Factors influencing bus-stop level ridership in the Arnhem Nijmegen City Region

TRAIL Research School, November 2014

Kasper Kerkman, Karel Martens, Henk Meurs
Institute for Management Research, Radboud University Nijmegen, the Netherlands

ABSTRACT

This paper explores the factors that influence transit ridership at the level of single transit stops for the local and regional bus transit system in the region of Arnhem-Nijmegen in the Netherlands. Direct transit ridership modeling was used to simultaneously explore the influence of spatial, population, and network characteristics on bus-stop level ridership (number of passengers boarding and alighting transit vehicles at a particular transit stop). Separate cross-sectional multiple regression models were built for two periods of time: March 2012 and March 2013. Between these periods, regional transit supply had changed considerably due to the start of a new tender period. The outcomes of these models are compared with a panel-data model, which relates the changes in ridership between both periods to the changes in transit supply characteristics. The results show that actual (short-term) ridership effects of supply changes are lower than regular cross-sectional direct ridership models forecast.

Keywords: direct transit ridership modeling; bus-stop level; cross-sectional regression; fixed-effects panel data

1. INTRODUCTION

Transit systems support a broad range of goals that include provision of mobility to the disadvantaged, access to employment or attraction centers, air pollution reduction, congestion reduction and the promotion of economic development. Understanding the factors that influence transit ridership is very important to achieve these goals and increase transit market potential. Previous research shows the effects of demographic, socio-economic and spatial variables, as well as the transit level of service, and the relative performance of transit compared to other transportation modes, on transit usage. Especially the density of population and employment, the diversity in land use, and the accessibility of the transit system are seen as important aspects. However, the literature also reveals some shortcomings. Not many studies take spatial, population, and network characteristics into account simultaneously. Furthermore, most research is conducted on an aggregate, usually system or region wide, level. To capture the micro-scale effects analyses at disaggregated scale are necessary. For instance, variation in land use characteristics between different parts of routes can best be captured and examined on a station level. In addition, changes in networks and stops might have specific effects on ridership, which cannot be captured by conventional cross-sectional analysis since it may take time to adjust to new networks.

Direct transit ridership modeling is a method capable of exploring the influence of the characteristics of the transit-stop and its surroundings, as well as transit supply at the stop level, on the number of passengers boarding and alighting transit vehicles at a particular transit stop (i.e. transit-stop level ridership). Usually, cross-sectional multiple regression method is used with station-based ridership as the dependent variable, and different characteristics of the station environment and transit level of service as independent variables (Cervero, 2006). Most direct transit ridership studies concern rail (Cervero, 2006; Kuby, Barranda, & Upchurch, 2004; Sohn & Shim, 2010) and bus rapid transit systems (Cervero, Murakami, & Miller, 2010; Currie & Delbosc, 2013; Estupiñán & Rodríguez, 2008). Recently, also a few direct ridership models have been developed for local bus transit systems in the United States (Chu, 2004; Dill, Schlossberg, Ma, & Meyer, 2013; Pulugurtha & Agurla, 2012; Ryan & Frank, 2009). Given the substantial differences between US and European cities, in terms of the quality of the transit systems, the general attitude towards transit, the role of bicycles in the transport system, and spatial patterns, it is relevant to examine the influence of land-use characteristics and transit supply on transit ridership in a European setting.

In the current study, we analyze the factors that influence transit ridership at the level of single transit stops, for the bus-based transit system of the region of Arnhem-Nijmegen in the Netherlands. We use cross-sectional (CS) multiple regression models with bus-stop ridership explained by spatial, socio-demographic, and transit supply characteristics. We will compare the outcomes with panel-data models for bus ridership, relating changes in ridership to changes in network characteristics. This comparison reveals the forecasting quality of CS models; in case the coefficients are about equal, CS model will predict changes relatively well; if coefficients deviate, it is interesting to find the reasons for this and to consider methods to improve the models.

In the remainder of this paper, we describe the methodology and data used to build the direct ridership models of the Arnhem Nijmegen City Region bus services, followed by a description of the results and conclusions.

2. REVIEW OF DETERMINANTS

The literature provides a rich understanding of the factors influencing transit ridership. In what follows, we provide a brief overview of the most important interrelationships discussed in the literature.

2.1 Potential demand

High transit ridership levels are only possible if there is enough (potential) demand for transit. Land use characteristics have a large influence on the potential demand. The most important land use factors in this context are often known under the headings Density, Diversity (land use mixture) and Design (e.g., provision of convenient sidewalks and other pedestrian amenities that encourage walking) (Cervero & Kockelman, 1997). Some researchers introduce a fourth D – accessibility to concentrated regional Destinations. Of these four D-factors, density in the transit corridor and activity intensity at the destination end of the corridor are the main quantifiable land use variables. For many station-level transit ridership studies, these factors form an essential part of the research. Especially population and employment density are commonly included, and show a positive effect on the number of boardings. Land use mix has been examined in several studies, but there is no consensus on its effect on boardings (Dill et al., 2013). Some studies demonstrated a small but statistically significant positive effect of pedestrian-friendly urban design on ridership (Chu, 2004; Ryan & Frank, 2009).

Other important potential demand factors are the socio-demographics. These factors encompass variables that describe the characteristics of the population living around transit stops. Especially income and car ownership are associated with transit use. Higher average (household) income and a higher level of car ownership around transit stations tend to result in lower transit ridership. Other socio-demographic variables included in some direct ridership studies, mostly concentrating on bus services, are ethnicity, age, gender and education level (Chu, 2004; Dill et al., 2013; Ryan & Frank, 2009), although these factors often have a rather small or non-significant effect on ridership.

2.2 Transit supply

Factors relating to the supply of transit services include travel time, costs, reliability, and comfort. Especially in demand forecasting studies, the quality of service is seen as an important variable (Polat, 2012). According to Cervero, "[a measure of transit service level] *is often the strongest single predictor of ridership*" (Cervero, 2006, p290).

In general, waiting or travel time and price are seen as the most important service level variables. When analyzing ridership at station-level, price and travel time seem not very usable, as they are mostly dependent on the entire transit system or the specific trip made and are therefore not addressable to station locations. Therefore, the average waiting time at a station seems to be the most relevant transit supply indicator. As a proxy for this, the most commonly used indicator of transit service-level used in station-level ridership studies is service frequency. A higher frequency of transit services indicates a shorter average waiting time, and shows consequently a positive effect on total ridership.

The local transit supply depends to a large extent on the potential demand. The endogeneity of transit supply characteristics in direct demand models may lead to the overestimation of the relative importance of transit supply for transit ridership (Taylor, Miller, Iseki, & Fink, 2009). This is because operators take demand into consideration when deciding upon supply characteristics. The endogeneity is especially large between potential demand and service frequency. This made researchers decide to exclude service frequency from their station-level ridership studies (Gutiérrez, Cardozo, & García-Palomares, 2011), or to use another proxy for service level (Cardozo, García-Palomares, & Gutiérrez, 2012). Estupiñán & Rodríguez (2008) used a two-equations simultaneous model to account for the interaction between transit supply and demand in their study on station boardings in Bogotá's BRT. Others did include service frequency directly in their models because of its strong predictive power of ridership (Blainey, 2010; Cervero, 2006; Guerra, Cervero, & Tischler, 2012). They did find indications of some bias in the results, as the frequency elasticity found seemed rather high. The other way around, however, as denser areas usually receive more frequent transit services, the exclusion of service frequency from ridership models likely results in exaggeration of the influence of urban densities on ridership (Cervero, 2006). Furthermore, service frequencies at stop level are the result of frequencies of routes. This means that stops 'en route' may receive high service frequencies despite low demand. When supply variables are excluded, the ridership effects of these high frequencies are falsely attributed to demand.

In the current research, we use panel data to correct for this endogeneity problem. Proper panel data analysis may reduce the endogeneity problem, especially when using a so-called fixed-effects model (Hsiao, 2003). This is because the bias is related to correlation of the error-terms in the demand and supply models. In panel data models these errors are decomposed into a time-invariant component and a variable, time-varying, component. It may be expected that a major proportion of the correlation between the errors can be attributed to the fixed time-invariant component, since operators cannot respond to quick time-varying specific circumstances. In case a fixed-effects model is used with differencing as estimation method, as is done in this paper, an important part of this correlation between the error terms, causing the biases, is cancelled. Conversely, a comparison between parameters in the CS-models and the panel data models may reveal the variables for which endogeneity is a problem.

2.3 Match between transit supply and land use patterns

Besides the relative influence of land-use characteristics, socio-demographics and transit supply on ridership, the quality of the match between the transit system and land use patterns is an important aspect for successful transit systems. Often there is, however, a disconnection between transport and land-use patterns (Badoe & Miller, 2000; Martens, 2000; Newman & Kenworthy, 1991). For example, residential and employment land uses have decentralized while metropolitan areas grew in spatial size and population. New patterns of urban form came into being, known as “polycentric” or “multi-nucleic” city regions, in which multiple locations attract development, employment, and population. At the same time, public transport systems remain largely oriented towards the traditional urban cores (downtowns or CBDs). This spatial mismatch between transit networks and spatial characteristics of urban regions is one of the causes of the declining share of the urban public transportation market because monocentric systems do not fit the travel patterns within polycentric cities. It is expected that if a transit system is well adjusted to the local land use patterns, this will yield higher transit usage.

3. DATA AND METHODOLOGY

3.1 Study area: Arnhem Nijmegen City Region

The area studied in this research is the Arnhem Nijmegen City Region (“Stadsregio Arnhem Nijmegen”) in the eastern part of The Netherlands. This is a collaboration of 20 municipalities, with a total size of more than 1000 square kilometers, and almost 750 000 inhabitants. The cities Nijmegen (168 000 inhabitants) and Arnhem (151 000 inhabitants), located only about 15 kilometers apart, form the core of the region. The local and regional public transportation within the region is mainly bus based. The bus transit services are operated by Hermes, a private company operating a tender for 10 years (using the brand name “Breng”), offering a more or less integrated network of bus services throughout the entire region.

The transit services within the region are largely focused on the city centers and main train stations of Arnhem and Nijmegen: all bus routes serve at least one of these locations. Local buses connect neighborhoods within the cities of Arnhem and Nijmegen, while regional buses connect the surrounding towns to the city centers and main train stations of Arnhem and Nijmegen. In December 2012, a new tender period of the bus transit in the region has started. The current transit network at that time was considered as reasonably successful and effective, and was supposed to form the basis of the services in the new tender period. However, to increase the efficiency of operations (higher cost coverage) and service levels, considerable changes in routes and timetables were made at this time.

3.2 Data

In this study, we analyze (change in) bus-stop ridership data over the months March 2012 (before the new tender commenced) and March 2013, for all bus-stops in the region. Bus-stop ridership is calculated as the sum of passengers entering and leaving a bus per stop as our dependent variable. The information about the ridership of each bus-stop is provided by the bus operator, Hermes. The data used are the total number of boardings and alightings per bus-stop within the region, on working days in the months March 2012 and March 2013. The data are based on the use of chip cards by passengers. It is estimated that 90% of all passengers used a chip card to travel on the bus system, so the chip card data give a reasonably accurate representation of the actual number of travelers.

The independent variables in our models are based on the results of previous direct ridership modeling studies (e.g., Cervero, 2006; Chu, 2004; Dill et al., 2013; Pulugurtha & Agurla, 2012; Ryan & Frank, 2009). The selected variables describe the local potential demand for transit and the transit supply at each bus-stop. Potential demand variables include socio-economic and demographic characteristics (household income, age levels), the number of potential travelers (sum of inhabitants, employers, students and train travelers), and the percent of the land used for a specific function (socio-cultural facilities, residential, or agricultural). These variables were calculated using the “key figures on districts and neighborhood 2012” and “bodemgebruik 2008” (land use) data files, published by the Dutch Central Bureau of Statistics (CBS, www.cbs.nl). ESRI ArcGIS 10.0 was used to define a 400 meter (.25 mile) straight line buffer around each bus-stop in the region, representing the area being served by a specific bus-stop. Using aerial interpolation, calculating the percentage of each neighborhood covered by a buffer and assigning the same share of socio-economic and land-use variables of the neighborhood to the buffer, the values for bus-stop service areas are determined.

The selected transit supply variables describe the local transit supply at each bus-stop, and include service levels and bus-stop characteristics. The General Transit Feed Specification (GTFS) data file (obtained from gtfs.ovapi.nl) is used as data source. This file contains the routes and timetables of the bus services in the entire region. A complicating factor is the endogeneity of potential demand and transit supply. Because operators take demand into consideration when deciding upon supply characteristics, including both potential demand and supply variables in the models may lead to the overestimation of the relative importance of transit supply for transit ridership. The exclusion of transit supply from the analysis, however, likely results in exaggeration of the influence of urban densities on ridership, as denser area’s usually receive more frequent transit services (Cervero, 2006). In the current research, we use panel data to correct for this endogeneity problem.

In addition to the variables describing the potential demand and transit supply individually, the interaction between these variables is likely to have a combined effect on ridership as well. Expected high ridership at locations with high supply might not be reached if the local potential demand is very low. In the same way, high potential demand will not result in high ridership if there is little supply. Therefore, an interaction variable is added to the analysis that describes the match between potential demand and local transit supply. This variable compares the ratio between the number of buses at a stop (stop frequency) and the number of potential travelers, with the average ratio of all bus-stops in the region (the average ratio is considered as optimal). This relationship is translated into an index between 0 and 1, where 1 indicates an optimal supply/demand match and 0 an extreme shortage or surplus of supply compared to the potential demand.

Although high correlation coefficients exist between several selected independent variables, all of them are below the danger level 0.7 (Clark and Hosking, 1986). A full overview of the variables used in the analysis and their data sources are presented in Table 1.

Table 1. Variables and data sources

Variable	Description	Data source	Calculation method
Ridership (dependent variable)			
Bus-stop ridership (logarithm)	Daily average number of passengers boarding and alighting a bus at each specific bus-stop	Hermes (bus operator)	Smartcard data
Potential demand (independent variables)			
Potential travelers (logarithm)	Sum of number of inhabitants, employees, students, and train travelers within service area	CBS	Aerial interpolation using 400m buffer
Income [x €1.000]	Average household income	CBS	
Percent elderly	Percent of the population aged 65 years or older	CBS	
Distance to urban centre [km]	Straight line distance between the bus-stop and the city centre of Arnhem or Nijmegen	GTFS	Euclidean distance using ArcGIS
LU: Residential	Part of the service area with residential land use	CBS	Share of 400m buffer area
LU: Agriculture	Part of the service area with agricultural land use	CBS	
LU: Socio-cultural facilities	Part of the service area with socio-cultural facilities as land use (education, healthcare, cultural)	CBS	
Transit supply (independent variables)			
Stop frequency (logarithm)	Average daily number of buses scheduled to serve each specific stop	GTFS	Timetables
Directions	Number of different destinations of buses serving each bust-stop	GTFS	Timetables, routes
Frequency per direction	Average number of buses per route at each stop	GTFS	[stop frequency] / [directions]
Direct connections	Average number of other stops on the routes of buses that serve each stop	GTFS	Timetables, routes
Competitive bus-stops	Number of other bus-stops with partly overlapping service area's	GTFS	Spatial join ArcGIS using 800m buffer
Bus terminus [1/0]	Start or end stop of at least one bus-route	GTFS	Timetables, routes
Transfer stop [1/0]	A transfer to another line is possible (i.e., at least two different bus lines serve the stop)	GTFS	
Bus station [1/0]	Bus-stop located in specific physical structure or area, which can contain one or more stops	GTFS	GTFS variable
Dynamic Information [1/0]	A dynamic passenger information system is installed at the bus-stop	City region	
Benches [1/0]	Seating facilities are present at the bus-stop	City region	
Supply / demand match			
Supply/demand index	Index describing how well the supply matches the potential demand (interaction between stop frequency and potential travelers)	f=stop frequency p=potential travelers	$i = \frac{1}{1 + \left\lvert\frac{\bar{a} - a}{\bar{a}}\right\rvert}$ where: $a = \frac{f}{p}$

3.3 Analysis method

We estimate two separate regression-based direct ridership models, using bus-stop ridership data of March 2012 and March 2013, respectively, as the dependent variable. As independent variables, the selected potential demand and transit supply characteristics are used. Ordinary least squares (OLS) regression method is used to build these cross-sectional (CS) multiple regression models. Logarithmic transformations are applied to correct for the skewed distribution of the dependent variable (bus-stop ridership), and the independent variables "stop frequency" and "potential travelers". Count data models (Poisson and Negative Binomial regression with log-link) were also tested, but because the results showed large similarities with the OLS models, we decided not to present the results here.

In addition to the cross-sectional models, the changes in the variables between both periods are used to study the effects of transit supply changes on ridership levels. Recall that much of the transit routes and timetables were redesigned as part of the new tender that commenced in December 2012. We use fixed-effects panel-data regression method (Hsiao, 2003; Meurs, 1990) to regress the changes in (natural logarithm of) bus-stop ridership (y_{March} 2013 - y_{March} 2012) to the changes in transit supply (x_{March} 2013 - x_{March} 2012). So basically, instead of analyzing the influence of the original independent variables on bus-stop ridership, we now explore how the *changes in* the dependent variables between March 2012 and March 2013 influenced the *changes in* bus-stop ridership in the same period of time. As we assume that there were no significant changes in land-use and population characteristics in this relatively short period of time, no potential demand variables are included in this model. Using this method, we can explore the actual influence of changes in transit supply on ridership. This approach reduces the endogeneity problem of transit supply and potential demand.

4. RESULTS

4.1 Effects of potential demand and transit supply

Two cross-sectional OLS regression models were estimated explaining the variance in monthly bus-stop level ridership in the Arnhem Nijmegen City Region using potential demand and transit supply variables: one for March 2012, and one for March 2013. The estimates and fit-measures of both models are displayed in Table 2.

Table 2. Regression models prediction (natural logarithm of) daily bus-stop ridership

	March 2012 (N=1232)			**March 2013 (N=1284)**		
Variable	***Coefficient***	***Std. coef.***	***t-value***	***Coefficient***	***Std. coef.***	***t-value***
(intercept)	-1,135		-2.78***	-1,705		-4,34***
Potential demand						
Potential travelers (logarithm)	,131	,091	3.09***	,113	,078	2,66***
Income	-,006	-,046	-2.66***	-,008	-,060	-3,45***
Percent elderly (65+)	-,007	-,028	-1.89*	-,006	-,026	-1,75*
Distance to urban centre	-,047	-,110	-5.93***	-,026	-,061	-3,34***
LU: Residential	,894	,153	7.28***	,759	,132	6,31***
LU: Agriculture	-,598	-,093	-4.33***	-,507	-,081	-3,73***
LU: Socio-cultural facilities	1,682	,076	4.84***	1,504	,069	4,39***
Transit supply						
Stop frequency (logarithm)	,807	,394	9.92***	,861	,445	11,71***
Directions	,044	,098	4.22***	,057	,126	5,32***
Frequency per direction	,012	,147	5.10***	,012	,145	4,91***
Direct connections	-,010	-,089	-6.24***	,002	,021	1,42
Competitive bus-stops	-,011	-,045	-2.17**	-,012	-,054	-2,71***
Bus terminus	,352	,057	3.67***	,348	,058	3,71***
Transfer stop	,211	,064	2.65***	,133	,041	1,78*
Bus station	,189	,053	2.73***	,177	,050	2,73***
Dynamic information	,494	,066	4.37***	,510	,066	4,36***
Benches	,330	,100	6.70***	,313	,097	6,59***
Supply/demand match						
Supply/demand index	,270	,039	2.11**	,464	,068	3,68***
Measure of fit	*Adjusted R^2= .772*			*Adjusted R^2= .762*		

* $p<0.10$; ** $p<0.05$; *** $p<0.01$

Most of the selected independent variables in the regression models are highly significant. Both models perform well, with an adjusted R^2 of .772 and .762, respectively. This means that the selected variables together explain 77.2% and 76.2% of the variance in bus-stop ridership. The models show large similarities. The directions of most coefficients are equal in both models, and largely in line with expectations. The values of most coefficients are very similar as well. The models only differ significantly at the “direct connections” variable. The direction of this coefficient changes from negative in 2012 to positive in 2013, although this variable is statistically not significant in the 2013 model. As the changes in routes between these periods resulted in more direct routes (i.e., less deviations), the negative side-effect of having much potential destinations (deviations, less direct, slower) seem to be less dominant in 2013 compared to the positive aspects (high connectivity). However, in order to draw conclusions on this subject, more information describing the directness from each stop (e.g. stop distance, travel speeds) should be added to the models. We will take up this issue in future work.

The standardized coefficients show that bus stop frequency is the variable with the largest relative influence on bus-stop ridership. Both the total frequency at each bus-stop and the frequency per direction have a large, positive contribution. This shows the importance of waiting time on transit ridership. As the models use a logarithmic transformation for both the ridership and the stop frequency, the coefficients found for stop frequency represent the elasticity. Although these elasticity’s (.81 and .86) are within the range usually found in other studies (Evans, 2004; Guerra et al., 2012), they are higher than the typical elasticity of between .3 and .4 found by Currie & Wallis (2008). In our models, the influence of service frequency on ridership is most likely overestimated due to the endogeneity between transit supply and potential demand. As more frequent services are usually planned at locations with high potential demand, the potential demand of a location is also represented in the service frequency. This effect is even larger, due to the fact that bus services are also being planned based on ridership levels of previous years. In this way, service levels can be increased (or decreased) based on unexpected high (or low) ridership due to (unobserved) potential demand. Due to this endogeneity between potential demand and service levels one cannot interpret the models’ coefficients of stop frequency as a direct effect of service frequency only.

Most transit supply variables included in the models have a positive and significant influence on ridership. Only the number of competitive bus-stops variable has, as expected, a negative influence on ridership. From the potential demand variables, the share of the service area with residential land use has relatively the largest, positive, influence on ridership. The coefficients for the potential travelers variable (number of inhabitants, employees, students, and train travelers) are lower than expected. This might be partly because of the inclusion of the land-use types as separate variables. Another aspect is the endogeneity between potential demand and transit supply as described before, resulting in an underestimation of the importance of potential demand.

The demand/supply index, an interaction variable between stop frequency and potential travelers indicating how good the supply matches the potential demand at each bus-stop, is significant and positive in both models. This shows that, next to the individual effects of stop frequency and potential travelers on bus-stop ridership, the balance between these variables has an additional influence on ridership. That is, if the stop frequency is too low or too high compared to the number of potential travelers, this has an extra negative effect on ridership. This shows the importance of careful planning of transit services, as both too much or too little services have a negative effect on efficiency.

4.2 Effects of transit supply changes on ridership

In addition to the cross-sectional models described above, the differences in the variables between both periods are used to study the effects of transit supply changes (due to the new tender that commenced in December 2012) on ridership. As we assume that land-use and population characteristics did not significantly change in this short period of time, no potential demand variables are included in this model. The coefficients of the resulting fixed-effects panel-data regression model (Table 3) show how changes in transit supply variables contribute to ridership change.

Table 3. Fixed-effects panel data regression model of changes in natural logarithm of daily bus-stop ridership between March 2012 and March 2013 (N=1210)

Variable	*Coefficient*	*Standardized coefficient*	*t-value*
(intercept)	.076		4.96***
Change in transit supply			
Stop frequency (logarithm)	.407	.296	7.53***
Directions	.028	.072	1.98**
Frequency per direction	.003	.096	1.99**
Direct connections	.004	.115	4.36***
Competitive bus-stops	-.004	-.014	-.52
Bus terminus	.176	.066	2.45**
Transfer stop	.117	.106	2.58**
Change in supply/demand match			
Supply/demand index	.446	.100	3.76***
Measure of fit		*Adjusted R^2= .169*	

** p<0.10; ** p<0.05; *** p<0.01*

Most of the variables describing the (change in) transit supply are highly significant. Only the change in number of competitive bus-stops is not significant. This is likely due to the low variation in this variable. Although the transit supply has significantly changed in this period, changes in the number and locations of bus-stops were very limited.

The adjusted R^2 of the panel-data model is .169, what means that 16.9% of the variance of change in ridership can be explained by the change in transit supply between March 2012 and March 2013. This shows that a large part of the change in ridership is caused by other factors than transit supply. The fact that bus-stops with no or little changes in transit supply were included in the regression also has a negative effect on the fit of the model, as variations in ridership for these stops cannot be explained by the independent variables. The significant and positive value of the intercept of the model shows an overall increase in ridership in the studied period, controlling for changes in transit supply. This is remarkable, as changes in services often initially result in decrease of patronage. This indicates that there is a positive trend in bus usage between the years, after controlling for changes in supply. This may reflect success of marketing and other factors not included in the models.

Similar to the cross-sectional models, the stop frequency has the largest individual influence on the changes in ridership. The value of its coefficient is, however, about half of the coefficients found in the cross-sectional models. In fact, the values of most coefficients in the panel data model are roughly half of the corresponding variables in the cross-section models. This is likely to be (at least partly) caused by the fact that travelers need time to adjust to changes in transit supply. It is reasonable to assume that in the first year after modifications of transit services, about half of the expected long-term effects will be reached. Moreover, the lower values of the coefficients in the panel data model might also be caused by an overestimation of the importance of transit supply due to the endogeneity between supply and potential demand in the cross-sectional models. As this endogeneity does not influence the results of the panel data model, the coefficients in this model are lower and more realistic. The influence of changes in the supply/demand index on changes in ridership is similar to its influence according to the cross-sectional models. This indicates that a better balance between supply and potential demand has a relatively large and direct effect on ridership.

5. CONCLUSIONS AND DISCUSSION

In this study, we analyzed the factors that influence transit ridership at the level of single transit stops, for the region of Arnhem-Nijmegen in the Netherlands. Direct transit ridership modeling was applied using OLS regression method, to analyze the importance of potential demand and supply on the number of passengers boarding and alighting buses at each specific bus-stop. Besides separate cross-section models for two periods of time (March 2012 and March 2013), we also build a panel-data model relating the changes in ridership between these periods to the changes in transit supply.

5.1 Cross-sectional models (March 2012 and March 2013)

The adjusted R^2 of the models for March 2012 and March 2013 (.772 and .762, respectively) show that bus-stop ridership can be explained for a large part using land use and socio-demographic information about the bus-stop surroundings, combined with information about the level of transit supply at the specific bus-stop. The influence of most variables is in line with the literature and intuition, suggesting that the variables describe what they are intended to capture. A large influence of transit supply on ridership is shown, with stop frequency as most important factor. As described before, however, the influence of stop frequency on ridership should not be considered straightforward, as endogeneity exists between transit supply and potential demand. The endogeneity most likely causes an overestimation of the importance of transit supply variables, and an underestimation of the influence of potential demand on ridership levels. An additional positive effect of a good balance between transit supply and potential demand is shown, stressing the importance of well-designed and planned transit services.

5.2 Panel data model (changes between March 2012 and March 2013)

Fixed-effects panel-data regression method is used to explore the influence of changes in transit supply between March 2012 and March 2013, on the changes in bus-stop ridership levels between these periods. The model shows the significant influence of most transit supply variables on ridership changes. There are large similarities between the cross-sectional and the panel data models in terms of the relative importance of most independent variables, but the values of most coefficients in the panel data model are only about half of the corresponding variables in the cross-sectional models. This is most likely due to the adjustment time (potential) travelers need to get used to the changes in transit services, as it usually takes a few years until the long-term effects of supply changes are fully visible. To further examine this adjustment effect, we aim to build new panel data models using data from later years, with the expectation that a model regressing the changes between March 2012 and March 2015 will be very similar to the cross-sectional models.

The large number of significant variables, and the similarities between the cross-sectional and the panel data models, show the potential of the use of panel data and the fixed-effects regression model to explore the influence of transit supply on ridership. Furthermore, the models can be used to simulate the effects of transit supply modification on local ridership levels in a fast and efficient way. However, the current model describes the relationships in the transit system of the Arnhem Nijmegen city region only. It cannot be generalized for use in other regions and transit systems. To be able to do so, more research is needed, and models of different transit systems should be built and compared.

5.3 Value of direct ridership models

Direct ridership modeling is an interesting method, capable of exploring the effects of many different characteristics on transit ridership, and simulating the effects of planning and policy implementations. However, there are a couple of concerns.

Firstly, the endogeneity between potential demand and supply has an influence on the results of modeling exercises. As transit supply levels are often partly determined by potential demand and previous ridership levels, the influence of transit supply on ridership is likely to be overestimated, while the importance of potential demand is often underestimated. An often used method to avoid this problem is the exclusion of either service frequency, or even supply variables in general. This means, however, excluding variables with high descriptive value. Moreover, as transit supply is often high where the potential demand is high, excluding transit supply variables likely results in overestimation of the influence of potential demand variables on ridership. A more optimal solution to the endogeneity problem would be to use more advanced estimation models which can account for the simultaneity between transit supply and potential demand, like two-stage regression method (Taylor et al., 2009). In the current study, we introduced the use of panel data models to reduce the endogeneity problem. The use of the fixed-effects model with differencing as estimation method reduces the endogeneity problem. This is because the bias is related to correlation of the error-terms. If the permanent unobserved errors are differenced out, a major part of the bias will be eliminated as well.

Secondly, the value of the results of direct ridership models cannot be generalized and used in other situations. There is currently too little knowledge on the differences in the micro-scale effects of the relationships between potential demand, transit supply, and ridership in different (geographical, cultural, political) situations. Specific policies on for example parking prices and car costs have large effects on transit ridership, and can vary widely between regions. As there is no knowledge on how this relates to the effects found in direct ridership models, results from direct ridership modeling are mainly regional specific.

ACKNOWLEDGEMENTS

This research was supported by the BRENG Kenniscentrum. We would like to thank Hermes (the local bus operator) for providing the ridership data that made this research possible.

REFERENCES

Badoe, D., & Miller, E. (2000). Transportation–land-use interaction: empirical findings in North America, and their implications for modeling. *Transportation Research Part D: Transport and Environment*, *5*.

Blainey, S. (2010). Trip end models of local rail demand in England and Wales. *Journal of Transport Geography*, *18*(1), 153–165. doi:10.1016/j.jtrangeo.2008.11.002

Cardozo, O. D., García-Palomares, J. C., & Gutiérrez, J. (2012). Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Applied Geography*, *34*, 548–558. doi:10.1016/j.apgeog.2012.01.005

Cervero, R. (2006). Alternative Approaches to Modeling the Travel-Demand Impacts of Smart Growth. *Journal of the American Planning Association*, *72*(3), 285–295.

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219.

Cervero, R., Murakami, J., & Miller, M. (2010). Direct ridership model of bus rapid transit in Los Angeles County, California. *Transportation Research Record: Journal of the Transportation Research Board*, *2145*(1), 1–7. doi:10.3141/2145-01

Chu, X. (2004). Ridership models at the stop level. *National Center for Transit Research, Florida Department of Transportation, Report No. BC137-31*.

Currie, G., & Delbosc, A. (2013). Exploring Comparative Ridership Drivers of Bus Rapid Transit and Light Rail Transit Routes. *Journal of Public Transportation*, *16*(2), 47–65.

Currie, G., & Wallis, I. (2008). Effective ways to grow urban bus markets – a synthesis of evidence. *Journal of Transport Geography*, *16*(6), 419–429. doi:10.1016/j.jtrangeo.2008.04.007

Dill, J., Schlossberg, M., Ma, L., & Meyer, C. (2013). Predicting Transit Ridership at Stop Level: Role of Service and Urban Form. *Transportation Research Board 92nd Annual Meeting*.

Estupiñán, N., & Rodríguez, D. a. (2008). The relationship between urban form and station boardings for Bogotá's BRT. *Transportation Research Part A: Policy and Practice*, *42*(2), 296–306.

Evans, J. (2004). *TCRP Report 95: Traveler Response to Transportation System Changes, Chapter 9: Transit Scheduling and Frequency*. Washington, D.C.

Guerra, E., Cervero, R., & Tischler, D. (2012). Half-Mile Circle. *Transportation Research Record: Journal of the Transportation Research Board*, *2276*(-1), 101–109. doi:10.3141/2276-12

Gutiérrez, J., Cardozo, O. D., & García-Palomares, J. C. (2011). Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. *Journal of Transport Geography*, *19*(6), 1081–1092. doi:10.1016/j.jtrangeo.2011.05.004

Hsiao, C. (2003). *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Kuby, M., Barranda, A., & Upchurch, C. (2004). Factors influencing light-rail station boardings in the United States. *Transportation Research Part A: Policy and Practice*, *38*(3), 223–247.

Martens, C. J. C. M. (2000). *Debatteren over mobiliteit: over de rationaliteit van het ruimtelijk mobiliteitsbeleid*. University of Nijmegen/Thela Thesis Publishers, Nijmegen/Amsterdam.

Meurs, H. (1990). Trip generation models with permanent unobserved effects. *Transportation Research Part B: Methodological*, *24*(2), 145–158. doi:10.1016/0191-2615(90)90025-T

Newman, P. W. G., & Kenworthy, J. R. (1991). Transport and urban form in thirty-two of the world's principal cities. *Transport Reviews*, *11*(3), 249–272.

Polat, C. (2012). The Demand Determinants for Urban Public Transport Services: A Review of the Literature. *Journal of Applied Sciences*, *12*(12).

Pulugurtha, S. S., & Agurla, M. (2012). Assessment of Models to Estimate Bus-Stop Level Transit Ridership using Spatial Modeling Methods. *Journal of Public Transportation*, *15*(1), 33–52.

Ryan, S., & Frank, L. (2009). Pedestrian environments and transit ridership. *Journal of Public Transportation*, 39–57.

Sohn, K., & Shim, H. (2010). Factors generating boardings at metro stations in the Seoul metropolitan area. *Cities*, *27*(5), 358–368. doi:10.1016/j.cities.2010.05.001

Taylor, B. D., Miller, D., Iseki, H., & Fink, C. (2009). Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas. *Transportation Research Part A: Policy and Practice*, *43*(1), 60–77. doi:10.1016/j.tra.2008.06.007