First-mile accessibility and public transport usage in Dutch peripheral regions

Master Thesis

MSc. Spatial, Transport and Environmental Economics

Vrije Universiteit Amsterdam

29 June 2024

Author: Lorenzo Tas

Student number: 2671236

E-mail address: l.j.tas@student.vu.nl

Supervisor: prof. dr. E.T. Verhoef

ABSTRACT

This study investigates to what extent the first-mile accessibility to a transit stop influences the use of public transport by travelers. For this purpose, data has been collected about smartcard check-ins at bus stops located in the Dutch provinces Drenthe and Groningen. In these peripheral regions, the bus network has seen a significant degradation since 2018. By use of GIS-software, the average distance to a bus stop from residents of the service area has been computed for the years 2015 up to 2022, forming the measure of first-mile accessibility in this study. Compared to previous studies, this analysis implements a broader definition of service areas and uses a more advanced measurement of distance. An OLS regression model with year-fixed effects is estimated to quantify the effect of distance to a bus stop on average yearly check-ins per resident of the service area. The elasticity of check-ins per capita with respect to the distance to a bus stop is found to be -0.77, suggesting a relatively strong relationship of distance to the bus stop on bus usage. Understanding how sensitive travelers are to the distance to a bus stop, could help public transport planners in designing more optimal routes.

Keywords: public transport, bus stops, travel behavior, first-mile, accessibility, distance

Table of contents

1. Introduction

Public transport plays a significant role in enhancing mobility in the Netherlands, by providing access to work, education, and other amenities for those who can't drive or can't afford a private vehicle. Furthermore, public transport is a sustainable form of transportation and it plays a role in reducing externalities caused by congestion (Rijksoverheid, 2024). The availability of transportation also plays a role in the prevention of social exclusion, for example for physically disadvantaged groups in society. (Currie & Delbosc, 2010). However, good accessibility by public transport is not a given in each corner of the country. Since 2018, provinces located in the north of the Netherlands have seen a significant degradation of public transport. To illustrate, in the province Groningen a number of 208 bus stops have been taken out of service, which is a decrease of 15 percent (RTV Noord, 2023). A recent analysis by Planbureau voor de Leefomgeving (PBL) on accessibility to amenities in the Netherlands suggests that the limited availability of public transport in suburban and rural areas forms the core of accessibility problems in the country, with consequences more severe than the increasing road congestion in the country (Bastiaansen & Breedijk, 2022).

Accessibility to amenities by public transport also entails the accessibility to transit stops, which is the first sequence of the journey. In transport literature, this aspect has received considerable attention, but is often focused on urban areas only (Ewing & Cervero, 2010; Mulley et al., 2018; Aman & Smith-Colin, 2020). This study investigates how much the first-mile accessibility influences ridership under the local population, focusing on two rural provinces in the Netherlands; Drenthe and Groningen.

A better understanding about what it is worth to bring public transport closer to people, can result in more optimal routing and bus stop placement by planners, contributing to the attractiveness of public transport. The central question of this research is: "*To what extent does the distance to the nearest bus stop affect bus usage by local residents?"*. To answer this question, GIS-software is used to compute the average distance to a bus stop from its service area. Data on the total number of check-ins at bus stops in Drenthe and Groningen forms the basis for bus usage of residents. Regression analysis will estimate the size of the relationship between the distance to a stop and bus usage, in the form of an elasticity.

The remaining part of the paper is organized as follows: Chapter 2 includes a review of literature about the analysis of public transport demand and its determinants. Chapter 3 contains a detailed description of the data used and how this data is processed, and specifies the model to be estimated. Chapter 4 presents the estimated results about the relationship between distance to a bus stop and bus usage, and includes checks for robustness. Finally, Chapter 5 contains a conclusion and devotes attention to the interpretation of the main result. Additionally, this chapter discusses limitations of this study and proposes opportunities for further research.

2. Literature review

2.1. Common methods in analyzing demand

In the transport literature a variety of methods are used to analyze the influence of service characteristics on the tendency of people to use public transport, or to use a specific type of public transport if multiple alternatives are present. A distinction can be made between studies that use stated preference (SP) techniques, and studies based on revealed preference (RP) that use data on observed travel behavior.

2.1.1. Choice experiments

Stated preference studies commonly take the form of discrete choice experiments. Borsje et al. (2023) have conducted a discrete choice survey in The Netherlands to obtain insight about respondents' preferences towards eight different attributes of Bus Rapid Transit (BRT), including service and vehicle characteristics. Results have been obtained through estimating a Multinominal Logit model. Bronsvoort et al. (2021) have carried out a between-mode SP survey targeted to bus users in Dutch rural areas, comparing conventional bus services with flexible alternatives, such as demand-responsive transport. For this last alternative, attributes like booking time and travel time uncertainty have been included in their choice experiment. To account for correlations between alternatives and heterogeneity in taste, they estimated Nested Logit (NL) and Mixed Logit models. Mulley et al. (2018) used a stated choice experiment to obtain insight about the trade-off between access distance and service frequency in various cities around the world. Also the effect of crowding levels in vehicles was included. They made use of an error component model that allows for the panel structure of the data, unlike a NL model (Mulley et al., 2018).

Bourgeat (2015) used a combined RP and SP approach to investigate participants' attitude towards bus service characteristics. Participants were first asked to provide information about perceived characteristics of their last bus trip. Subsequently, they were asked to indicate their preference towards varying service configuration sets, and how likely a certain configuration could replace the trip they initially described. Through this approach, SP elasticities could be compared to 'real-life' elasticities from RP models to ensure the external validity of bus choice elasticities (Bourgeat, 2015).

2.1.2. Demand models and endogeneity

The current study will take a RP approach using data on actual bus passenger demand and different explanatory variables. The use of actual data would allow for a more realistic estimation of this study's variable of interest, distance to bus stop. In contrast, the SP experiment by Mulley et al. (2018) uses only four different possible values for access distance. A RP setup also prevents possible hypothetical bias. This section gives an overview of causal analyses of public transport demand in empirical studies that use actual data.

A wide range of studies have been written about the determinants of public transport demand. Taylor & Fink (2003) and Holmgren (2007) provide an overview of explanatory factors, such as service level variables, income, car ownership, and fares. Both studies discuss potential estimation problems

arising from endogenous supply variables that explain demand. As stated by Holmgren (2007), the number of trips made likely influences the supply of vehicle-kilometres, and the latter therefore should be treated as endogenous. Simultaneous or reverse causality as a form of endogeneity in models is a problem because the error term is correlated with one or more independent variables. This results in biased OLS coefficient estimates, which could lead to drawing incorrect conclusions about the relationship between variables (Stock & Watson, 2020).

Studies have used different methods to deal with endogeneity in models that explain passenger demand. Taylor et al. (2009) analyzed factors influencing ridership using Two Stage Least Squares (2SLS) regression, to address the simultaneity between service supply and consumption. Initially, they regress a measure of supply on all its exogenous determinants, excluding demand, and obtain estimated values. Next, they regress demand on the estimated supply along with other exogenous determinants of demand. According to Chen et al. (2011), however, an instrumental variable approach would not be able to measure an effect of demand on supply if there is one. Other studies made use of the concept of Granger-causality (Holmgren, 2005; Rayaprolu & Levinson, 2024). This method allows to test for the existence of a feedback relationship, relying on the data's time dimension (Granger, 1969).

2.2. Determinants of public transport demand

This section provides an overview of findings and methodological choices made in studies that investigated the drivers of ridership in public transport. In this review the main attention is given to accessibility to transit, service frequency, and socio-demographic characteristics.

2.2.1. Accessibility to transit

Accessibility in public transport has multiple dimensions. One share of literature in this field focuses on the accessibility from residential areas to the city center or to jobs by public transport, sometimes in comparison with other modes (Rayaprolu & Levinson, 2024; Liu et al., 2022; Conwell et al., 2023; Aman & Smith-Colin, 2020). A selection of them incorporated the first stage of a journey; the access to a transit stop. This type of transit accessibility is the focus of the current research. Aman & Smith-Colin (2020) created a comprehensive public transit accessibility (CPTA) score per census tract in Dallas, USA. This measure includes connectivity to destinations as well as connectivity to the network, but also service frequency. A disadvantage is that connectivity to the network is not weighted by population density patterns within the bus stop service area. Rayaprolu & Levinson (2024) controlled for first-mile accessibility in their model that explains station-level ridership, through counting the population within 15-minute walking time to the train station. However, this approach does not exploit variation in access time or distance within the service area of the transit stop. A similar method is used in a Dutch study by Kerkman et al. (2015), who analyzed factors explaining transit ridership at stop level in the Arnhem-Nijmegen urban area. The authors have used a circular 400-meter buffer around each stop to define its service area. This study did not consider potential endogeneity between transit supply and potential demand.

Other studies have used the distance to a transit stop as a variable accessibility measure that explains transit use. Ewing & Cervero (2010) conducted a meta-analysis about associations between travel and built environment. Based on three studies that focus on urban areas in the USA, they computed a weighted average elasticity of transit use with respect to distance to nearest transit stop of -0.29, indicating that people make less use of transit when they live further away from a transit stop. Bento et al. (2003) recognized that households that plan to use public transport are likely to locate themselves close to a transit stop, forming a source of endogeneity. Mulley et al. (2018) investigated the trade-off between walking distance to a bus stop and frequency using a stated choice experiment. They find that travelers in US and UK cities are prepared to walk 370 to 475 meters further to a transit stop if headways are reduced by 10 minutes. This suggests that transit users are less sensitive to distance when a higher level of service is provided.

2.2.2. Frequency

Service frequency is an important factor related to transit use, because it has a substantial impact on total travel time, and thus on the willingness to use public transport. The underlying dynamics between demand, frequency, and waiting time has been described in the theory of Mohring (1972). Frequency influences user costs in the form of waiting time at the origin stop, as well as transfer time between vehicles (Hörcher & Tirachini, 2021; Mohring, 1972; Iseki & Taylor, 2009; Litman, 2008).

Borsje et al. (2023) concluded from their discrete choice survey about BRT in the Netherlands that frequency is the most valued attribute of bus services, followed by service hours, stop spacing, and reliability on arrival. The most preferred service formula, the 'Conventional' type, is characterized by moderate frequency and short stop spacing (Borsje et al., 2023). This last characteristic is closely related to the distance to a transit stop. Graham et al. (2009) found in a panel data analysis on metro networks that frequency has a greater impact on demand compared to other factors such as fares. Although these studies suggest an association from frequency to demand, other studies give indication for an opposite direction of effect. Börjesson et al. (2017) found that optimal level of frequency is subject to differences in demand between peak and off-peak periods during the day. Bar-Yosef et al. (2013) study the presence of vicious cycles of bus lines when demand is low. They state that low demand triggers the transit operator to cut expenses through reducing frequency. Existing travelers might then opt for alternative types of transport, which decreases demand further, et cetera. Thus with regard to analyzing frequency, a major endogeneity issue may be present.

2.2.3. Socio-demographic characteristics

The literature also indicates that socio-demographic factors influence public transport demand. With regard to age, Kerkman et al. (2015) finds that the percentage of elderly in a service area (above 65 years old) is negatively related to bus ridership. This suggests that it is mainly the younger generations that travel by bus, which is logical because they travel more frequently for work or education. Numerous studies have been written about the positive relationship between employment and public transport use,

such as Chen et al. (2011) and Taylor & Fink (2003). Average household income appears to have a negative effect on ridership (Kerkman et al., 2015; Dargay & Hanly, 2002). A reason for this is that households with a higher income are more likely to own a car (Nolan, 2010). Graham et al. (2009) raise the point that the effect of income on ridership is generally positive in studies that also include car ownership in their model. This can be explained by the observation that people are more involved in activities that require transportation as incomes increase. But if car ownership is excluded, the negative effect takes over because income then picks up the car ownership effect on transit use. Graham et al. (2009) adds that most studies have not included both of the variables income and car ownership due to multicollinearity.

2.2.4. Fares

Fares are a relevant factor in explaining public transport demand, because they directly influence the cost of using public transport. However, there is a general agreement in the literature that variation in fares impacts demand significantly less than the level of service (Graham et al., 2009; Brechan, 2017; Kain & Liu, 1999). Kerkman et al. (2015), who analyze stop-level ridership similarly to the current research, state that fares are not very useable when analyzing ridership at stop level. As the price is mostly determined by the length of a trip, it is hard to address fares to stop locations. On top of that, fares would not affect travelers with a free-travel subscription, such as students, which comprise a relatively large share of ridership.

2.3. Research gap

This literature review reveals that most studies about the factors influencing public transport demand primarily concentrate on urban areas. The current study will focus on two 'rural' provinces in the Netherlands; Drenthe and Groningen. In these areas, few travel options are available compared to the larger cities in the country, which have metro or tram systems. Leaving private types of transport aside, areas of lower density are reliant on a relatively sparse bus network. Due to large variations in residents' proximity to bus stops, the mentioned area of study could provide a better setting to analyze to what extent people are prepared to bridge large distances to arrive at a bus stop.

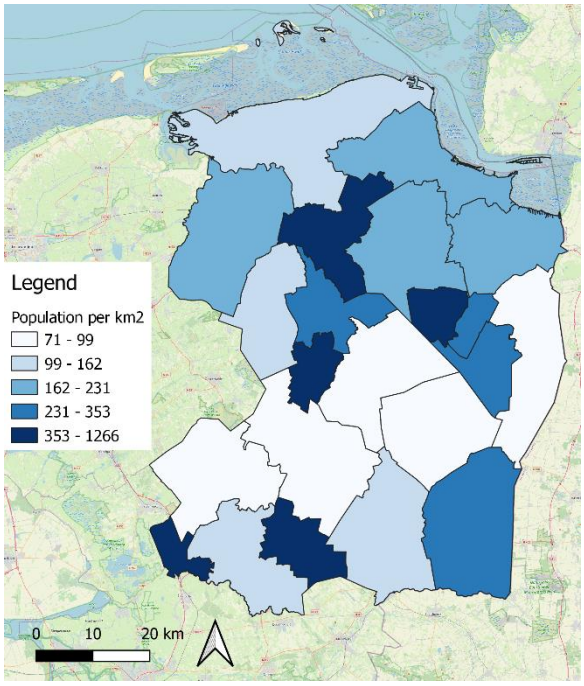
Another difference to previous studies that analyzed distance to a bus stop and demand, is that the current study will allocate the whole study area to the service area of a specific bus stop. Kerkman et al. (2015), who also analyzed stop-level ridership in a part of the Netherlands, used a 400-meter range as service areas. Rayaprolu & Levinson (2024) counted the population living within 15 minutes walking time from a station to measure accessibility. It is plausible that service areas can reach larger than the ranges these studies have chosen, especially when travelers use a bicycle or even a car for the first mile. The current study presumes variability in service area sizes, and accounts for the distribution of the population within the service area of a transit stop to obtain a measure for accessibility.

3. Data and methodology

3.1. Study area

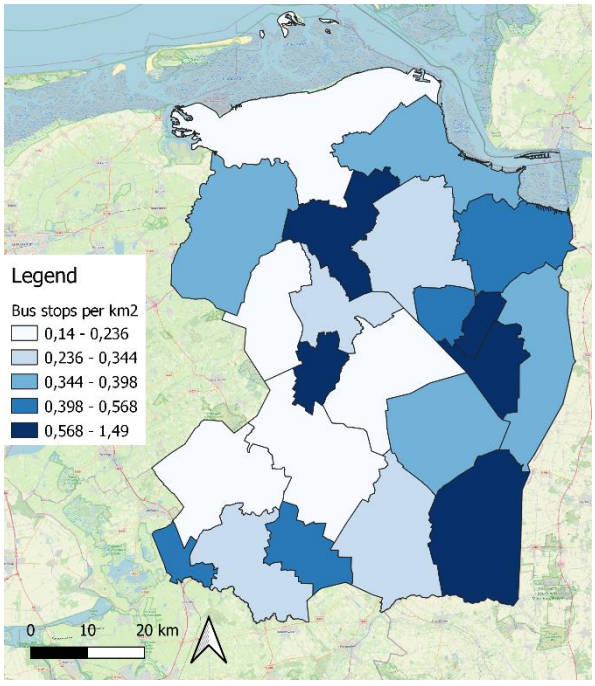
The Dutch provinces Drenthe and Groningen form the study area of this research. These provinces are situated in the northeast of the Netherlands. These areas are characterized by a relatively low population density compared to the average national density. While the Netherlands has an average population density of 529 inhabitants per square kilometer, Drenthe and Groningen have a density of 191 and 260 per square kilometer respectively (CBS, 2024-1). A convenient property of this study area is that public bus transport in Drenthe and Groningen is arranged as one entity, allowing for integrated transport data that covers both provinces.

Within these provinces, clear disparities exist in population density as well as the density of bus stops. Figure 1 illustrates the population density per municipality, and Figure 2 illustrates the density of bus stop locations, both in 2022. On the maps the municipalities are divided into equally sized quantiles. Although the color patterns differ slightly between the maps, it can be inferred that more densely populated areas generally have a higher density of bus stops. The differences in the density of bus stops over municipalities indicate that there exists significant variation in accessibility to public transport in the study area.



LAYER SOURCES: Population data, CBS (2024-1); Municipality borders, CBS & BRK (2024); Background map, OpenStreetMap.

Figure 1: *Population density per municipality in Drenthe and Groningen in 2022*



LAYER SOURCES: Bus stops 2022, OV-bureau Groningen Drenthe; Municipality borders, CBS & BRK (2024); Background map, OpenStreetMap.

Figure 2: *Density of bus stops per municipality in Drenthe and Groningen in 2022*

3.2. Study period

The study period of this research consists of the years 2015 up to and including 2022. This is the largest possible period for which data is available about public transport usage (check-ins), as well as population counts at raster level, which is used to calculate the average distance to a bus stop. Detailed information about this variable of interest can be found in section 3.5.4. The indicated period also covers the Covid-19 pandemic, which mainly took place in the years 2020 and 2021. These years are characterized by, for example, small commuting flows, increased education from home, and scaled-back public transport provision. Recorded travel behavior in the years 2020 and 2021 is therefore not representative for normal patterns. Hybrid forms of work or education also remained prevalent after these years. To account for such possible time-varying trends, time-fixed effects are used for model estimations. In order to assess whether including the years 2020, 2021, and 2022 have an impact on the results, a robustness check is performed that runs regressions excluding these years, presented in section 4.2.

3.3. Data sources

This section provides a description of data sources that contain the data for the main analysis. A distinction is made between public transport data and GIS data.

3.3.1. Public transport data

- **Excel file "Incheckers per halte per lijn", provided by OV-bureau Groningen Drenthe**
  For each the years 2014-2023 it contains the total number of check-ins by smartcard ('OV-chipkaart') at each bus stop serviced by Qbuzz, which is the bus transport provider in Drenthe and Groningen. Check-ins are also reported per system route number that services the bus stop in the relevant year. The year 2023 is incomplete and only covers the first quarter of that year.
- **Current timetables Qbuzz**
  Due to the unavailability of historic data on frequency of bus routes in the area under study, frequency per bus route is approximated using current frequencies. Current timetables, of year 2024, are provided online by the bus transport provider Qbuzz (2024).

3.3.2. GIS data

- **CBS 100 meter-raster map of the Netherlands with statistics**
  This shapefile obtained from CBS (2024-2) consists of 100-by-100-meter raster cells with statistics on, amongst other things, demographics, housing, and proximity to amenities. It only includes raster cells for which at least one value is recorded for a variable. Detailed data is available for the years 2015-2023. However, more recent years do not yet include all statistics.

  As described by Malekzadeh & Chung (2020), data with high resolution is important for the accuracy of transit accessibility measurements. The allocation of service areas in this study, described in section 3.5.4, becomes inaccurate for areas with a high density of bus stops

when lower resolution is used. A resolution of 500 by 500 meter, which CBS provides as well, would not be appropriate in this setting.

- **Transport features OpenStreetMap**

  To locate bus stops and train stations on the map in the software of QGIS, shapefiles with transport features in Drenthe and Groningen have been downloaded from Geofabrik (2024). These files include features such as bus stops, railway stations, and ferry terminals.

- **Land use OpenStreetMap**

  Shapefiles about land use types in Drenthe and Groningen have been used for the selection of the bus stops for the study sample. The selection of observations is described in section 3.7. This data has also been acquired from Geofabrik (2024).

3.4. Variable choice

As the dependent variable of this analysis, *average check-ins per capita* is the variable that measures the average degree of bus usage in a service area. Scaling check-in numbers by the local population allows for comparison across service areas and is likely to create a more natural distribution of values. The independent variable of interest *average distance to bus stop* will measure how far away the inhabitants of the service area live, on average, from the bus stop. *Total frequency* at a bus stop will be included as a control variable because service frequency is commonly taken as an important determinant of public transit demand in the literature. Additionally, *average distance to train station* is included in the analysis. This variable accounts for the possibility that inhabitants of a bus stop's service area prefer to travel directly from their residence to the train station, instead of first taking the bus. Travelers could do this because, for example, a train station is closer or they avoid transfer time.

Data on two socio-demographic control variables is collected. Firstly, the *fraction of 15 to 65 years old* is used because this group is assumed to travel more frequently due to the number of commuters and students in this group. Secondly, a variable expressing (household) income would fit in the analysis to account for the likeliness of owning one or more cars, which forms an alternative for public transport (Nolan, 2010). However, income data is not offered by CBS on a 100-meter raster level, and merely as broadly classified categorical data on 500-meter level. An explanation is that (household) income is considered as privacy-sensitive information for smaller spatial units. As an alternative for income, *average WOZ-value* is used to approximate income patterns. The WOZ-value is the property value estimated by the municipality and is used for tax purposes. Studies such as Tu et al. (2018) and Boelhouwer et al. (2004) confirm the relationship between income and house prices in the Netherlands.

3.5. Data preparation

3.5.1. Check-ins per capita

For each of the years 2015-2022, the total number of check-ins at a bus stop has been calculated as the sum of the check-ins over all bus connections that stop at this location, excluding check-ins of night services and school services. Afterwards, this number is divided by the population in the service area of

the respective bus stop, to obtain an expression for trips per person. What defines a service area is described in section 3.5.4.

3.5.2. Number of bus routes

Transfer passengers are assumed not to live in a bus stop's service area, and ideally check-in records from this type of passengers should be excluded from the analysis. However, the number of transfer check-ins is not separately recorded in the available dataset. The number of bus routes that service a bus stop is used as a selection criterion to assess which bus stops are less likely used for transfers between bus routes. This number is calculated by counting the number of check-in values, excluding the values from night and school services. In the regression analysis only bus stops with a maximum of three routes will be examined, in order to limit the number of bus stops that largely rely on transfer traffic and to still maintain a sizeable dataset. The impact of using a different number of bus routes on the results will be assessed as a robustness check in section 4.2.

3.5.3. Total frequency

Timetables of bus routes are used to deduce the frequency of each bus route included in the check-ins dataset. Using the current timetables of Qbuzz, frequency per hour is collected manually for each bus route present in the smartcard check-ins data of the year 2023, which is the most recent dataset of bus routes and bus stops. Specifically, the frequency is taken during the late afternoon on Thursdays. A frequency of zero is recorded for bus routes that only operate in the morning (3 cases), or for routes with less than two executions in the afternoon (3 cases). Subsequently, for each bus stop in the 2023 data, total frequency per hour is calculated as the sum of frequencies of bus routes for which the bus stop has a check-ins value recorded. Bus routes that are night or school services are disregarded. The resulting value for total frequency per bus stop will be used for the years 2015-2022. For bus stops in the years 2015-2022 that are not serviced anymore in 2023, the number of bus routes present in the specific year is used as an approximation for the frequency per hour in that year. It is assumed that bus routes that service a bus stop have a low frequency of one per hour before a bus stop is abolished.

3.5.4. Average distance to bus stop

The average distance of residents to a bus stop is calculated using the software of QGIS. Bus stop locations in Drenthe and Groningen have been filtered out from the transport features layer of OpenStreetMap. In this layer bus stops only have a name attribute. Because of the likeliness of multiple bus stops with the same name in the study area, municipalities are added as an attribute used to distinguish between these bus stops. Duplicates of the same name-municipality combination have been deleted, because bus stop locations on both sides of the road are included in the layer, and one location is needed. It is assumed that this brings minimal error for the calculation of distance to the bus stop.

The other dataset used for the calculation of the average distance, is the CBS 100-meter raster layer containing population per raster cell. Using the 'Distance to nearest hub' tool in QGIS, the distance from the center of each raster cell to the closest bus stop is calculated using a straight line. An important

assumption underlying this approach is that travelers start their bus trip at the closest bus stop to their residence. All raster cells with the same allocated bus stop are combinedly defined as a bus stop's service area. Figure 3 and Figure 4 provide visual examples of how service areas are created in QGIS, for both an urban part of the study area and a rural part, using data for 2022. There is a possibility that the closest bus stop at straight-line distance is not the closest bus stop over road, because of the low density of roads. Figure 4 depicts such as situation, where some raster cells on the right half of the map are closest to bus stop "A" over road, but are not allocated to this bus stop. This forms a source of inaccuracy in the service area allocation.

Average distance to a bus stop from its service area is calculated as a population-weighted average;

$$ADB_{i,t} = \frac{\sum_{j \in S_{i,t}} (p_{j,t} \times d_{i,j,t})}{\sum_{j \in S_{i,t}} p_{j,t}} \quad (1)$$

where:

- $ADB_{i,t}$ is the average distance to bus stop i in year t;
- $p_{j,t}$ is the population in 100m-raster cell j in year t;
- $d_{i,j,t}$ is the distance from cell j to bus stop i in year t;
- $S_{i,t}$ is the set of raster cells for which bus stop i is the closest bus stop in year t

Due to variation in operational bus stops and raster populations over time, a bus stop's average distance value is likely to vary over time as well.

The use of straight-line distance allows for efficient calculation. Computing the shortest distance or travel time over road would give a more realistic value for the average distance to a bus stop. This is a common way of measuring distance to transit (Ewing & Cervero, 2010). However, it is accompanied by a high computational demand in the GIS system, given the large number of raster cells for which a route needs to be calculated (for reference: 35,114 cells in 2022). This is the main reason why straight-line distance is preferred. Straight-line distance underestimates the actual distance. Because of the underestimated values, the use of straight-line distance is expected to yield more conservative results. If a significant effect is found of an understated distance on check-ins per capita, it is likely also found using actual travel distance. In rural areas the deviation is expected to be larger due to the lower density of roads compared to densely populated areas.

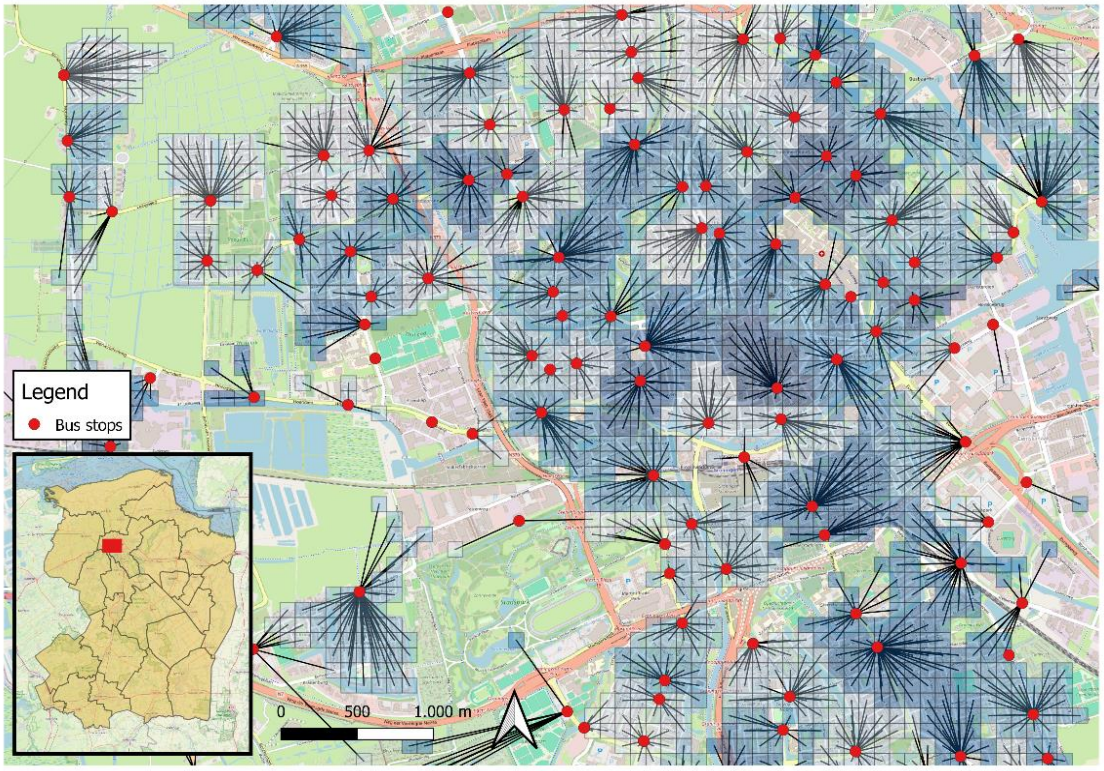
3.5.5. Average distance to train station from bus stop service area

The average distance from a bus stop's service area to a train station is calculated similarly as the distance to a bus stop. For each raster cell, the straight-line distance to the closest train station is computed using the "Distance to nearest hub" tool in QGIS. The population-weighted average distance from a bus stop's service area to a train station is described as;

$$ADT_{i,t} = \frac{\sum_{j \in S_{i,t}} (p_{j,t} \times d_{j,t})}{\sum_{j \in S_{i,t}} p_{j,t}} \quad (2)$$

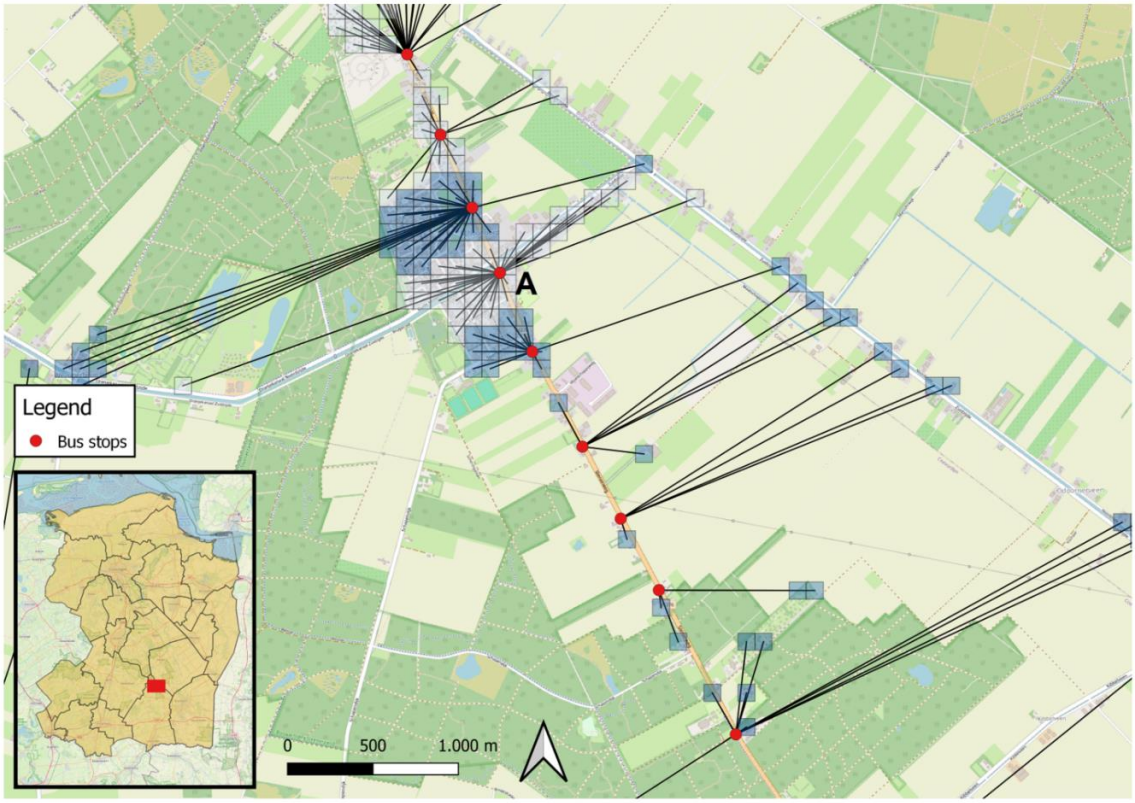
where:

- $ADT_{i,t}$ is the average distance to a train station from the service area of bus stop i in year t;
- $p_{j,t}$ is the population in 100m-raster cell j in year t;
- $d_{j,t}$ is the distance from cell j to the closest train station in year t;
- $S_{i,t}$ is the set of raster cells for which bus stop i is the closest bus stop in year t



LAYER SOURCES: Bus stops, Geofabrik (2024); Raster cells, CBS (2024-2); Background map, OpenStreetMap.

Figure 3: *Service area allocation in the city Groningen in 2022. The 'Distance to nearest hub' tool in QGIS has allocated each raster cell to the closest bus stop. Black lines connect raster cells to the closest bus stop, and distance is measured along the line. Different shades of blue form the service area for the connected bus stop.*



LAYER SOURCES: Bus stops, Geofabrik (2024); Raster cells, CBS (2024-2); Background map, OpenStreetMap.
Figure 4: *Service area allocation around the village Schoonoord in 2022. The 'Distance to nearest hub' tool in QGIS has allocated each raster cell to the closest bus stop. Black lines connect raster cells to the closest bus stop, and distance is measured along the line. Different shades of blue form the service area for the connected bus stop.*

3.5.6. Other variables

Data for the control variables *fraction 15 to 65 years old* and *average WOZ-value* is extracted from the CBS raster dataset. The fraction of 15-65 year olds is calculated by first summing the number of 15-25 year olds, 25-45 year olds, and 45-65 year olds in a bus stop's service area. The resulting number is then divided by the population in the service area. An issue with this data is that population counts per separate age group in a given raster cell are only recorded for a minimum of 5 people in an age group. On top of that, all age records are rounded to fives. These two characteristics can lead to measurement error for the fraction of 15 to 65 year olds, especially for low-density service areas. The use of data on a higher level, such as a 500-meter raster level, could have prevented missing values for age groups. However, age data on higher levels would not be able to cover the exact same service area as the variables for distance to transit. The average WOZ-value of a bus stop's service area is calculated as the mean of the average WOZ-values observed at each 100-meter raster cell within the service area. Values of 2016 are extrapolated to 2015. Due to the potential inaccuracy of the age variable and the fact that the WOZ-value only approximates income patterns, regressions are run both with and without these controls to evaluate whether they enhance the regression model.

3.6. Initial descriptive statistics

Table 1 presents initial descriptive statistics for the variables involved in the analysis, primarily used to inspect the range of values. A total of 1,827 different bus stops are identified over the years 2015 to 2022 in the provinces Drenthe and Groningen, together accounting for 13,093 observations. For a couple of variables, a minimum value of 0 is recorded.

An explanation for zero-values for the number of routes is that bus stops have been included in the dataset that are solely serviced by night or school services in a certain year (135 obs., 40 stops). Because check-ins from these services are excluded from the check-ins count, these observations have the value 0 for the number of check-ins per stop. This subsequently causes check-ins per capita to obtain the value 0. A service area population of 0 is possible for bus stops which are not located near residential areas. Examples are bus stops near the recreational area Lauwersmeer, a lake in the north of the Groningen province. These observations are automatically excluded from model estimations because an average distance is missing. Bus stops with a frequency of 0 are bus stops which are, in 2024, only serviced in the morning or less than two times seen by a bus in the afternoon.

With regard to maximum values, a value for check-ins per capita of over 36,000 is remarkable. Figure 5 includes a histogram for check-ins per capita per year to illustrate the range of high values. Although most values are concentrated at the lower side of the spectrum, there are also values visible above 1,000. High values likely occur at bus stops where check-ins do not solely replicate journey starts by the local population. Examples are bus stops where passengers transfer, or travel to home after an activity. Another variable with a high maximum is the fraction of 15 to 65-year-olds. It is highly unlikely that 100 percent of a service area's population belongs to this age category. The same holds for 0 percent, which is the minimum value. This is caused by the inaccurate age data in service areas that primarily consist of sparsely populated raster cells.

Table 1: Initial descriptive statistics

This table presents the number of observations (N), the mean value, the standard deviation (sd), the minimum value (min), and the maximum value (max) for the listed variables.

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
stopID	13,093	915.2	526.3	1	1,827
year	13,093	2,018	2.299	2,015	2,022
check-ins per stop	13,093	9,185	109,157	0	5,058,223
service area population	13,071	629.7	812.9	0	12,680
check-ins per capita	13,071	34.28	570.7	0	36,216
av. distance to bus stop (m)	13,071	504.77	543.3	18.88	6,491
total frequency	13,093	2.164	3.165	0	68
routes	13,093	1.703	1.803	0	43
av. distance to station (m)	13,071	6,427	4,874	180.9	22,938
fraction 15 to 65	13,071	0.387	0.211	0	1
av. WOZ-value (x €1000)	11,056	206.6	76.89	24	866.3

3.7. Selecting observations

In order to analyze how bus stop usage per capita is affected by the average distance from a service area to its bus stop, ideally the number of recorded check-ins should be in line with starts of a journey from travelers' living area. For this purpose, a selection of observation is made according to several criteria, which is expected to limit the number of extremely high values for check-ins per capita. An observation may apply to multiple of these criteria:

- *The bus stop is serviced by a maximum of 3 bus routes.* If a relatively high number of bus routes service a bus stop, the bus stop is more likely to be used for transfers. The maximum number 3 is a personal choice. In section 4.2, robustness checks are performed to assess the effect of choosing a different number on the results. A number of 1,310 observations that do not apply to this criterium are dropped.
- *The bus stop is not located inside an industrial, commercial, or retail area.* This criterium is assessed in the QGIS software using the land use layer. These three land use categories are selected because bus stops located in these type of areas are likely to record many check-ins from travelers leaving work, as well as from those leaving leisure activities like shopping. An additional 186 observations that are located in the three land use types are dropped.
- *The bus stop is not named "Station"*. Bus stops located at a train station are likely to record many check-ins from travelers that transfer from train to bus, as opposed to people living around that bus stop. An additional 73 observations that have the bus stop name "Station" are dropped.

On top of the measures taken to limit high values for check-ins per capita, one additional selection criterium is used. Observations that record a bus stop with zero routes are dropped as well. These bus stop locations are unserved by regular lines (other than night and school services) and are not regarded as an option for travelers. Another 134 observations are excluded for this reason. However, the involved bus stops were included in the allocation of service areas. This likely caused an underestimation of service areas sizes for bus stops in the vicinity of those with zero routes.

Figure 6 includes a histogram for check-ins per capita after applying the selection criteria. Compared to figure 5, the highest values now do not reach as high, with a maximum value of around 3,600. The selection of observations as described above has the intended effect. Still a number of values above 100 check-ins per capita remain, which is notably high given that it represents an average value for the whole service area population. After a manual investigation of bus stops with a high value for check-ins per capita, logical explanations can be given for high values. Examples of such bus stops are: bus stops at hospitals which are not assigned a land use type, bus stops in the vicinity of businesses with a small service area population, bus stops at P+R locations, and bus stops at event locations. As it appeared inevitable within the capabilities that observations of this sort remained in the dataset, a check for robustness will be done, running a regression without the 5% highest values for check-ins per capita. The robustness analysis is detailed in section 4.2.

3.8. Adjusted descriptive statistics

Table 2 displays descriptive statistics of the relevant variables after the selection measures have been applied. Somewhat more than 1,000 observations have been left out after selection. Focusing on the outcome variable check-ins per capita, the minimum and the mean have changed next to the maximum value. The minimum value is non-zero because observations without a regular service have been excluded. Low values near the minimum of 0.000288 can occur in cases where a large service area is allocated to a hardly used bus stop. The mean has decreased from approximately 34 as initially reported in Table 1, to a value of 10 after selection, which is considered more realistic. The descriptive statistics for the variable of interest average distance to bus stop have remained stable and show acceptable values.

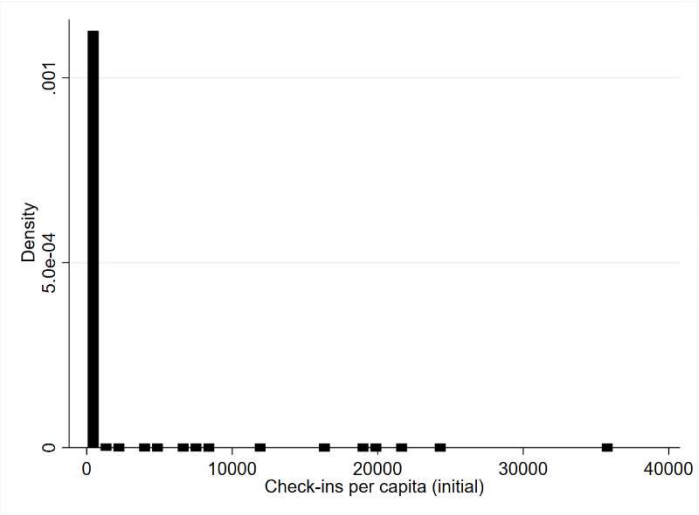


Figure 5: *Histogram with distribution of check-ins per capita per year using the initial dataset.*

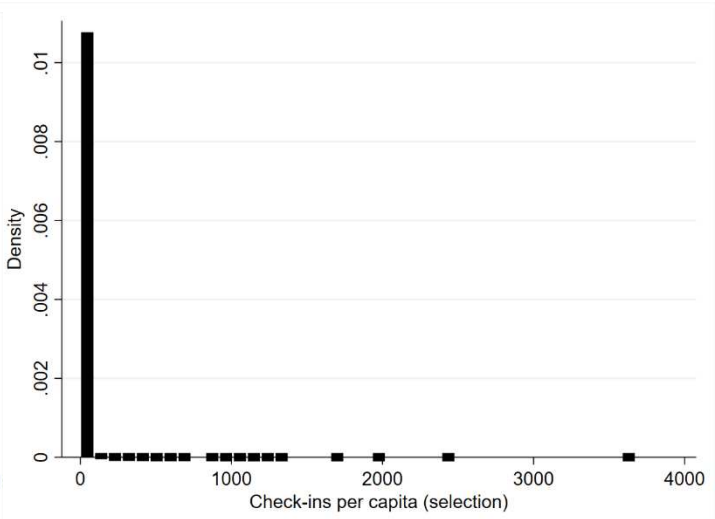


Figure 6: *Histogram with distribution of check-ins per capita per year after applying selection criteria as described in section 3.7.*

Table 2: Adjusted descriptive statistics

This table presents the number of observations (N), the mean value, the standard deviation (sd), the minimum value (min), and the maximum value (max) for the listed variables, after selection.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max
stopID	11,814	924.1	532.8	1	1,827
year	11,814	2,018	2.300	2,015	2,022
check-ins per stop	11,814	3,096	7,878	1	153,761
service area population	11,814	588.5	767.6	0	9,680
check-ins per capita	11,793	10.23	63.70	0.000288	3,676
av. distance to bus stop (m)	11,793	513.0	549.8	18.88	6,491
total frequency	11,814	1.743	1.153	0	14
routes	11,814	1.391	0.624	1	3
av. distance to station (m)	11,793	6,719	4,831	214.6	22,938
fraction 15 to 65	11,793	0.374	0.204	0	1
av. WOZ-value (x €1000)	9,858	208.1	77.79	24	866.3

3.9. Model choice

3.9.1. Time-fixed effects model

To estimate the effect of distance on relative bus stop usage, a fixed-effects model is estimated with only time-fixed effects. Time-fixed effects are important because it allows for time-varying factors that influence relative bus stop usage over time, but are constant across the whole study area. An important example is the effect of the Covid-19 crisis on travel behavior. Additionally, it could capture the effect of economic conditions or changes in policy by the public transport operator. A time-fixed effects model is estimated in the form of Ordinary Least Squares (OLS) with time dummies that capture year-specific effects. The use of entity or spatial fixed effects has been considered as well. Using fixed effects at bus stop level would make the variable of frequency drop because this variable is constant over time in this analysis for a given bus stop. Fixed effects at municipality level would not be suitable, because during the period 2015-2022 merges of municipalities have taken place in the provinces Drenthe and Groningen. As such, potential relevant characteristics of a bus stop's municipality may not be constant. A final possibility is fixed effects at town level. Perhaps, this could account for the quality or density of infrastructure which influences the choice of using a bus service, assuming it stayed relatively constant over years. Another source could be the number of amenities in a town which increases the need to travel if they are limited. However, the above examples are likely to have some variation over time. Due to the lack of compelling reasons to include town-fixed effects, it is not included in the models for the main analysis. Nevertheless, the effect of including town-fixed effect on the results is tested as a robustness check in section 4.2.

3.9.2. Endogeneity considerations

In order for OLS to produce consistent and unbiased estimates in a fixed effects model, a couple of assumptions need to hold. Hanck et al. (2024) list four important assumptions, which are now briefly assessed in the context of this research.

The first assumption is that the error term has a conditional mean of zero, which implies that regressors should be exogenous. Omitted variable bias and simultaneous causality are the main sources for violation of this assumption. For this research, the main concern is simultaneous causality concerning the variables check-ins and average distance to the bus stop. As described in section 2.1.2, the general problem of endogeneity between supply and demand characteristics in public transport is commonly addressed in the literature. It is imaginable that the density of active bus stops in a specific area, which directly affects average distance, is based on spatial variations in demand. Additionally, endogeneity may be present with regard to location choice of households that (plan to) use public transport, as described by Bento et al. (2003). The use of an instrumental variable has been considered to avoid the endogeneity problem. However, a relevant instrument that affects the average distance (or density), but which is unrelated to transit usage has not been found, given the research setting.

In this research, number of check-ins is divided by the service area population, creating a more behavioral measure of transit use. If average distance were endogenous with respect to total check-ins, but total check-ins is not highly correlated with check-ins per capita, then it suggests that the endogeneity problem would be limited when using check-ins per capita as the outcome variable. Pearson's pairwise correlations are estimated to check this, which is a coefficient between -1 and 1. Table A1 in the appendix presents correlations between all pairs of relevant variables. The correlation coefficient between total check-ins and check-ins per capita is 0.19. This correlation is regarded as sufficiently low. Therefore the described endogeneity problem is assumed to be limited when using check-ins per capita as the outcome variable.

3.9.3. Other OLS assumptions

Following Hanck et al. (2024), a second assumption is that variables are independently and identically distributed (i.i.d.) across entities, which is justified if entities are selected under simple random sampling. In this research, however, no random selection of bus stops has taken place. As such this assumption does not hold. A third assumption is that large outliers are unlikely. As described in section 3.7, effort has been made to reduce the number of high values for check-ins per capita. Still, a range of high values remained in the sample. To reduce the effect of the skewed distribution, the natural logarithm of the variable check-ins per capita is used in the analysis. A last assumption is that there is no perfect multicollinearity between variables. It is made sure that no two variables in the model exploit the same information. This is confirmed in Table A1, where no correlations exist of (near) 1 or -1.

3.10. Model specification

Before determining the exact model equation, the use of logarithmic transformations is considered for the continuous variables in the model. Taking the natural logarithm helps to reduce the skewness of a variable's distribution and to linearize relationships. Additionally, estimated parameters of a log-log model allow for the convenient interpretation of elasticities. The choice to use logarithms will be highlighted for the variables of interest: check-ins per capita and average distance to bus stop. Figure 6 and Figure 7 display the distributions for check-ins per capita and the natural logarithm of check-ins per capita, respectively. Comparing the figures, the transformation evidently helped to spread out the observations captured in the large bin in Figure 6. Figure 8 and Figure 9 display the distribution of the variable average distance to bus stop, respectively before and after transformation. In Figure 8 the skewness of this variable is clearly visible when expressed in its original scale. After the logarithmic transformation, the distribution remains slightly right-skewed in Figure 9, but has largely improved as well. Since the distribution has become significantly closer to a normal distribution for both variables of interest, the logarithm of both check-ins per capita and average distance to a bus stop is used in the remainder of the analysis.

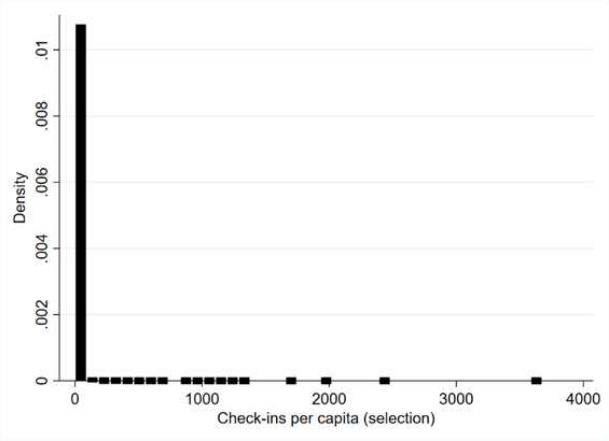


Figure 6 (repeated): *Histogram with distribution of check-ins per capita per year.*

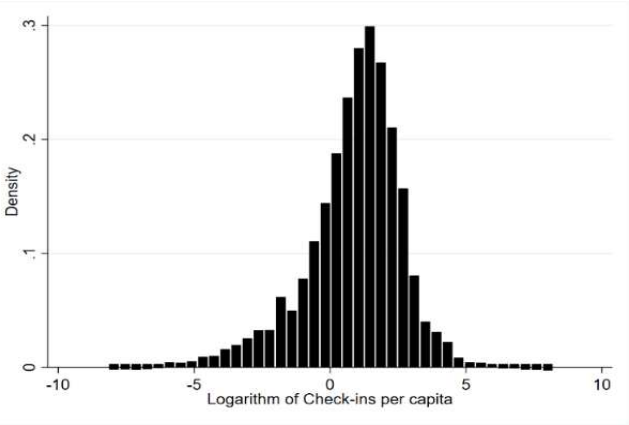


Figure 7: *Histogram with distribution of the natural logarithm of check-ins per capita per year.*

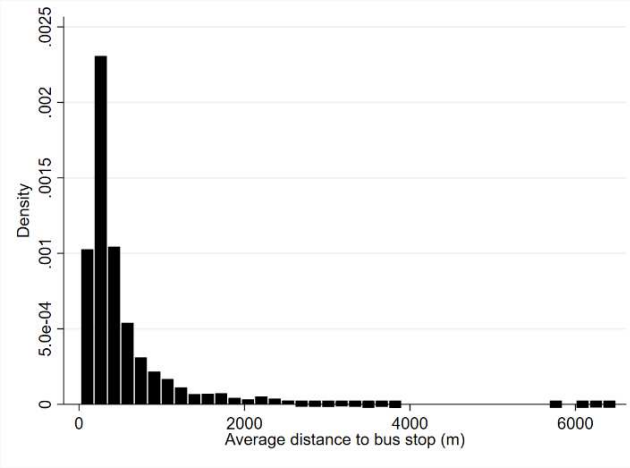


Figure 8: *Histogram with distribution of the average distance to a bus stop.*

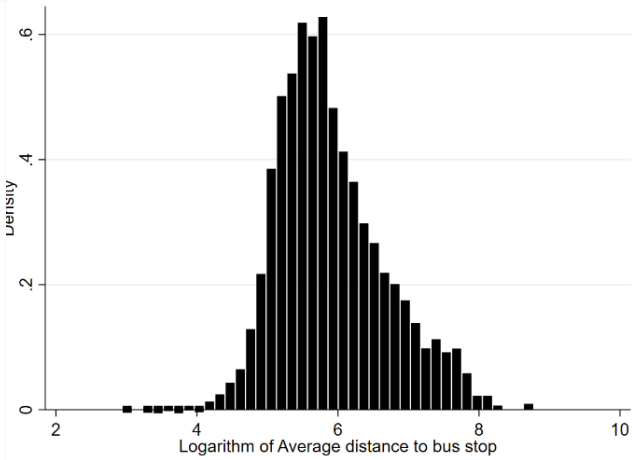


Figure 9: *Histogram with distribution of the natural logarithm of the average distance to a bus stop.*

The most comprehensive model to be estimated is specified as follows;

$$logY_{it} = \beta_0 + \beta_1 logADB_{it} + \beta_2 Freq_{it} + \beta_3 logADT_{it} + \beta_4 logWOZ_{it} + \beta_5 age1565_{it} + \sum_{t=2016}^{2022} \gamma_t Year_t + \varepsilon_{it} \quad (3)$$

where: Y is the number of check-ins per capita, $logADB$ is the natural logarithm of the average distance to a bus stop, $Freq$ is the total frequency at a bus stop, $logADT$ is the natural logarithm of the average distance to a train station, $logWOZ$ is the natural logarithm of the average WOZ-value, and $age1565$ is the fraction of inhabitants from 15 to 65 years old. Additionally, a summation of year dummies is included, with the year 2015 as the reference category. Subscript i represents a bus stop with its service area and t is the year.

Four separate models are estimated. The first model will only contain frequency and accessibility variables as explanatory variables. This model is used as a benchmark to assess the effect of the following model extensions: adding year-fixed effects, adding the socio-demographic controls for age and WOZ-value, and finally a model including both year-fixed effects and the socio-demographic controls. In the estimated models standard errors are clustered at bus stop level, which allow for autocorrelation between subsequent observations over time. Additionally, they are robust to heteroskedasticity.

4. Results

4.1. Regression results

Four OLS regression models are estimated to explain the variability in check-ins per capita for a bus stop's service area. The estimations of the four models are shown in Table 3. A first observation is that all independent variables in the models have a highly significant relationship with check-ins per capita, with p-values below 0.001. Regarding goodness of fit, measured by the adjusted R-squared, Model 3 and Model 4 appear to perform best in explaining variation of the dependent variable with a value of approximately 0.25. This is not surprising because Model 3 and 4 have additional control variables included. Although the models do not explain a high share of the variation, 0.25 is considered as an acceptable value.

The benchmark model, Model 1, is solely built up by the frequency variable and variables for the distance to the bus stop and distance to a train station from the bus stop's service area. Both the independent variable of interest, average distance to bus stop, and the dependent variable check-ins per capita are expressed as the natural logarithm. This allows for the convenient interpretation of elasticity. According to Model 1, a 1% increase in the average distance to a bus stop leads to a decrease in check-ins per capita by 0.52%, ceteris paribus. Model 2 includes year-fixed effects as an addition to Model 1. When comparing the estimated coefficients of both models, it can be concluded that including year-fixed effects has little to no effect on the estimations. This insight also holds true for the frequency and distance variables when comparing Model 3 and 4, which differ only in the inclusion of year-fixed effects. Yet, year-specific effects are relevant to account for the variation in transit use over time across the study area, for example due to the Covid-19 crisis. Therefore it is considered as an important element in the model.

Model 3 and 4 are estimated including the socio-demographic controls for fraction of 15 to 65 and average WOZ-value. With regard to the variable of interest average distance to bus stop, the estimated parameter has become more negative compared to Model 1 and 2. In the case of Model 4, a 1% increase in the average distance to a bus stop leads to a decrease in check-ins per capita of 0.77%, ceteris paribus. A possible explanation for the change in coefficient is the presence of omitted variable bias in the first two models. In these models the coefficient for average distance to bus stop is smaller

in magnitude (less negative). This variable may have absorbed a part of the effect of control variables that are positively related to the dependent variable check-ins per capita. The average WOZ-value complies to this, and could be responsible if it shares a positive correlation with average distance to bus stop. This is the case according to Table A1, denoting a low but positive correlation coefficient of 0.1. Another possible explanation for the stronger negative effect in Model 3 and 4 could source from a negative correlation between the age variable and average distance to bus stop, which is -0.3 as given in table A1.

Concerning the control variables for average distance to a train station and total frequency, they both have a highly significant positive effect on check-ins per capita, ceteris paribus. A higher bus frequency and a larger distance to a train station both make the bus relatively more attractive compared to other modes. The coefficients for the socio-demographic controls both do not have the expected sign. Based on the literature review, income was expected to negatively affect transit ridership. The variable average WOZ-value, which is used as a proxy for income, shows a significant positive relationship with check-ins per capita. The fraction of 15 to 65 year old was expected to positively affect transit use due to the number of students and commuters, but the estimated relationship is negative.

Out of the four models, Model 4 with time-fixed effects and socio-demographic controls is considered the best model. As mentioned, time-fixed effects are important to control for study area-wide variation in transit use over time. The socio-demographic controls for WOZ-value and fraction of 15 to 65 year are considered as an improvement of the model because of the limited number of variables available and the risk of omitted variable bias. Additionally, the estimated coefficients are highly significant. Although the variables may not perfectly measure the intended characteristics of the service area, which is explained in sections 3.4 and 3.5.6, the above-mentioned reasons to include them are considered more important.

Table 3: Estimation results

Dependent variable: logarithm of check-ins per capita. This table presents the estimated parameters for each of the four models. Stars indicate the level of significance. Standard errors are reported in parenthesis. The adjusted R-squared and number of observations used in the estimations are reported at the bottom of the table. All models are estimated using OLS.

	(1) Base model	(2) Time-FE	(3) S-D controls	(4) Time-FE & S-D
log(av.distance bus stop)	-0.517***	-0.517***	-0.762***	-0.772***
	(0.050)	(0.050)	(0.056)	(0.056)
total frequency	0.545***	0.545***	0.555***	0.551***
	(0.029)	(0.029)	(0.030)	(0.030)
log(av.distance station)	0.457***	0.456***	0.367***	0.364***
	(0.040)	(0.040)	(0.043)	(0.043)
Year-fixed effects	**No**	**Yes**	**No**	**Yes**
Soc.-dem. controls:	**No**	**No**	**Yes**	**Yes**
log(WOZ)			0.320***	0.488***
			(0.095)	(0.112)
fraction 15 to 65			-0.912***	-0.878***
			(0.246)	(0.246)
cons	-0.879*	-1.004*	-0.126	-0.914
	(0.436)	(0.437)	(0.702)	(0.749)
N	11799	11799	9860	9860
Adj. R^2	0.186	0.194	0.243	0.255

Robust (clustered) standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.2. Robustness checks

In order to estimate the effect of average distance to a bus stop on check-ins per capita, choices have been made regarding the appropriate study sample. The robustness of the estimated results to changes in the study sample is assessed in this section. On top of that, the effect of including town-fixed effects in the regression model is analyzed. Only the coefficient for average distance to a bus stop is examined; the variable of interest. The results are reported in Table 4 for all four model specifications. Table A3 up to A7 in the appendix show the full regression results of the robustness checks.

To obtain reliable estimation results given the purpose of this study, a sample has been created such that check-in numbers largely represent the start of journeys from travelers' residential areas. Section 3.7 describes under what criteria this sample is selected. For the selection of observations, a personal judgement is made regarding the maximum number of bus routes that should service a bus stop, in order to limit check-ins from transfers. This maximum number is set at three bus routes. To assess the effect of this choice on the results, the same regressions are run using observations with a maximum of two routes, as well as a maximum of four routes. The top two rows of Table 4 include the estimated coefficients for average distance to bus stop, using these maximum values. When compared to the coefficients of the main results in Table 3, minimal differences are observed. Excluding bus stops

serviced by three routes appears to strengthen the negative elasticity of check-ins with respect to distance by 0.01 to 0.02 percentage-points (for a 1% percent increase in distance). The difference between a maximum of four routes at a bus stop and three routes is negligible.

A third robustness check is performed by dropping the 5% highest observations for check-ins per capita. As described in section 3.7., the taken selection measures failed in removing all observations with high values. The 95% percentile has a value of 26.6 check-ins per capita, so all observations with values above 26.6 have been removed. Focusing on the most extensive models, Models 3 and 4, the negative coefficient for average distance to bus stop has weakened by around 0.01, which is also limited. As a fourth robustness check, all observations from the years 2020 up to 2022 have been excluded to more precisely assess people's sensitivity to distance, uninfluenced by coronavirus measures and hybrid working trends. This has made the negative relationship of distance on check-ins per capita stronger by around 0.02 to 0.03 percentage-points (for a 1% increase in distance). Finally, the inclusion of town-fixed effects has a relatively large impact on the results. For Model 4, the coefficient changed from -0.772 to -0.597, indicating a weakening of the effect of distance. A possible reason is that the fixed effects capture relevant factors that were previously hidden in the distance variable.

Table 4: Robustness average distance to bus stop

This table presents the estimated coefficient for log- average distance to bus stop for each robustness check (top to bottom) and for each model specification (left to right). Starts indicate the level of significance. Standard errors are reported in parentheses. For each model estimation, the adjusted R-squared and number of observations (N) is given in italics.

	(1) Base model	(2) Time-FE	(3) S-D controls	(4) Time-FE & S-D
2 routes	-0.531***	-0.530***	-0.780***	-0.790***
(1)	(0.051)	(0.051)	(0.057)	(0.058)
	N = 10904, R2 = 0.178	*N = 10904, R2 = 0.185*	*N= 9020, R2 =0.185*	*N= 9020, R2 =0.185*
4 routes	-0.519***	-0.519***	-0.763***	-0.773***
(2)	(0.049)	(0.049)	(0.056)	(0.056)
	N = 12202, R2 =0.203	*N = 12202, R2 =0.211*	*N = 10250, R2 =0.261*	*N = 10250, R2 =0.273*
5% highest check-ins out	-0.549***	-0.549***	-0.750***	-0.759***
(3)	(0.046)	(0.046)	(0.055)	(0.055)
	N =11218, R2 = 0.185	*N =11218, R2 = 0.192*	*N =9569, R2 = 0.222*	*N =9569, R2 = 0.233*
2015-2019	-0.525***	-0.527***	-0.794***	-0.794***
(4)	(0.056)	(0.056)	(0.064)	(0.064)
	N =7418, R2 = 0.175	*N =7418, R2 = 0.179*	*N =6130, R2 = 0.230*	*N =6130, R2 = 0.232*
town-fixed effects	-0.449***	-0.448***	-0.594***	-0.597***
(5)	(0.049)	(0.049)	(0.058)	(0.058)
	N =11795, R2 = 0.471	*N =11795, R2 = 0.479*	*N =9858, R2 = 0.514*	*N =9858, R2 = 0.523*

Robust (clustered) standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5. Conclusion and discussion

5.1. Conclusion

This study aimed at answering the following research question: *"To what extent does the distance to the nearest bus stop affect bus usage by local residents?"*. For this purpose, data has been collected on smartcard check-ins at bus stops and population numbers in the Dutch peripheral provinces Drenthe and Groningen, over a range of eight years. By the use of GIS-software, service areas for bus stops have been defined, and the average distance from each bus stop to its service area has been computed. As opposed to the Dutch study from Kerkman et al. (2015), the current study uses a more comprehensive measure for accessibility, using larger service areas and taking into account population patterns in these areas. The relationship between the average distance and the average number of check-ins per capita has been tested using OLS, controlling for service levels and socio-demographic characteristics of the local population. The elasticity of check-ins per capita with respect to the distance to a bus stop is found to be -0.77, from which it can be concluded that there exists a strong dynamic between distance to the bus stop and bus usage. However, this obtained result does not fully answer the exact research question, which focuses on bus usage of local residents. Efforts have been taken to select bus stops which are expected to be mainly used for journey starts by local residents, but the presence of check-ins by non-locals cannot be ruled out in many occasions. This issue implies that the calculated values for check-ins per capita are likely overestimated, which is a first limitation of this study. In light of the performed robustness checks, the estimated elasticity is still assumed to be a fair indication for the common variability between check-ins per capita and distance to a bus stop.

5.2. Interpretation of result

The estimated negative relationship between check-ins per capita and average distance is not surprising, because a larger distance to a bus stop makes other modes of transport more attractive, for instance because of convenience or due to shorter travel time. In the current study the main focus is laid on the size of the effect. One elasticity value of -0.77 has been obtained, which is an average for the study area. The provinces Drenthe and Groningen are covered by urbanized areas as well as rural areas. In the context of US urban areas, Ewing & Cervero (2010) computed an elasticity of transit use with respect to distance to the nearest stop of -0.29. Although travel behavior and transit availability in the US may not be comparable to Dutch standards, it is a sign that potential travelers are relatively less sensitive to distance in densely populated areas, which have a higher density of transit stops. For example, a 10%-change in distance to a bus stop can be more impactful for residents living on the countryside than for residents in a city, because the absolute change is larger. Mulley et al. (2018) have shown that travelers in urban areas are prepared to walk further to a bus stop if a higher frequency of service is provided. High frequency services are likely to be found in densely populated areas with a high concentration of demand. This could be another explanation why the elasticity value of check-ins per capita with respect to the distance to a bus stop of -0.77 is likely an overestimation for more urbanized areas. The estimated

model in this study did not account for a potential interaction between distance and frequency. However, for the analysis a large number of bus stops with a high total frequency have been filtered out, because the maximum number of routes at a bus stop was set at three to limit transfer check-ins. The potential interaction found by Mulley et al. (2018) may therefore be limited in this study.

5.3. Limitations

Apart from the probable overestimation of check-ins per capita, other limitations can be identified in this study. One limitation is related to the measurement of the variable of interest, the average distance to a bus stop. The distance from residential locations to the closest bus stop is measured "as the crow flies", but in reality people travel over roads. This has led to underestimated values for the average distance to a bus stop, especially in areas with a low density of roads. On top of that, it might not be exactly the distance to the bus stop that is important for people, but the travel time. The use of actual distance or travel time, instead of straight-line distance, is expected to have affected the allocation of service areas.

Another limitation of this study is the limited number and inaccuracy of control variables, with total frequency as a first example. Because of limited historical data for this important driver of demand, total frequency at a bus stop for all years is estimated based on the current frequency of the routes that see the bus stop. While a bus stop's service level can vary over years, changes in frequency or headways at a bus stop are not accurately represented, which impacts the reliability of the estimated effect of frequency on check-ins per capita. Also the use of the socio-demographic control variables can be questioned. The variable for age is not accurately measured in many occasions, which sourced from the detailed scale of the CBS data at 100-meter raster level. On top of that, this dataset does not include variables about income of the population, which resulted in the choice of using WOZ-value as a proxy. However, the use of data at 100-meter raster level was necessary to define service area borders and to compute average distance to the bus stop as accurately as possible.

A final limitation of the study is the restricted external validity of the main result. The estimated elasticity of check-ins per capita with respect to the distance to a bus stop is an average for the whole study area, which has urban regions as well as rural regions. The result could hold for study areas with similar contrasts in population density as in Drenthe and Groningen, and comparable organization and quality of public transport. Other peripheral provinces in the Netherlands could be a candidate. However, the applicability to urban areas in isolation, or to different countries is considered to be limited.

5.4. Further research

Opportunities for further research are proposed that follow from limitations of the current study. In order to more accurately analyze the sensitivity of distance for the local population, a method is needed to isolate journey starts from the residential area. This is not fully accomplished in this study. A new research attempt could be enhanced using data on individual trip level, including the check-in location and the timing of the check-in. A smartcard ID in the data would be useful to identify all check-ins on a

day for an individual, where the earliest one is likely to be the journey start from the residence. This would enable the identification of check-ins from local residents, creating more reliable estimates for the relationship between the average distance and check-ins per capita. However, this is merely a description of the ideal dataset, and such information can be sensitive due to privacy concerns. Alternatively, stated preference techniques can form an outcome, with surveys conducted in the surrounding area of a bus stop, by asking participants about their transit usage and measuring the distance from their property to the bus stop. However, deviations are possible between actual travel behavior and recorded behavior.

Regardless of the method, further research attempts could measure travelers' sensitivity to distance to a transit stop solely in non-urban areas. Most existing literature is focused on urban areas. Given the signs from literature that the elasticity estimated in the current study is likely an overestimate for urban areas, a study focused on non-urban areas only is expected to obtain different results. This information would be helpful for the public transport operator to design efficient bus routes, and at the same time serving rural populations as well as possible.

6. References

Aman, J.J.C., & Smith-Colin, J. (2020). Transit Deserts: Equity analysis of public transit accessibility. *Journal of Transport Geography*, *89*, 102869. https://doi.org/10.1016/j.jtrangeo.2020.102869

Bastiaansen, J., & Breedijk, M. (2022). *Toegang voor iedereen? Een analyse van de (on)bereikbaarheid van voorzieningen en banen in Nederland*. Planbureau voor de Leefomgeving. https://www.pbl.nl/publicaties/toegang-voor-iedereen

Bento, A., Cropper, M., Mobarak, A., & Vinha, K. (2003). *The Impact of Urban Spatial Structure on Travel Demand in the United States*. The World Bank. https://doi.org/10.1596/1813-9450-3007

Boelhouwer, P., Haffner, M., Neuteboom, P., & Vries, P. (2004). House prices and income tax in the Netherlands: An international perspective. *Housing Studies*, *19*(3), 415–432. https://doi.org/10.1080/0267303042000204304

Börjesson, M., Fung, C. M., & Proost, S. (2017). Optimal prices and frequencies for buses in Stockholm. *Economics of Transportation*, *9*, 20–36. https://doi.org/10.1016/j.ecotra.2016.12.001

Borsje, R., Hiemstra-van Mastrigt, S., & Veeneman, W. (2023). Assessing passenger preferences for Bus Rapid Transit characteristics: A discrete choice experiment among current and potential Dutch passengers. *Research in Transportation Economics*, *100*, 101307. https://doi.org/10.1016/j.retrec.2023.101307

Bourgeat, P. (2015). A Revealed/stated Preference Approach to Bus Service Configuration. *Transportation Research Procedia*, *6*, 411–423. https://doi.org/10.1016/j.trpro.2015.03.031

Brechan, I. (2017). Effect of Price Reduction and Increased Service Frequency on Public Transport Travel. *Journal of Public Transportation*, *20*(1), 139–156. https://doi.org/10.5038/2375-0901.20.1.8

Bronsvoort, K., Alonso-González, M., Van Oort, N., Molin, E., & Hoogendoorn, S. (2021). Preferences toward Bus Alternatives in Rural Areas of the Netherlands: A Stated Choice Experiment. *Transportation Research Record*, *2675*(12), 524–533. https://doi.org/10.1177/03611981211029919

CBS (2024-1). Dashboard Bevolking. *Centraal Bureau voor de Statistiek*. https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/regionaal/

CBS (2024-2). Kaart van 100 bij 100 meter met statistieken. *Centraal Bureau voor de Statistiek.* https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/kaart-van-100-meter-bij-100-meter-met-statistieken

CBS & BRK (2024). Wijk- en Buurtkaart 2024. *Centraal Bureau voor de Statistiek*, *Basis Registratie Kadaster.* https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/wijk-en-buurtkaart-2024

Chen, C., Varley, D., & Chen, J. (2011). What Affects Transit Ridership? A Dynamic Analysis involving Multiple Factors, Lags and Asymmetric Behaviour. *Urban Studies*, *48*(9), 1893–1908. https://doi.org/10.1177/0042098010379280

Conwell, L. J., Eckert, F., & Mobarak, A. M. (2023). More Roads or Public Transit? Insights from Measuring City-Center Accessibility. *NBER Working Papers*, *30877*. https://doi.org/10.3386/w30877

Currie, G., & Delbosc, A. (2010). Modelling the social and psychological impacts of transport disadvantage. *Transportation*, *37*(6), 953–966. https://doi.org/10.1007/s11116-010-9280-2

Dargay, J. M., & Hanly, M. (2002). The demand for local bus services in England. *Journal of Transport Economics and Policy*, *36*(1), Article 1.

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment: A Meta-Analysis. *Journal of the American Planning Association*, *76*(3), 265–294. https://doi.org/10.1080/01944361003766766

Geofabrik (2024). Netherlands. OpenStreetMap data extracts. *Geofabrik GmbH.* https://download.geofabrik.de/europe/netherlands.html

Graham, D. J., Crotte, A., & Anderson, R. J. (2009). A dynamic panel analysis of urban metro demand. *Transportation Research Part E: Logistics and Transportation Review*, *45*(5), 787–794. https://doi.org/10.1016/j.tre.2009.01.001

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Hanck, C., Arnold, M., Gerber, A. & Schmelzer, M. (2024). *Introduction to Econometrics with R.* 10.5. University of Duisburg-Essen. https://www.econometrics-with-r.org/10.5-tferaaseffer.html

Holmgren, J. (2005). Demand and Supply of Public Transport—The problem of Cause and Effect. In *Competition & Ownership in Land Passenger Transport* (pp. 405–421). Elsevier. https://doi.org/10.1016/B978-008044580-9/50116-1

Holmgren, J. (2007). Meta-analysis of public transport demand. *Transportation Research Part A: Policy and Practice*, *41*(10), 1021–1035. https://doi.org/10.1016/j.tra.2007.06.003

Hörcher, D., & Tirachini, A. (2021). A review of public transport economics. *Economics of Transportation*, *25*, 100196. https://doi.org/10.1016/j.ecotra.2021.100196

Iseki, H., & Taylor, B. D. (2009). Not All Transfers Are Created Equal: Towards a Framework Relating Transfer Connectivity to Travel Behaviour. *Transport Reviews*, *29*(6), 777–800. https://doi.org/10.1080/01441640902811304

Kain, J. F., & Liu, Z. (1999). Secrets of success: Assessing the large increases in transit ridership achieved by Houston and San Diego transit providers. *Transportation Research Part A: Policy and Practice*, *33*(7), 601–624. https://doi.org/10.1016/S0965-8564(99)00009-9

Kerkman, K., Martens, K., & Meurs, H. (2015). Factors Influencing Stop-Level Transit Ridership in Arnhem–Nijmegen City Region, Netherlands. *Transportation Research Record*, *2537*(1), 23–32. https://doi.org/10.3141/2537-03

Litman, T. (2008). Valuing Transit Service Quality Improvements. *Journal of Public Transportation*, *11*(2). https://doi.org/10.5038/2375-0901.11.2.3

Liu, L., Zheng, B., Luo, C., Bedra, K. B., & Masrabaye, F. (2022). Access to City Center: Automobile vs. Public Transit. *International Journal of Environmental Research and Public Health*, *19*(9), Article 9. https://doi.org/10.3390/ijerph19095622

Malekzadeh, A., & Chung, E. (2020). A review of transit accessibility models: Challenges in developing transit accessibility models. *International Journal of Sustainable Transportation*, *14*(10), 733–748. https://doi.org/10.1080/15568318.2019.1625087

Mohring, H. (1972). Optimization and Scale Economies in Urban Bus Transportation. *The American Economic Review*, *62*(4), 591–604.

Mulley, C., Ho, C., Ho, L., Hensher, D., & Rose, J. (2018). Will bus travellers walk further for a more frequent service? An international study using a stated preference approach. *Transport Policy*, *69*, 88–97. https://doi.org/10.1016/j.tranpol.2018.06.002

Nolan, A. (2010). A dynamic analysis of household car ownership. *Transportation Research Part A: Policy and Practice*, *44*(6), 446–455. https://doi.org/10.1016/j.tra.2010.03.018

Qbuzz (2024). Reizen. *Qbuzz.* https://www.qbuzz.nl/gd/lijn

Rayaprolu, H., & Levinson, D. (2024). Co-evolution of public transport access and ridership. *Journal of Transport Geography*, *116*, 103844. https://doi.org/10.1016/j.jtrangeo.2024.103844

Rijksoverheid. (2024). *Openbaar vervoer.* Rijksoverheid. https://www.rijksoverheid.nl/onderwerpen/openbaar-vervoer

RTV Noord. (2023, February 25). *Ruim 200 bushaltes verdwenen in Groningen*. RTV Noord. https://www.rtvnoord.nl/nieuws/1000917/ruim-200-bushaltes-verdwenen-in-groningen

Stock, J.H. & Watson, M.W. (2020). *Introduction to Econometrics, Global Edition.* (4th ed.) Pearson

Taylor, B. D., & Fink, C. N. Y. (2003). *The Factors Influencing Transit Ridership: A Review and Analysis of the Ridership Literature*. University of California Transportation Center, Working Papers. https://escholarship.org/uc/item/3xk9j8m2

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. https://doi-org.vu-nl.idm.oclc.org/10.1016/j.tra.2008.06.007

Tu, Q., de Haan, J., & Boelhouwer, P. (2018). House prices and long-term equilibrium in the regulated market of the Netherlands. *Housing Studies*, *33*(3), 408–432. https://doi.org/10.1080/02673037.2017.1346786

7. Appendix

Table A1: Pearson correlation coefficients

This table presents the Pearson correlation coefficient between variable pairs. A star indicates statistical significance of the correlation at 5%-level.

VARIABLES	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
check-ins per stop [1]	1.000								
service area population [2]	0.351*	1.000							
check-ins per capita [3]	0.191*	-0.060*	1.000						
av. distance to bus stop (m) [4]	-0.110*	-0.030*	0.021*	1.000					
total frequency [5]	0.5192*	0.300*	0.116*	-0.136*	1.000				
routes [6]	0.1974*	0.133*	0.073*	-0.094*	0.403*	1.000			
av. distance to station (m) [7]	-0.132*	-0.268*	-0.007	0.175*	-0.251*	-0.064*	1.000		
fraction 15 to 65 [8]	0.340*	0.536*	-0.063*	-0.312*	0.356*	0.156*	-0.363*	1.000	
av. WOZ-value (x €1000) [9]	-0.015	-0.044*	0.053*	0.097*	0.062*	-0.023*	0.013	-0.197*	1.000

Table A2: Full results main analysis

Dependent variable: logarithm of check-ins per capita. This table presents the estimated parameters for each of the four models from the main analysis. Stars indicate the level of significance. Standard errors are reported in parenthesis. The adjusted R-squared and number of observations used in the estimations are reported at the bottom of the table. All models are estimated using OLS.

	(1)	(2)	(3)	(4)
	Base model	Time FE	S-D controls	Time FE & S-D
log(av. dist bus)	-0.517***	-0.517***	-0.762***	-0.772***
	(0.050)	(0.050)	(0.056)	(0.056)
total frequency	0.545***	0.545***	0.555***	0.551***
	(0.029)	(0.029)	(0.030)	(0.030)
log(av. dist. station)	0.457***	0.456***	0.367***	0.364***
	(0.040)	(0.040)	(0.043)	(0.043)
2015.year		0		0
		(.)		(.)
2016.year		0.148***		0.00629
		(0.032)		(0.024)
2017.year		0.257***		0.147***
		(0.035)		(0.031)
2018.year		0.289***		0.177***
		(0.038)		(0.036)
2019.year		0.328***		0.195***
		(0.042)		(0.041)
2020.year		-0.104*		-0.303***
		(0.040)		(0.040)
2021.year		-0.0964*		-0.381***
		(0.043)		(0.051)
2022.year		0.225***		-0.0531
		(0.043)		(0.054)
log(WOZ)			0.320***	0.488***
			(0.095)	(0.112)
fraction 15-65			-0.912***	-0.878***
			(0.246)	(0.246)
_cons	-0.879*	-1.004*	-0.126	-0.914
	(0.436)	(0.437)	(0.702)	(0.749)
N	11799	11799	9860	9860
R^2	0.186	0.195	0.243	0.256

Robust (clustered) standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A3: Full results: 2 routes

Dependent variable: logarithm of check-ins per capita. This table presents the estimated parameters for each of the four models, using bus stops that see two routes at maximum. Stars indicate the level of significance. Standard errors are reported in parenthesis. The adjusted R-squared and number of observations used in the estimations are reported at the bottom of the table. All models are estimated using OLS.

	(1) Base model	(2) Time FE	(3) S-D controls	(4) Time FE & S-D
log(av. dist bus)	-0.531***	-0.530***	-0.780***	-0.790***
	(0.051)	(0.051)	(0.057)	(0.058)
total frequency	0.572***	0.573***	0.582***	0.579***
	(0.033)	(0.032)	(0.034)	(0.034)
log(av. dist. station)	0.484***	0.485***	0.393***	0.392***
	(0.042)	(0.042)	(0.046)	(0.046)
2015.year		0		0
		(.)		(.)
2016.year		0.122***		-0.0342
		(0.035)		(0.027)
2017.year		0.279***		0.159***
		(0.038)		(0.035)
2018.year		0.313***		0.200***
		(0.041)		(0.039)
2019.year		0.331***		0.196***
		(0.045)		(0.044)
2020.year		-0.0713		-0.284***
		(0.043)		(0.042)
2021.year		-0.0717		-0.364***
		(0.046)		(0.054)
2022.year		0.247***		-0.0379
		(0.047)		(0.057)
log(WOZ)			0.309**	0.466***
			(0.098)	(0.115)
Fraction 15-65			-0.881***	-0.855***
			(0.257)	(0.257)
_cons	-1.088*	-1.239**	-0.266	-1.004
	(0.452)	(0.454)	(0.729)	(0.776)
N	10904	10904	9020	9020
r2_a	0.178	0.185	0.235	0.247

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A4: Full results: 4 routes

Dependent variable: logarithm of check-ins per capita. This table presents the estimated parameters for each of the four models, using bus stops that see four routes at maximum. Stars indicate the level of significance. Standard errors are reported in parenthesis. The adjusted R-squared and number of observations used in the estimations are reported at the bottom of the table. All models are estimated using OLS.

	(1) Base model	(2) Time FE	(3) S-D controls	(4) Time FE & S-D
log(av. dist bus)	-0.519***	-0.519***	-0.763***	-0.773***
	(0.049)	(0.049)	(0.056)	(0.056)
total frequency	0.531***	0.531***	0.543***	0.539***
	(0.027)	(0.027)	(0.028)	(0.028)
log(av. dist. station)	0.445***	0.444***	0.357***	0.354***
	(0.039)	(0.039)	(0.042)	(0.042)
2015.year		0		0
		(.)		(.)
2016.year		0.147***		0.0128
		(0.030)		(0.022)
2017.year		0.250***		0.139***
		(0.034)		(0.030)
2018.year		0.286***		0.176***
		(0.037)		(0.034)
2019.year		0.336***		0.201***
		(0.041)		(0.040)
2020.year		-0.105**		-0.303***
		(0.039)		(0.039)
2021.year		-0.107**		-0.387***
		(0.041)		(0.050)
2022.year		0.208***		-0.0660
		(0.042)		(0.053)
log(WOZ)			0.300**	0.472***
			(0.094)	(0.110)
Fraction 15-65			-0.913***	-0.878***
			(0.243)	(0.243)
_cons	-0.733	-0.855*	0.101	-0.703
	(0.429)	(0.430)	(0.699)	(0.745)
N	12202	12202	10250	10250
r2_a	0.203	0.211	0.261	0.273

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A5: Full results: excluding observations with 5% highest check-ins values

Dependent variable: logarithm of check-ins per capita. This table presents the estimated parameters for each of the four models, using only observations up to the 95th percentile of check-ins per capita. Stars indicate the level of significance. Standard errors are reported in parenthesis. The adjusted R-squared and number of observations used in the estimations are reported at the bottom of the table. All models are estimated using OLS.

	(1) Base model	(2) Time FE	(3) S-D controls	(4) Time FE & S-D
log(av. dist bus)	-0.549***	-0.549***	-0.750***	-0.759***
	(0.046)	(0.046)	(0.055)	(0.055)
total frequency	0.496***	0.497***	0.510***	0.507***
	(0.027)	(0.027)	(0.030)	(0.030)
log(av. dist. station)	0.434***	0.433***	0.358***	0.355***
	(0.038)	(0.038)	(0.042)	(0.042)
2015.year		0		0
		(.)		(.)
2016.year		0.149***		0.00912
		(0.032)		(0.024)
2017.year		0.243***		0.141***
		(0.036)		(0.032)
2018.year		0.272***		0.163***
		(0.039)		(0.036)
2019.year		0.311***		0.169***
		(0.043)		(0.041)
2020.year		-0.0712		-0.268***
		(0.040)		(0.040)
2021.year		-0.0671		-0.350***
		(0.043)		(0.051)
2022.year		0.241***		-0.0376
		(0.043)		(0.053)
log(WOZ)			0.281**	0.434***
			(0.092)	(0.108)
Fraction 15-65			-0.728**	-0.701**
			(0.239)	(0.239)
_cons	-0.558	-0.687	0.0196	-0.695
	(0.412)	(0.414)	(0.678)	(0.723)
N	11218	11218	9569	9569
r2_a	0.185	0.192	0.222	0.233

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A6: Full results: years 2015-2019

Dependent variable: logarithm of check-ins per capita. This table presents the estimated parameters for each of the four models, using only observations from the years 2015-2019. Stars indicate the level of significance. Standard errors are reported in parenthesis. The adjusted R-squared and number of observations used in the estimations are reported at the bottom of the table. All models are estimated using OLS.

	(1) Base model	(2) Time FE	(3) S-D controls	(4) Time FE & S-D
log(av. dist bus)	-0.525***	-0.527***	-0.794***	-0.794***
	(0.056)	(0.056)	(0.064)	(0.064)
total frequency	0.544***	0.542***	0.529***	0.528***
	(0.033)	(0.033)	(0.033)	(0.033)
log(av. dist. station)	0.466***	0.464***	0.363***	0.362***
	(0.044)	(0.044)	(0.047)	(0.047)
2015.year		0		0
		(.)		(.)
2016.year		0.149***		0.00690
		(0.032)		(0.024)
2017.year		0.258***		0.148***
		(0.035)		(0.031)
2018.year		0.291***		0.178***
		(0.038)		(0.036)
2019.year		0.329***		0.190***
		(0.042)		(0.041)
log(WOZ)			0.594***	0.570***
			(0.126)	(0.128)
Fraction 15-65			-0.709**	-0.718**
			(0.269)	(0.269)
_cons	-0.839	-1.002*	-1.243	-1.222
	(0.488)	(0.488)	(0.838)	(0.842)
N	7418	7418	6130	6130
r2_a	0.175	0.179	0.230	0.232

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A7: Full results: including town-fixed effects

Dependent variable: logarithm of check-ins per capita. This table presents the estimated parameters for each of the four models, including town-fixed effects. Stars indicate the level of significance. Standard errors are reported in parenthesis. The adjusted R-squared and number of observations used in the estimations are reported at the bottom of the table. All models are estimated using OLS.

	(1)	(2)	(3)	(4)
	Base model	Time FE	S-D controls	Time FE & S-D
log(av. dist bus)	-0.449***	-0.448***	-0.594***	-0.597***
	(0.049)	(0.049)	(0.058)	(0.058)
total frequency	0.496***	0.497***	0.499***	0.502***
	(0.041)	(0.041)	(0.041)	(0.041)
log(av. dist. station)	0.342***	0.341***	0.274**	0.270**
	(0.097)	(0.097)	(0.099)	(0.100)
2015.year		0		0
		(.)		(.)
2016.year		0.113***		0.00156
		(0.029)		(0.023)
2017.year		0.236***		0.145***
		(0.033)		(0.029)
2018.year		0.238***		0.179***
		(0.035)		(0.032)
2019.year		0.296***		0.231***
		(0.038)		(0.037)
2020.year		-0.132***		-0.232***
		(0.037)		(0.038)
2021.year		-0.162***		-0.230***
		(0.039)		(0.058)
2022.year		0.174***		0.110
		(0.041)		(0.061)
log(WOZ)			-0.266*	-0.115
			(0.110)	(0.151)
Fraction 15-65			-1.040***	-0.985***
			(0.246)	(0.248)
_cons	-0.223	-0.314	2.920**	2.124
	(0.868)	(0.870)	(1.032)	(1.115)
N	11795	11795	9858	9858
r2_a	0.471	0.479	0.514	0.523

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$