rail stations and number of park and ride lots are insignificant as well, though *comp*'s effect (-603.70) is larger in magnitude than most others and has the expected sign—nearby rail lines compete with each other, siphoning off passengers.

However, these results all rely on the assumption that supply is exogenous to transit demand, that there is no simultaneous causality between ridership and supply. Indeed, supply has a quantitatively small effect—adding 3000 “seats” (the equivalent of 30 Green line trolleys) to a station's service would increase predicted station ridership by only 27 passengers per day—but is very significant, with a p-value of .000. Supply seems to influence ridership, almost beyond a doubt, so if ridership also causes supply, the resulting endogeneity could introduce serious bias and inconsistency into the results.

I test formally for the endogeneity of supply using a two-step procedure, first regressing supply on all exogenous variables and instruments and then adding the residuals from this regression to the structural demand equation. The residuals are significant at the 5% level (p-value=.027), confirming that supply is indeed endogenous and that only a simultaneous equations model can produce unbiased coefficient estimates.

Model 2. 2SLS

Accordingly, I proceed by estimating a two-equation simultaneous model using two stage least squares. As explained in the **Data** section, *branch, heavyr,* and *dist* are all presumed to influence the supply of transit service. However, logically, none of these variables should affect demand, i.e. ridership. The fact that a station is on a branch or is closer or farther from downtown should not intrinsically influence ridership; of course, these variables are correlated with factors such as density, walkability, service supply, and headway, but once such factors-which vary between branch and non branch stations and as one moves further out from downtown-have been controlled for in the demand regression, neither *branch* nor *heavyr* should appear in the demand equation or be correlated with the error term. To claim the same for *heavyr*, one must assume that passengers have no intrinsic preference for heavy over light rail vehicles, holding other included factors such as capacity and frequency of service (i.e. headway) constant—however, if, for example, heavy rail vehicles have a smoother ride or more comfortable seats, *heavyr* should appropriately be included in the demand regression, since I have no measure of ride quality or seats. I will proceed by assuming that no such preference for heavy over light rail exists, once the other variables that I include in demand are controlled for. Then, *heavyr* will not be included in the demand equation and will likely not be correlated with the demand error term. Therefore, *Branch, heavyr,* and *dist* can all be used as instruments to correct for the endogeneity of supply in a 2SLS process. Peforming a Hausman test of overidentifying restrictions, setting H_0: all instruments exogenous, yields a chi-square p-value of .427. Thus, we fail to reject H_0, which simply indicates that the three instruments could be either all exogenous or all endogenous and create biases in the same direction. Thus, while the Hausman test is by its nature inconclusive, the result does yield some additional confidence in the validity of the instruments.

Expanding the model into a system of two equations yields:

$$Rid_i = \beta_0 + \alpha_1\, supp_i + \beta_1\, bus_i + \beta_2\, comp_i + \beta_3\, downt_i + \beta_4\, speed_i + \beta_5\, par_i + \beta_6\, headw_i + \beta_7\, popdens_i + \beta_8\, emp_i + \beta_9\, ws_i + \beta_{10}\, pcfborn_i + \beta_{11}\, noveh_i + \beta_{12}\, pcyoung_i + \beta_{13}\, medinc_i + u_{i1}$$

$$Supp_i = \delta_0 + \alpha_2\, rid_i + \delta_1\, heavyr_i + \delta_2\, branch_i + \delta_3\, dist_i + u_{i2}$$

Once again, a Breusch-Pagan test finds no evidence of heteroskedasticity (p-value=1.000).

Table 3. 2SLS

Model 2: Demand Equation, 2SLS

dependent variable: ridership

Variable	Description	Coefficient	Standard Error
supp	Weekday station entries	0.0066***	0.00173
bus	# feeder bus routes	361.64***	73.783
comp	Other rail stops w/i 0.5 mi	-598.44*	355.28
speed	average speed of trip to downtown	14724.85***	5003.67
par	# park and ride spots	0.16911	0.6131
headw	rush hour headway	-117.132	206.78
popdens	Population density in PCA	0.00437	0.0277
emp	employed workers in census tract	0.0434*	0.0237
ws	WalkScore	62.091**	25.442
downt	dummy for downtown location	3868.47***	1105.71
pcfborn	% foreign born in census tract	-13.97	24.82
noveh	% households with no vehicle in tract	17.51	20.65
pcyoung	% under age 35 in PCA	58.97**	23.82
medinc	Median income in PCA	0.0233**	0.0105
constant		-12180.66***	2997.3
R^2		0.7827	
Prob > X^2		0	
N		117	

* indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level

Sources: see Table 1

The simultaneous model confirms the role of average trip speed to downtown in influencing station ridership; speed is now significant even at the 1 percent level, and its effect is even larger: increasing average trip speed by 30mi/hr increases predicted station ridership by approximately 7,362 riders per day, ceteris paribus, rather than 5,613 as predicted by the OLS regression. My result thus closely tracks that of previous literature underscoring the importance of travel time, especially time spent actually on a transit vehicle, in determining ridership (Peng 1997; Thompson, Brown and Bhattacharya 2012; McFadden 1974; Boarnet and Crane 2001; Choi et al. 2012).

Median income and the percentage of residents under 35 continue to positively affect transit ridership in the 2SLS regression, as both are still significant at the five percent level, and their effects resemble those predicted by OLS quite closely.

In examining the effect of the Walk Score, however, estimating a 2SLS model does appreciably affect the results. A station's Walk Score is now almost significant at the 1 percent level (p-value=.015), and its coefficient is naturally still positive but 28 percent larger in magnitude than the corresponding coefficient in the OLS regression. Now, a ten point increase in *ws* will increase predicted daily ridership by approximately 621 passengers, controlling for other relevant factors, chiefly because mixed use, pedestrian friendly neighborhoods make the walk to a transit stop more pleasant and allow riders to accomplish errands along the way or on separate trips without a car.

2SLS provides no more evidence in favor of my hypothesis with regards to headway than does OLS; headway is still negative and insignificant. This result may well reflect the fact that headway does not vary at all between stations on the same line and only slightly between stations on different lines.

The results on the remaining built environment control variables also change little; population density is still insignificant, while employment in the given station's census tract is now significant at the 10 percent level (p-value=.067) and still positively affects ridership. Unsurprisingly, the downtown dummy variable remains significant at the 1 percent level; all else equal, the predicted ridership for a station downtown will be 3868.47 passengers per day higher than that of an otherwise identical station located outside of the downtown area. Here, the data again concretely reveals the substantial influence of the "omitted factors" which *downt* represents (Zhao et al. 2014; Cervero 1996; Cervero 2002; Choi et al. 2012**).**

In a similar vein, neither any of the other controls nor the supply variable see any substantial changes in significance, sign, or magnitude. Feeder bus routes remain highly positively correlated with ridership; the relevant coefficient, which is still significant at the 1 percent level, is now 7 percent larger in magnitude, compared to the OLS results. The number of stations on competing routes within a half mile radius barely attains significance at the 10 percent level (p-value=.092), and the number of park and ride spots is insignificant as ever. Both the percentage of foreign-born residents and the number of vehicles per household fail to attain significance at any conventional level. Multicollinearity might well explain the insignificance of *noveh*, however. *Pcfborn, noveh, pcyoung,* and *medinc* are still jointly significant at the five percent level, while the coefficient on supply is as significant yet small in magnitude as ever. In sum, though statistical testing suggests that supply is endogenous and that the instruments may well be valid, the use of 2SLS to estimate the derived demand equation for MBTA rail transit does not affect the results or the conclusions in any practically significant way.

Finally, I estimate two variations on the simultaneous equations model described above.

Table 4.

Functional Form Variations

		Model 3: Demand, 2SLS, with interactions		Model 4: Demand, 2SLS, ws^2	
		dependent variable: ridership		*dependent variable: rid*	
Variable	**Description**	**Coefficient**	**Standard Error**	**Coefficient**	**Standard Error**
supp	Weekday station entries	0.0061***	0.00177	0.00669***	0.00172
bus	# feeder bus routes	374.52***	73.36	361.66***	73.602
comp	Other rail stops w/i 0.5 mi	-551.45	351.92	-576.5	356.13
speed	average speed to downtown	14337.66***	5135.63	14269.72***	5026.78
par	# park and ride spots	0.2912	0.6112	0.13678	0.6134
headw	rush hour headway	-80.547	210	-95.086	208.8
popdens	Population density in PCA	0.00707	0.0274	0.00312	0.0277
emp	employed workers in tract	0.0461**	0.0235	0.0428*	0.0236
ws	WalkScore	-123.64	117.62	-50.434	180.32
downt	dummy for downtown location	3975.77***	1098.24	3717.93***	1126.12
pcfborn	% foreign born in tract	-8.196	24.711	-12.62	24.87
noveh	% carless households in tract	10.109	21.813	12.78	21.988
pcyoung	% under age 35 in PCA	-75.464	137.046	58.957**	23.757
medinc	Median income in PCA	-0.0628	0.0458	0.0218**	0.0108
wsinc	Ws * inc	0.00118*	0.00061		
agews	Age * ws	1.889	1.684		
ws^2	WalkScore quadratic			0.7734	1.2303
constant		1122.31	9268.36	-8064.533	7164.68
R^2		0.788		0.7838	
Prob > X^2		0		0	
N		117		117	

* indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level

Sources: see Table 1

First, I generate two interaction terms, WalkScore*Income (*wsinc)* and Age*WalkScore (*agews)*, to test whether the effect of pedestrian friendliness and walkability depends on income or age and vice versa. *Agews* is insignificant, indicating that the effect of age on ridership does not seem to depend on the value of *ws* and that the effect of WalkScore on ridership does not depend on the percentage of young people in the neighborhood. The fact that *agews* and age as well as *agews* and *ws* are both jointly insignificant seems to confirm this conclusion. This disqualifies any number of intuitive hypotheses; one might think, for example, that younger people might be more willing to walk farther distances to use transit and that therefore they would flock to transit in greater numbers, even relative to the rest of the population, in more walkable neighborhoods than in more car-oriented areas. On the other hand, one could plausibly assume that younger people would be more willing to make sacrifices, e.g. unpleasant walks, to use transit, and that lower WalkScores would in fact increase the effect of the percentage of young people on daily ridership.

Unlike those of *age*, the effects of median income seem to interact in more complex ways with the effect of walkability (*ws*). While *ws* is now insignificant, WalkScore*Income is in fact significant at the 10 percent level, and almost so at the 5 percent level. WalkScore*Income and WalkScore are jointly insignificant, WalkScore*Income and median income are jointly significant at the 5 percent level (p-value=.0117). The coefficient on this interaction term, .00118, indicates that the effect of median income on ridership is larger, the greater the WalkScore in a neighborhood. For example, when *ws* = 50, a $10,000 increase in median income yields a decrease in daily ridership of 38 persons, while an equivalent increase in median income would cause an increase in daily ridership of 552 persons when *ws* = 100. Thus, we gain additional insight into the effect of income; though the coefficient on *medinc* in this regression is negative, the effect of income in fact depends on walkability. In less walkable neighborhoods, median income is negatively correlated with transit use, but in more walkable areas, the opposite is true. Intuitively, this divergence makes sense—where the built environment orients itself around the auto, those who can afford to own several cars shy away from transit use, but where neighborhood design complements transit, the rich use transit in greater numbers. This effect may reflect the fact that the rich travel more, or that the destinations to which they travel (i.e. downtown offices) are better served by the transit system. The effect of *ws*, therefore, will be more positive, the greater the median income in the neighborhood. While the effect of WalkScore does not seem to depend on income, the effect of income does increase with walkability.

Most other variables change neither significance nor sign. Employment, which was significant only at the 10 percent level in the original 2SLS regression, is now significant at the 5 percent level and positively correlated with ridership. Oddly, percent young is now insignificant, possibly due to multicollinearity or because the interaction term *wsinc* may have somehow accounted for age—for example, walkable areas may have more wealthy young people, and therefore, when I omitted the interaction, *pcyoung* picked up this interplay between walkability and income.

In addition to the interaction model, I estimated a model adding a quadratic of WalkScore; this quadratic term and the level of WalkScore are jointly significant at the 5 percent level (p-value=.0423). Hence, the marginal effect of walkability increases in value as the WalkScore increases; though the coefficient on *ws* is now negative, the combined effect of the two terms will be positive and increasing for all $ws > 8.06$, and no MBTA station has WalkScore lower than 8.06. At low levels of walkability, intuitively, minor improvements will not increase ridership much, but once an area is already reasonably walkable, improvements in pedestrian infrastructure and mixed-use development can yield more significant improvements in ridership. This result provides support for the many public policy initiatives, from zoning reform to pedestrian crossing rebuilding, that improve walkability in an already walkable area.

As for the other controls, the number of competing stops becomes insignificant even at the 10 percent level, but no other control variables experiences changes in sign or significance or major changes in magnitude, relative to the original 2SLS regression.

Limitations

First, this study has failed to take account of the possible endogeneity of the built environment. Development—and particularly the type of mixed-use, dense development that increases WalkScores and of course, population density—often springs up around transit stations. In fact, the very idea of transit oriented development implies that the causal relationship between mixed-use, walkable development and transit ridership goes both ways (Cervero 2007). Thus, except in the case of new transit systems, a simultaneous equations model that also treats WalkScore and/or population density as endogenous might better capture the complex relationships between ridership, supply, and the built environment (Estupinan and Rodriguez 2008). The data available for this paper, however, contained no plausible instruments for the built environment; land use regulations, local utility and road infrastructure, local political attitudes, and the geologic composition of the local ground might all serve as instruments to estimate the effect of *ws* or *popdens*, but these variables are all either nonexistent or difficult to obtain for individual neighborhoods. Thus, simultaneity between *popdens* and *rid* or *ws* and *rid* may have caused biased coefficients in my regressions.

Secondly, the proxies used to represent the relative time prices of auto and transit travel – the average speed of a transit trip to downtown from each station and the headway – are crude ones. If the speed of car travel from any point in the MBTA system to downtown were constant and all riders traveled from outlying stations to downtown, my proxy would exactly model this time differential. Obviously, these assumptions would apply exactly only in a perfectly monocentric city with equal traffic congestion throughout. The question, then, should be whether the part of the "true" time price of travel from each station that is not explained by my proxies is correlated with any other independent variables. If so – if, for example, more walkable stations see a higher proportion of trips that involve transfers and non-downtown destinations – then all of my coefficient estimates may be biased. However, assuming that the most MBTA rail passengers spend the majority of their trip on one rail line, the speed and headway factors could model the general "speed" and thus relative time advantage over car travel for that entire line. Additionally, because of the monocentric orientation of the MBTA rail system, one can probably assume that most people use the rail system primarily for trips to the center, opting for bus or car travel when traveling from one outlying area to another.

As in all econometric studies, omitted variables pose a substantial concern. I do not control for residents' preferences for different transit modes, as suggested by Boarnet and Crane (2001), Kitamura (1997), and Cervero (2005). Including this variable would have required conducting an extensive survey, and the omission of this factor means that all coefficients may be biased. In particular, my analysis may overestimate the effect of walkability and mixed-use development, if indeed these types of developments attract people who are already predisposed to use transit. The coefficients of age and income may also well overestimate the impact of these variables if younger and higher income people are more likely to innately prefer transit use over driving. Likewise, omitting the percentage of residents who commute by transit in each station's neighborhood (Goetzke 2008) may bias all coefficients, as could omitting the percentage of commuters in each pedestrian catchment area who commute to downtown. Naturally, since the MBTA is so highly monocentric, neighborhoods with a higher share of downtown workers would see higher transit use. Finally, time constraints prevented me from collecting data on crime and education; higher crime would naturally deter transit use, while education might serve as a proxy for the preferences mentioned above.

Conclusion

The results of my regressions do not represent a radical departure from previous literature; however, this paper confirms that commonly held hypotheses relating to transit use also apply in Boston and suggests several new relationships. Both the OLS and 2SLS regressions strongly confirm my hypothesis that speed will positively influence transit ridership, indicating that the time cost of transit plays a huge role in travelers' mode choice decisions. The insignificance of the headway variable, however, complicates these findings, and I thus find no evidence of my hypothesis that longer headways will reduce transit ridership. This oddity may be the result of relatively similar headways systemwide. Furthermore, the data reveals that income does seem to be positively correlated with transit ridership in Boston, though the effect is small in magnitude, contradicting most previous research (Estupinan and Rodriguez 2008; Peng 1997; Cervero and Kockelman 1997; Boarnet and Crane 2001; Chu 2004; Kitamura et al. 1997; Thompson, Brown, and Bhattacharya 2012). Accordingly, the common assumption (Glaeser, Kahn, and Rappaport 2008), at least in the US, that transit riders tend to be low income people with no other choice seems not to apply in the dense, transit-filled metropolis of Boston. My paper confirms my own hypothesis – and Boarnet and Crane (2001) as well as Cervero (2007)'s findings – that neighborhoods with a greater proportion of young people see higher transit use. This result could indicate that younger adults in fact prefer transit more so than older generations or that families use transit in greater numbers than childless households. Finally, echoing Cervero (2002), Cervero (1995), Cervero and Kockelman (1997), and Estupinan (2008), I find that pedestrian friendliness and mixed land uses, as represented by the Walk Score framework, have an unambiguous positive effect on transit ridership across the MBTA rail system, and the Walk Score seems to have a greater influence in higher-income as well as already more walkable areas. Interestingly, controls often thought to influence transit ridership, including the percentage of foreign born residents, percentage of households with no vehicles, and population density, turn out to be insignificant.

Admittedly, the conclusions of direct demand models can often seem overly abstract. An assessment of the driving forces of demand can, however, give policymakers and transit executives a clearer picture of the effectiveness—or lack thereof—of various demand-maximizing policies. For example, since speed clearly plays a major role in riders' mental calculus, upgrading infrastructure to increase travel speed should be a major priority. Immediately concluding that headway does not matter would be a mistake; in order to truly isolate the influence of headway, one would have to examine a system where headways vary widely, e.g. where some lines run every 5 minutes and some every 10 or 20. Policymakers might reasonably conclude, however, that small differences in headway of 1 to 2 minutes, as seen during rush hour on the MBTA, do not appreciably influence ridership. In a similar vein, the fact that median income has a slight positive effect on ridership confirms that the MBTA is not the typical low income rider focused transit system; in fact, the T might actually be engaging in the opposite sort of bias, providing somehow more attractive service to higher-income neighborhoods or providing monocentric service that better suits high income riders' needs. The positive coefficient on income might also suggest that the MBTA should increase their marketing and outreach to low income populations, if indeed these populations have some innate preference for the automobile relative to their higher income peers. Finding that younger neighborhoods see more ridership could have two explanations: either families or young people, or both, are predisposed to use transit. Either way, the MBTA might do well to consider how the rail system could better serve the middle-aged as well as senior citizens. Finally, since the regressions reveal a major role for walkability and mixed use development, the MBTA as well as the Metropolitan Area Planning Council (of Boston) can rest assured that they are not pursuing transit oriented development in vain. Of course, the possible endogeneity of the built environment could invalidate these findings, or perhaps TOD simply attracts people who already, exogenously, want to use transit (Cevero 2007). From a policy perspective, however, this nuance may be irrelevant; Cervero himself explains governments have an important role to play in smoothing housing market distortions such as zoning constraints to provide transit "aficionados" with a suitable place to live. Regardless of the true mechanism behind TOD's effect, such development increases ridership in the MBTA system.

Nevertheless, these conclusions represent only a start in the quest to deconstruct the mechanisms behind transit usage in Boston. Future research might collect data on instruments that could be used to estimate a three equation simultaneous model where supply, ridership, and the Walk Score are all endogenous. Including omitted variables, such as preferences and trip destinations, or adding variables for percentage of childless young people and percentage of families to replace the *age* variable, could improve the results, as could a more accurate measure of employment density. The ideal model, then, might use data on station to station ridership—including all aforementioned omitted variables—in a three equation model.

Appendix: Matrix of Correlations

	par	headw	pcfborn	noveh	emp	speed	pcyoung	popdens	medinc	ws	bus	comp	downt	dist	heavyr	branch	supp
par	1.0000																
headw	0.2445	1.000															
pcfborn	0.1285	0.0686	1.0000														
noveh	-0.1769	-0.1681	0.1307	1.0000													
emp	-0.0638	-0.3187	-0.0386	0.3122	1.0000												
speed	0.3104	-0.0103	0.2278	-0.4934	0.0083	1.0000											
pcyoung	-0.2005	-0.0728	-0.1085	0.6215	-0.0205	-0.5042	1.0000										
popdens	-0.2741	-0.0393	0.0043	0.6145	0.0187	-0.5093	0.5658	1.0000									
medinc	0.0715	0.0213	-0.2322	-0.5616	0.0796	0.3022	-0.5625	-0.4836	1.0000								
ws	-0.2040	-0.1680	-0.1061	0.5834	0.3845	-0.4154	0.4056	0.5060	-0.2317	1.0000							
bus	0.1364	-0.1718	0.1939	0.0379	0.1254	0.4245	0.0019	-0.1106	-0.1358	0.0576	1.0000						
comp	-0.2299	-0.1126	-0.2153	0.4932	0.4935	-0.4415	0.1958	0.2473	-0.0307	0.4796	-0.1706	1.0000					
downt	-0.1014	-0.5080	-0.0218	0.3173	0.6517	-0.0807	-0.0656	0.0565	0.1277	0.4274	0.0574	0.4845	1.0000				
dist	0.4156	0.4344	-0.0651	-0.5827	-0.4631	0.3441	-0.3662	-0.4188	0.3671	-0.7110	-0.0909	-0.5218	-0.5169	1.0000			
heavyr	0.2513	-0.0579	0.3725	-0.0886	0.1434	0.5761	-0.2232	-0.2525	-0.0617	-0.0397	0.5059	-0.1695	0.1738	-0.2005	1.0000		
branch	0.0171	0.6304	-0.1855	0.0836	-0.2362	-0.4636	0.2901	0.2702	0.0174	0.0621	-0.3969	0.1098	-0.3176	0.2873	-0.6656	1.0000	
supp	0.1768	-0.3543	0.2165	0.0115	0.3366	0.4957	-0.0825	-0.1506	-0.0227	0.2010	0.5114	-0.0543	0.3647	-0.3582	0.8410	-0.6730	1.0000

Sources: see Table 1

References

Anderson, Michael L. 2013. “Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion.” *NBER Working Paper Series* w18757.

Boarnet, Marlon, and Randall Crane. 2001. “The Influence of Land Use on Travel Behavior: Specification and Estimation Strategies.” *Transportation Research Part A: Policy and Practice* 35 (9): 823–45. doi:10.1016/S0965-8564(00)00019-7.

Brown, Jeffrey R., and Gregory L. Thompson. 2008. “Examining the Influence of Multidestination Service Orientation on Transit Service Productivity: A Multivariate Analysis.” *Transportation* 35 (2): 237–52. doi:http://dx.doi.org/10.1007/s11116-007-9140-x.

Cervero, Robert. 1996. “Mixed Land-Uses and Commuting: Evidence from the American Housing Survey.” *Transportation Research Part A: Policy and Practice* 30 (5): 361–77. doi:10.1016/0965-8564(95)00033-X.

———. 2002. “Built Environments and Mode Choice: Toward a Normative Framework.” *Transportation Research Part D: Transport and Environment* 7 (4): 265–84. doi:10.1016/S1361-9209(01)00024-4.

———. 2007. “Transit-Oriented Development’s Ridership Bonus: A Product of Self-Selection and Public Policies.” *Environment and Planning A* 39: 2068–85.

Cervero, Robert, and Kara Kockelman. 1997. “Travel Demand and the 3Ds: Density, Diversity, and Design.” *Transportation Research Part D: Transport and Environment* 2 (3): 199–219. doi:10.1016/S1361-9209(97)00009-6.

Choi, Jinkyung, Yong Jae Lee, Taewan Kim, and Keemin Sohn. 2012. “An Analysis of Metro Ridership at the Station-to-Station Level in Seoul.” *Transportation* 39 (3): 705–22. doi:http://dx.doi.org/10.1007/s11116-011-9368-3.

Chu, Xuehao. 2004. “Ridership Models at the Stop Level”. NCTR-473-04, BC137-31. Tampa: National Center for Transit Research: Center for Urban Transportation Research. http://www.nctr.usf.edu/pdf/473-04.pdf.

Deka, Devajyoti. 2014. “The Living, Moving and Travel Behaviour of the Growing American Solo: Implications for Cities.” *Urban Studies* 51 (4): 634–54. doi:10.1177/0042098013492233.

Duranton, Gilles, and Matthew Turner. 2009. “The Fundamental Law of Road Congestion: Evidence from US Cities.” *University of Toronto, Department of Economics, Working Papers, 2009, 70 Pp.*, 70.

Estupiñán, Nicolás, and Daniel A. Rodríguez. 2008. “The Relationship between Urban Form and Station Boardings for Bogotá’s BRT.” *Transportation Research Part A: Policy and Practice* 42 (2): 296–306. doi:10.1016/j.tra.2007.10.006.

Frank, Lawrence D., and Gary Pivo. 1994. "Impacts of Mixed Use and Density on Utilization of Three Modes of Travel: Single-Occupant Vehicle, Transit, and Walking." *Transportation Research Record*, no. 1466: 42–52.

Glaeser, Edward L., Matthew E. Kahn, and Jordan Rappaport. 2008. "Why Do the Poor Live in Cities? The Role of Public Transportation." *Journal of Urban Economics* 63 (1): 1–24. doi:10.1016/j.jue.2006.12.004.

Goetzke, Frank. 2008. "Network Effects in Public Transit Use: Evidence from a Spatially Autoregressive Mode Choice Model for New York." *Urban Studies* 45 (2): 407–17. doi:10.1177/0042098007085970.

Google. 2014. Google Maps. Accessed March 23, 2014. https://maps.google.com/.

Homan, Timothy R. 2014. "Millennials Choose Public Transit Over Private Cars." *The Fiscal Times*, March 12, sec. Business + Economy. http://www.thefiscaltimes.com/Articles/2014/03/12/Millennials-Choose-Public-Transit-Over-Private-Cars.

Kitamura, Ryuichi, Patricia L. Mokhtarian, and Laura Laidet. 1997. "A Micro-Analysis of Land Use and Travel in Five Neighborhoods in the San Francisco Bay Area." *Transportation (1986-1998)* 24 (2): 125.

"MBTA Subway Map: Interactive Street Map." 2014. Online. Boston, MA: Massachusetts Bay Transportation Authority. Accessed April 14. http://www.mbta.com/schedules_and_maps/subway/.

McFadden, Daniel. 1974. "The Measurement of Urban Travel Demand." *Journal of Public Economics* 3 (4): 303–28. doi:10.1016/0047-2727(74)90003-6.

Peng, Zhong-Ren. 1997. "A Simultaneous Route-Level Transit Patronage Model: Demand, Supply, and Inter-Route Relationship." *Transportation* 24 (2): 159–81.

"Ridership and Service Statistics." 2010. 13th Edition. Boston: Massachusetts Bay Transportation Authority. http://www.mbta.com/uploadedfiles/documents/bluebook%202010.pdf.

"System Map." 2014. Online. Boston, MA: Massachusetts Bay Transportation Authority. Accessed April 14. http://www.mbta.com/schedules_and_maps/system_map/.

Thompson, Gregory, Jeffrey Brown, and Torsha Bhattacharya. 2012. "What Really Matters for Increasing Transit Ridership: Understanding the Determinants of Transit Ridership Demand in Broward County, Florida." *Urban Studies* 49 (15): 3327–45. doi:10.1177/0042098012443864.

U.S. Census Bureau. American Community Survey, 2006-2010. Prepared by Census Transportation Planning Products. Accessed March 23, 2014. http://ctpp.transportation.org/Pages/5-Year-Data.aspx.

U.S. Census Bureau. American Community Survey, 2008-2012. Prepared by Social Explorer. Accessed March 23, 2014. http://www.socialexplorer.com/.

Walk Score. 2014. WalkScore.com. Accessed March 23, 2014. http://www.walkscore.com/.

Walk Score. "Walk Score Methodology." 2014. WalkScore.com. Accessed April 14, 2014. http://www.walkscore.com/methodology.shtml.

Zhao, Jinbao, Wei Deng, Yan Song, and Yueran Zhu. 2014. "Analysis of Metro Ridership at Station Level and Station-to-Station Level in Nanjing: An Approach Based on Direct Demand Models." *Transportation* 41 (1): 133–55. doi:http://dx.doi.org/10.1007/s11116-013-9492-3.