

---

# Impacts of New Light Rail Transit Service on Riders' Residential Relocation Decisions

*You-Lian Chu, Ph.D.  
Parsons Corporation*

*Yi Deng, Ph.D.  
MTA New York City Transit*

*Rongfang (Rachel) Liu, Ph.D.  
New Jersey Institute of Technology*

---

## Abstract

*Using the rider survey data collected from Hudson-Bergen Light Rail Transit System in New Jersey, this paper investigated the residential relocating decisions of the riders who have been riding the LRT for a period of time. Using the Heckman's sample selection model, the paper extends the current mobility literature by describing not only a rider's likelihood to move as a result of the new LRT service but also the movers' orientation toward their residence distances to LRT stations. Information on the socio-economic characteristics of the movers and their residence distances to LRT stations would help planners and developers identify areas where housing growth associated with specific characteristics of the riders will occur, and plan for these areas to provide affordable housing and amenities for relocating residents.*

**Key Words:** *Light Rail Transit, Residential Moving Behavior, Sample Selection Model*

## Introduction

Transportation, social, and economic impacts of light rail transit (LRT) service on the communities have been the subjects of many studies including accessibility, property value, dislocation, transit-focused development, auto ownership, and mode shares. However, very little work has been devoted to the examination of LRT's influences on residential moving behavior and the spatial redistribution of riders' residences. In

2008, New Jersey Transit conducted a platform survey along the northern segment of the Hudson-Bergen Light Rail Transit (HBLRT). The survey included the questionnaire about the extent to which the HBLRT service affected riders' moving decisions and their daily living. The survey results showed that 65% of the riders had actually moved in response to the new LRT service. For those riders who had moved, 46% and 24% of them consider, respectively, the LRT as a "very important" and "somewhat important" factor in their moving decision-making process. Based on the HBLRT survey data, this paper focuses on the residential relocating decisions of the riders who have been riding the HBLRT for a period of time. Two relocating decisions were analyzed, including whether or not a rider would move as a result of the new LRT service (discrete choice) and, having decided to move, how far a rider would live away from the nearest LRT stations (continuous choice).

The objective of this paper is to use an econometric model to jointly investigate the rider's discrete move choice and continuous distance choice. A major characteristic of such mixed discrete-continuous models is that the continuous dependent variables are censored because data on residence distance to the station usually exhibit certain number of observations clustered at zero (when no move occurs), with the rest of the observations being positive (when move occurs). In either case, the commonly used ordinary least squares (OLS) regression on the entire sample will produce biased parameter estimates because the residuals do not have a mean zero. Yet, estimation of OLS with a truncated sample created by dropping observations with zero values will also yield inconsistent estimates because of potential sample selection bias (Greene 2000). A typical model used to account for the censored nature of the data is the Tobit model (Tobin 1958) which analyzes censored data in a regression context. However, the Tobit model not only restricts the exact same variables affecting the rider's move and distance decisions but also restricts the relative effects of those variables to be equal in both decisions. To relax these restrictions, this paper adopted a Heckman-style two-stage sample selection model (Heckman 1979).

## **Literature Review**

Very few studies have associated LRT service with respect to residential moving behavior. At most, Cao and Schoner (2013) investigated transportation impact of the Hiawatha LRT in the Twin Cities by exploring motivations for people moving into the LRT corridor and the socio-economic/demographic characteristics of the movers. They found that compared to non-movers (defined as residents who have lived in the Hiawatha corridor before the opening of LRT), Hiawatha movers are more likely to be well-educated, younger, employed, and renters. There are no significant differences in household size, income, share of female, and number of cars per driver between movers and non-movers in the Hiawatha corridor. Important relocating factors considered by movers include easy access to transit station, job accessibility, affordable and high quality living unit, and safe neighborhood.

Instead of focusing on LRT, a number of studies investigated the impact of a new rail service on residential location choice. Orchieng et al. (2002) conducted a survey analysis

of the residential relocations that followed the New Jersey Transit Midtown Direct rail improvement. The survey shows that improvement caused 8% of the riders to relocate their residence within five months of the start of the service. In addition, 15% of the respondents stated they would relocate if they could save between 31-45 minutes on their one-way commute. In the decision to choose residence, the most important attributes include accessibility to work, school, and services, neighborhood security, real estate values, and traffic congestion concerns. The above survey was also used by Holguin-Veras et al. (2002) to analyze the impacts of transit accessibility changes upon residential location choice. The result indicates that the decision to change residence is affected by overall accessibility (for all modes) rather than transit accessibility only. Riders also take into account the overall characteristics of the commute (including travel time, comfort, convenience, among other) while making residential choice decisions, as opposed to the sole consideration of travel time.

Using travel data from the San Francisco Bay Area, Cervero and Duncan (2008) employed a nested logit model to jointly estimate the traveler's decision to live near a train station (within ½ mile radius) and the decision to routinely take rail to work. The research reveals that station-area residents are most likely to be lower-income households, younger individuals, non-traditional households (traditional household is defined as two adults between the ages of 25 and 54 years with at least one child), Asian-Americans and Hispanics, and have lower levels of auto ownership. Also instrumental in the choice to live near transit is job accessibility via both transit and highway networks (Note: Many rail stations in the Bay Area have good highway access.) Lund (2006) conducted a survey of households who moved to transit-oriented developments (TODs) that were served by rail (including light, heavy, and commuter rail system in the San Francisco Bay Area, Los Angeles, and San Diego). Only about one-third of respondents reported access to transit as one of the most important reasons for choosing to live in a TOD. Other important reasons include cost of housing, quality of the living environment, and access to shops and services. Those who reported that their choice of residence location was motivated in part by access to transit were more likely to use transit than those who did not. Olaru, et al. (2011) also evaluated how households consider TOD characteristics in their residential location decisions with regard to new Mandurah railway line stations in Western Australia. Whereas "stage of life" (a continuous evolution of family structure) is found to influence household relocation decisions significantly, other important factors include affordable and safe locations, proximity to shops, services, and transit, and prospects for increased real estate prices.

From the review of literature, it appears that previous studies largely concentrated on the exploration of the characteristics of station-area residents and their main reasons for moving to the station area. Therefore, the primary contribution of this paper is to extend the current mobility literature by describing not only a rider's likelihood to move as a result of the new LRT service but also the movers' orientation toward their residence distances to LRT stations. Although recent studies (Holguin-Veras, et al. 2002; Lund 2006; Olaru 2011) claimed that access to transit was not always viewed as the most important factors for people to make relocation decisions, the HBLRT survey

showed that 65% of the riders had actually moved in response to the new LRT service. This result suggests that improved transit access can be a dominant relocation factor in certain conditions, particularly when the new LRT system can easily connect to existing transportation infrastructure (Note: The HBLRT provides connections to PATH subway, ferry service, bus stops, suburban commuter rail, and park and ride lots). This multi-modal connectivity would provide transit riders more options for more destinations in general and improves job accessibility to lower and midtown Manhattan and Newark in particular.

## **Model Description**

In this paper, two mobility (move and distance) choices were analyzed using the Heckman's sample selection model (Heckman 1979). The model contains two latent variable equations: the move equation (first stage) and distance equation (second stage). The move equation uses a binary probit model to determine whether or not a rider would make a residential move as a result of the new LRT service. For those riders who have decided to move, the distance equation uses a linear regression to calculate how far a rider would live away from the nearest LRT station. To control for sample selectivity, the second stage regression appends the inverse Mills ratio (Greene 2000) calculated from the linear predictions of the probit model as an additional independent variable. A more detailed description of the model structure and its estimation method is included in Appendix 1.

## **Data Description**

The primary rider data were drawn from New Jersey Transit's 2008 platform survey conducted at seven stations along the 5.5-mile northern segment of the Hudson-Bergen Light Rail Transit (HBLRT) between Tonnelle Avenue in Jersey City and the Hoboken Terminal in Hoboken. As shown in Figure 1, the seven stations include Tonnelle Avenue, Bergenline Avenue, Port Imperial, Lincoln Harbor, 9th Street, 2nd Street, and Hoboken Terminal. The survey technique was to distribute self-administered questionnaires at every station and ask passengers to return them by mail or to drop boxes at stations. The target population consisted of typical weekday boarding passengers traveling southbound from the selected stations during the time period 4:45 AM – 4:00 PM. Surveying in one direction is to reduce the likelihood of asking the same riders to fill in the questionnaire twice. (Note: Because questionnaires were distributed to southbound passengers only, the Hoboken Terminal would act as a collection station.) The sampling strategy adopted a census approach, in which an attempt was made to distribute a survey form to every boarding passenger. If the survey form was refused, it was put to the side not to be reused. The final count of the refused survey forms and field survey records confirm the total boarding volumes at each station and the total number of forms that were handed out.

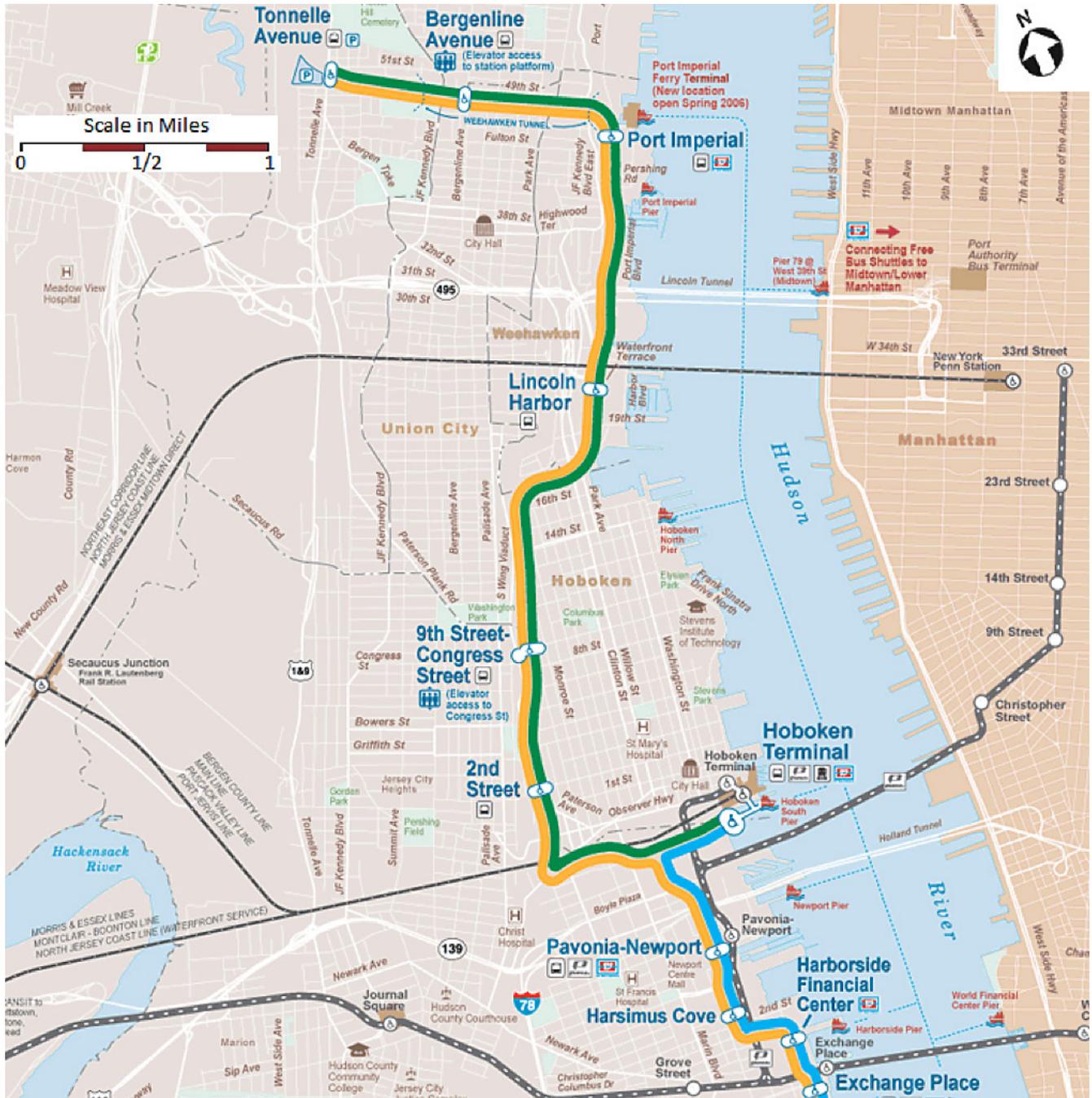


FIGURE 1. The Northern Segment of the HBLRT

The survey consisted of 39 questions covering the information on riders' travel patterns, customer satisfaction ratings, workplace locations, residential relocations, and socio-economic and demographic (SED) characteristics. To understand how the HBLRT service affects riders' moving decisions, one of the questions is that "If you moved within the last five years, how important was the availability of the HBLRT service in your decision to move?" Four categories are available for respondents to mark the importance of the HBLRT, including "Very important", "Somewhat important", "Somewhat unimportant", and "Not important at all". Survey results show that 46% of the movers considered LRT "very important" and 24% of the movers consider LRT "somewhat important". The above responding results also indirectly revealed that approximately 65% of the riders had made the decision to move after the HBLRT service began operation. Because riders who had made the decision to move also provided their new address information, geocoding of these addresses into a GIS-based database system allowed us to calculate the Manhattan distance, in miles, from a rider's residence to LRT station.

After the responses were reviewed and errors were corrected, the final data collection effort resulted in 1,023 usable survey forms, achieving a response rate of approximately 19%. The distribution of usable survey forms among the six stations surveyed matches the distributions of the total boarding volumes, implying a representative sample of passengers entering the northern segment of the HBLRT system. The 1,023 responses would achieve a system-wide confidence level of 90% with a  $\pm 8.5\%$  margin of error for the survey results. A 8.5% margin of error means that if the survey was repeated, survey responses would lie  $\pm 8.5\%$  of the initial survey responses 90 percent of the time, for each survey question. Or, we are 90% confident that the actual values are  $\pm 8.5\%$  of the reported values.

## **Variable Specification**

The Heckman's sample selection model contains two latent variable equations: the move and distance equations. The dependent variable of move equation is a binary variable that has the value of one if a rider's move was observed and zero otherwise. The dependent variable of distance equation is the amount of road network distance (in miles) from a rider's residence to the nearest LRT station. The distance variable would have the value of 0 if a rider did not make a move. For model estimation, the natural log of the distance was used to reduce the effects of the skewed nature of the distance variable as well as guaranteeing the nonnegativity of the distance predictions. To control for sample selection bias in the distance equation, it is suggested there should be at least one independent variable that appears in the move equation but not the distance equation (Wooldridge 2009).

The main category of the independent variables used in the analysis is individual (rider) and household socio-economic and demographic (SED) characteristics, including household income, age of individual rider, number of children less than 18 years old, race, ethnicity, automobile ownership, and tenure (own or rent). These variables were used to account for the taste variations in mobility choices between different

household groups, as well as capturing the effects of the life-cycle stage of a household and its orientation toward changing housing needs.

Whereas most of the SED variables are self-explanatory, household income was defined as a set of three income categories (from low, middle, to high) to encompass approximately 30, 40, and 30 percent respectively of all households. To allow for a nonlinear relation between income and dependent variables, each income category was specified as a dummy variable with value of one if a rider belongs to that category and 0 otherwise. Only low- and middle-income categories were entered into the model due to the collinearity problem in model estimation. Similarly, the age variable also was classified by three age categories:  $\text{age} < 34$ ,  $35 \leq \text{age} \leq 55$ , and  $\text{age} > 55$ . Young-age ( $\text{age} < 34$ ) category was used as the base category and, therefore, only middle- and old-age were entered into the model.

In addition to SED variables, access mode, workplace location, and length of time using LRT were also included in the models. The choice of these variables was based either on the findings from previous research, or on the statistical tests of the parameter estimates. Finally, constants were included to capture the mean effects of unobserved or unmeasured variables that affected rider relocation decisions, so that the random component of the model would be forced to have zero mean. Because rider information on relocation and transportation costs was not available, the constants would represent largely the costs and, therefore, their signs were expected to be negative.

## **Estimation Results**

The parameter estimation results of the model are presented in Table 1. All the coefficient estimates for variables about which hypotheses were formulated have the expected signs. A positive coefficient for a variable in the move equation means that the likelihood of move increases with an increase in the value of that variable. Similarly, a positive coefficient for a variable in the distance equation means that the distance from a residential location to its nearest transit station increases with an increase in the value of that variable. Most individual coefficient estimates are significantly (at the 95% confidence or more) different from zero, implying that the selected variables can adequately explain variations in the move/distance data. Because the correlation coefficient  $\rho$  is statistically significant, the error term  $e_i$  in Equation 1 (in Appendix 1) is correlated with the error term  $\varepsilon_i$  in Equation 2. Therefore, unobserved factors driving riders' moving also have impacts on riders' distance choice, indicating that the effort to construct a Heckman sample selection model is worthwhile.

In the remainder of this section, model estimation results are discussed with emphasis on the effect of the independent variables on rider's mobility decisions. Based on the magnitude of the coefficient estimates, independent variables of age, income, walk access and renter in the move equation have relatively large effects on the rider's decision to move or stay. Similarly, independent variables of age, income, walk access, bus access and automobile ownership in the distance equation have relatively large effects on the rider's decision to determine the residence distance to LRT station.

As shown in Table 1, the constants are negative and highly significant, indicating diminishing utility for riders to make a residential relocation decision. The age variables show that younger adults exhibit a greater tendency than middle-aged and senior riders to move after the introduction of a new LRT service. This is possibly due to the fact that the life-cycle stage of the younger people experience more changes due to marriage and divorce. Furthermore, younger people experience more moves because of lower residential moving costs and more opportunities to get a better residence offer. Regarding residence distance to LRT station, the distance equation shows that senior riders are more likely than younger riders to live nearby LRT stations.

The value of income coefficients in the move equation is smaller with respect to high/middle-income riders and, therefore, increase in income would decrease the probability of a rider making a residential move as a result of the new LRT service. The magnitude of income coefficients decreases with increasing household income but at a decreasing rate, indicating a nonlinear effect of income on rider's move propensity. With regard to residence distance to LRT station, high-income riders are more likely than low-income riders to live closer to LRT stations once they have decided to make a residential relocation. This seems logical because of increases in the market value of residential properties located near transit stations. Note that the variable of middle-income was purposely not included in the distance equation in order to account for selection bias (Wooldridge 2009).

LRT riders who access stations by walking are more likely to move than riders who access by cars or buses. Furthermore, riders with bus access have a lower probability than riders with car access of making a move decision. This is possibly because that bus-access riders already have a direct bus connection to a LRT station and may lose bus availability and convenience after residential relocations. Regarding the residence distance to LRT stations, riders with walking or bus access show a greater propensity to live closer to LRT stations. In contrast, car-access riders are prone to live farther to LRT stations.

The presence of children has a negative effect on the rider's moving propensity and, therefore, as the number of children increases, the riders are less likely to move. Lower mobility rates are expected because the presence of children usually increases the cost of moving. In respect to residence distance to LRT station, a positive children coefficient indicates that the riders with more children are prone to live farther to LRT stations. Riders with a low car ownership level are prone to make a residential move as a result of a new LRT service. This is presumably because the riders with fewer cars are more likely captive to transit and, therefore, would like to move and live in housing within the walking distance of LRT station. For those riders who have decided to move, a lower car ownership level would be associated with riders living in housing near stations than in housing farther away.

Riders with rented housing are more likely to move than homeowners. For the renters higher mobility rates are expected, as the moving costs are lower. Regarding the residence distance to LRT stations, homeowners show a greater propensity than renters to live closer to the stations. This would suggest that the proposed housing developments near the stations should focus on sales rather than rental markets. Hispanic riders have the lower propensity for moves than other ethnic riders. However,

Hispanic movers are prone to like proximity to the stations than other ethnic movers. Similarly, white riders are less likely than nonwhite riders to move, but white movers are more likely than nonwhite movers to live near the stations.

Riders who work in New York City are prone to move than riders who work in New Jersey. For those riders who have decided to move, New York City workers are more likely than New Jersey workers to live further away from the stations. This result indicates that the LRT's regional connectivity might reduce total travel times of the New York City workers and, therefore, they could choose a new residence further away from the stations. Finally, length of time using the LRT service has a negative effect on the rider's moving propensity, implying a lower probability of relocating as the overall length of time reported by riders using LRT service increases. This result may suggest that the earlier the riders decide to move, the more likely they would view transit access (rather than housing cost or quality) as the top reason for moving. Once the riders have decided to move, the longer the riders use LRT service, the less likely they would live further away from LRT stations.

**TABLE 1.**  
Move and Distance Model  
Estimation Results

Model	Move		Distance	
Variable	Coefficient	t-statistic	Coefficient	t-statistic
Constant	-2.42	-7.15	-6.89	-6.93
High income	-0.97	-3.53	-1.57	-1.86
Middle income	-0.89	-4.48	--	--
Age > 55	-2.82	-8.31	-3.97	-3.54
35 < Age < 55	-0.77	-3.93	-1.10	-2.03
Children < 18 years	-0.46	-1.59	0.83	1.65
Automobile ownership	-0.29	-4.78	1.22	3.12
Walk access	0.75	4.64	-3.67	-4.61
Bus access	-0.46	-3.31	-2.36	-5.66
Renter	0.49	4.47	1.05	3.34
Hispanic/Latino	-0.39	-3.94	-1.01	-2.65
White	-0.21	-2.01	-0.69	-2.03
Work in New York City	0.33	2.89	0.68	1.81
Length of time using LRT service	-0.18	-4.24	-0.38	-2.65
$\sigma_{\epsilon}$	--	--	0.91	2.94
$\rho$	--	--	0.69	3.62
Number of Observations	1,023			
Log likelihood at convergence	-1,899			

## Conclusion and Discussion

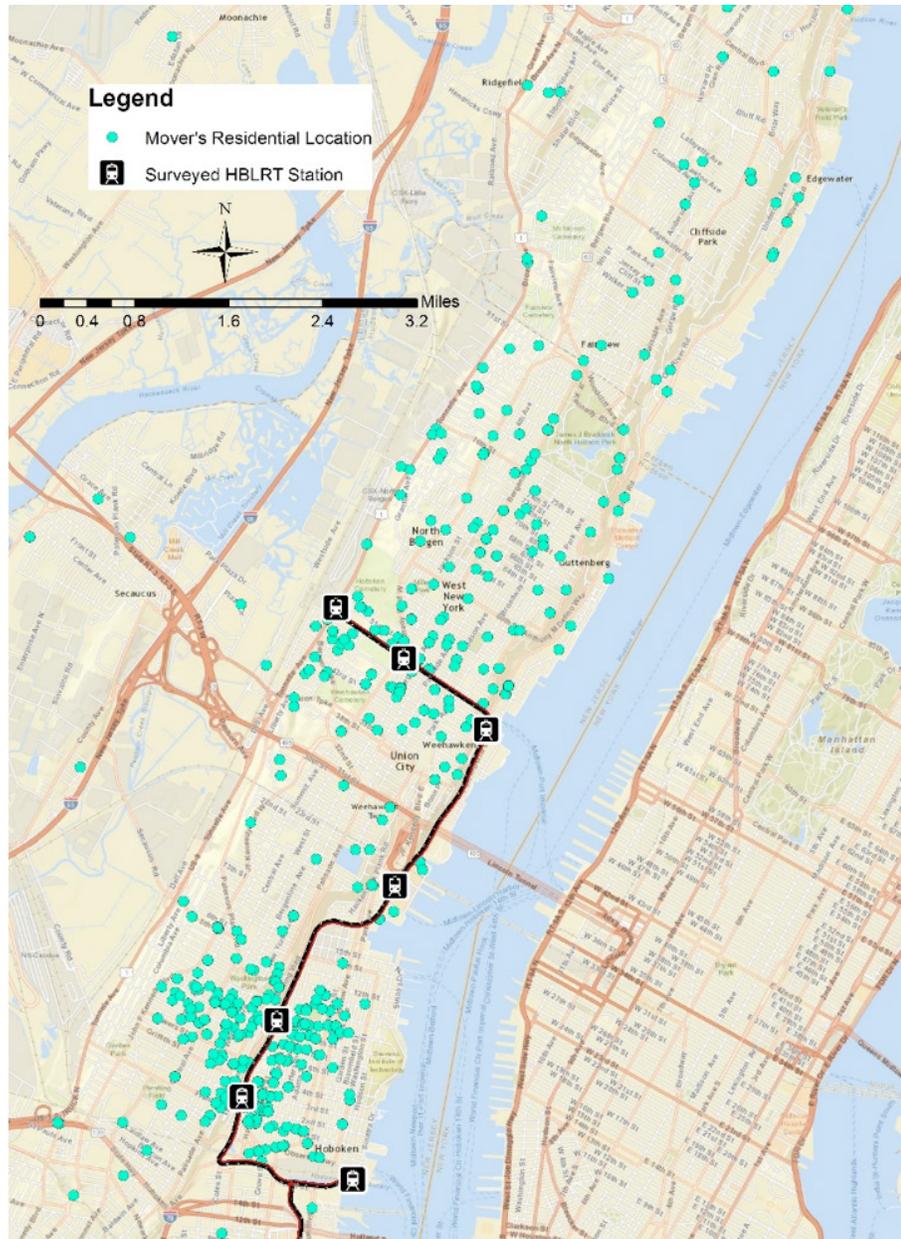
Using the HBLRT rider survey data, this paper conducted a cross-sectional analysis of the residential relocating decisions for riders who have been riding the HBLRT for a period of time. This transit investment opens new residential opportunities because the HBLRT provides the regional connectivity that makes it a viable transportation

option for relocated residents to enjoy easy access to a diverse array of destinations. Two relocating decisions were analyzed, including whether or not a rider would make a residential move as a result of the new LRT service (move choice) and, having decided to move, how far a rider would live away from the nearest LRT station (distance choice). Using socio-economic and demographic (SED) data from the survey, the Heckman sample selection model was used to explore the characteristics of the riders most likely to move and the movers' orientation toward their residence distances to LRT stations. Note that although the current model is largely employed to explain past riders' decisions, it can be used to forecast relocating decisions of the riders who are starting to use the HBLRT or a new LRT system if their SED data are available. This relocation forecasting would provide useful information to help planners/developers plan out housing and amenities for relocating residents, as explained later.

Regarding the effects of demographic characteristics on the move and distance choices, the model reveals that senior riders, Hispanic origin riders, white riders, or riders with more children are less likely to move as a result of the new LRT service. For those riders who have actually moved in response to the new service, senior riders, Hispanic origin riders, white riders, or riders with few children are prone to live near LRT stations. In respect to the effects of socio-economic characteristics, the model indicates that riders with high household income, home ownership, and high auto ownership level have the lower propensity for moves. For those riders who have decided to move, high-income riders, homeowners, or riders with low auto ownership level like proximity to LRT stations. The model also indicates that the riders who access stations by walking are more likely to move than riders who access by cars or buses. Walking-access riders show a greater propensity (than car- and bus-access riders) to live closer to LRT stations. Compared to riders who work in New Jersey, riders who work in New York City are prone to move and incline to live further away from the stations. Finally, the longer riders have used LRT services, the less likely they would move and more likely they would live near LRT stations.

By knowing the characteristics of riders most likely to move and the movers' residence distances to LRT stations, both planners and developers can identify areas where housing growth associated with specific characteristics of riders will occur, and plan for these areas to provide housing and amenities for relocating residents. More specifically, planners would understand the role and influence of LRT on residential moving behavior and the spatial redistribution of movers' residences. According to survey results, the movers' residential locations are shown as green circles in Figure 2. Although our model is not a residential location model, it can estimate how far a mover would live away from a LRT station. The model shows that 34% of the movers choose to live within 0.5 mile road network distance of a LRT station; 63% live within 1 mile of a station; 77% live within 2 miles of a station; 87% live within 5 miles of a station; and 13% live beyond 5 miles of a station. The average distance from movers' residences to LRT stations is 2.4 miles and the median is 0.7 mile. This information would allow planners to undertake planning efforts at specific locations by providing housing and amenities for relocating riders. Planning efforts may include the land use plan related to the type, location, density and intensity of development, and the transportation plan related to designation of a system of motorized and non-motorized travel that supports the land use plan.

**FIGURE 2.**  
Distribution of Mover  
Residential Locations



On the one hand, survey results show that 56%, 24%, and 20% of the movers would access stations by walking, buses, and cars, respectively. On the other hand, the model indicates that the average distances between movers' home and stations are different along the mover's choice of access mode. The average distances for walking, bus, and car movers are 0.51, 2.53, and 7.59 miles, respectively. As described before, approximately 34% and 63% of the movers choose to live within 0.5 and 1 mile road distance of a LRT station. All of this information suggests that planners consider the integration of land use and transit development (e.g., developing transit-friendly neighborhoods to support transit infrastructure) within 0.7 or 0.8 mile road distance of a LRT station.

By knowing the characteristics of movers and their residence distances to LRT stations, developers would know the challenges and opportunities for developing specific housing types within various road distances of a LRT station to serve different income levels, family structure, and ownership. For those areas nearby the stations, for example, useful modeling information for the developers include:

- Whereas high/middle-income riders and homeowners are less likely to move, they are associated with those riders who would live closer to the stations provided they decide to move. In contrast, low-income riders and renters have a higher propensity to relocate, but are associated with those riders who would live farther to the stations. This implies that the greater share of new housing units being created near stations should be planned for sales (rather than rental) markets and for high- and middle-income households.
- Senior riders are less prone to relocate, but are associated with those riders who would like proximity to stations once they decide to change residence. In similar conditions, riders with more children and higher automobile ownership levels are less likely to move and more prone to live farther to stations. This result indicates that the new housing units (and their associated amenities) being created near stations should be designed for accommodating smaller household size, senior residents, and households with lower auto ownership level.

Therefore, instead of developing a diversity of housing types and price ranges in all areas, each area's development should be tailored to meet the housing demand associated with specific characteristics of the movers, including ages, races, income levels, ownership, and family status. Once knowing the characteristics of the movers in an area, planners and developers can develop policies and plans to offset the rising cost of land, keep housing affordable, and yet allow high quality housing to be built.

## **References**

- Cao, J. and J. Schoner. 2013. *Transportation Impact of Transitways: A Case Study of Hiawatha Light Rail Transit in Minneapolis*. CTS Report 13-13, Center for Transportation Studies, University of Minnesota.
- Cervero, R. and M. Duncan. 2008. *Residential Self Selection and Rail Commuting: A Nested Logit Analysis*, Working Paper, University of California at Berkeley.
- Greene, W. H. 2000. *Econometrics Analysis*. Prentice-Hall, Upper Saddle River, NJ.
- Heckman, J. J. 1979. Sample Selection Bias as a Specification Error. *Econometrica*, 47: 153-161.
- Holguin-Veras, J., R. E. Paaswell, W. Aviles, V. Ochieng, and C. McKnight. 2002. *Multivariate Analysis of the Relationship between Transit Accessibility and Residential Location Choice*. Working paper No. 11, University Transportation Research Center, New York, February.

Lund, H. 2006. Reasons for Living in a Transit-Oriented Development, and Associated Transit Use. *Journal of the American Planning Association*, Vol. 72, 357-366.

Ochieng, V., H. S. Levinson, R. E. Paaswell, J. H.-Veras, and C. McKnight. 2002. *The Residential Relocation Impacts of Midtown Direct A Descriptive Analysis*. Working paper No. 10, University Transportation Research Center, New York.

Olaru, D., B. Smith, and J.H.E. Taplin. 2011. Residential Location and Transit-Oriented Development in a New Rail Corridor. *Transportation Research*, 45A: 219-237.

Tobin, J. 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 26: 24-36.

Wooldridge, J. M. *Econometrics*. India Edit. Michigan: Cengage Learning, 2009, pp. 459-461.

## Appendix 1. Model Structure

The Heckman's sample selection model contains two latent variable equations: the move equation (decision to move or stay) and distance equation (residence distance to the nearest LRT station). They are defined as follows:

$$d_i^* = \alpha' z_i + e_i, \quad i = 1, 2, \dots, N \tag{1}$$

$$d_i = 1, \text{ if } d_i^* > 0; \quad d_i = 0, \text{ if } d_i^* \leq 0$$

$$y_i^* = \beta' x_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad n < N \tag{2}$$

$$y_i = y_i^*, \text{ if } d_i = 1; \quad y_i = 0, \text{ if } d_i = 0$$

where  $d_i^*$  and  $y_i^*$  are the latent dependent variables for individual rider  $i$ ;  $N$  and  $n$  are the numbers of observations in the full and selected samples, respectively (Note: The selected sample includes uncensored observations ( $d_i = 1$ ) only);  $z_i$  and  $x_i$  are vectors of observed explanatory variables;  $\alpha$  and  $\beta$  are vectors of parameters to be estimated; and  $e_i$  and  $\varepsilon_i$  are the error terms assumed to be correlated through a correlation coefficient  $\rho$ , independently of  $(z_i, x_i)$ , and bivariate normally distributed, with zero mean and unknown covariance matrix.

$$\begin{bmatrix} e_i \\ \varepsilon_i \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & \sigma_\varepsilon^2 \end{bmatrix} \right) \tag{3}$$

where  $\sigma_\varepsilon^2$  is normalized to 1 for identification purposes. The latent variable  $d_i^*$  in Equation 1 represents the rider  $i$ 's propensity to make a residential move as a result of the new LRT service, and it is continuous and unobserved. Instead, we observe the binary realization  $d_i$ , which takes the value  $d_i = 1$  when  $d_i^* > 0$  (move occurs), and  $d_i = 0$  when  $d_i^* \leq 0$  (no move occurs). The latent variable  $y_i^*$  in Equation 2 contains the Manhattan distance information (i.e., the shortest road distance (in mile) from residence to the nearest LRT station) of those riders for which the realization variable  $d_i = 1$ , that is  $y_i = y_i^*$  when  $y_i^* > 0$ , otherwise their information is unobservable ( $y_i = 0$ ).

The standard procedure to estimation of the model is the Heckman two-step estimator. The resultant estimates are consistent but not asymptotically efficient under the normality assumption. More efficient estimates can be obtained using the full information maximum likelihood (FIML) approach (Greene 2000). The log-likelihood for the full sample of observations is given as follows:

$$\ln L = \sum_{d_i=0} \ln \Phi(-\alpha' z_i) + \sum_{d_i=1} \left[ -\ln \sigma_\varepsilon + \ln \phi\left(\frac{y_i - \beta' x_i}{\sigma_\varepsilon}\right) + \ln \Phi\left(\frac{\alpha' z_i + \rho \sigma_\varepsilon^{-1} (y_i - \beta' x_i)}{\sqrt{1 - \rho^2}}\right) \right] \quad (4)$$

Maximization of this function produces simultaneous estimation of the parameters of both the move and distance equations (i.e.,  $\alpha$ ,  $\beta$ ,  $\rho$ , and  $\sigma_\varepsilon$ ). Compared with the Heckman procedure, the FIML estimator is computationally intensive and difficult to numerically find the maximum values. Hence, good starting values already close to the true parameter values become very important. In this study, the final values of the Heckman procedure were used as the starting values for the FIML procedure.

## About the Authors

**YOU-LIAN CHU** ([you-lian.chu@parsons.com](mailto:you-lian.chu@parsons.com)) is currently a senior supervising transportation planner/engineer with Parsons in New York City. He received a Master of Architecture from Yale University, a Master of City and Regional Planning from Harvard University, and his Ph.D. in Transportation from the University of Pennsylvania.

**YI DENG** ([yi.deng2@nyct.com](mailto:yi.deng2@nyct.com)) is a principal transportation planner in the Division of Operational Planning at MTA New York City Transit. She received her Ph.D. in civil engineering from New Jersey Institute of Technology and a Master degree in Transportation from Tongji University. She is a current member of the TRB Committee on Automated Transit Systems.

**RONGFANG (RACHEL) LIU** ([RLiu@njit.edu](mailto:RLiu@njit.edu)) is a professor in the Department of Civil and Environmental Engineering, New Jersey Institute of Technology (NJIT). Her research and consultation interests are in the areas of multimodal and intermodal transportation planning. She has published extensively in various journals and conference proceedings and her new book is *Automated Transit: Planning, Operation, and Applications* (Wiley, 2016), the first in-depth, comprehensive analysis on automated and/or driverless transit.