Evaluating Factors Driving Bike Sharing Demand

Jasmin Classen

Computational Social Science Seminar Paper, June 2020

1 Motivation & Research Gap

In the wake of a changing climate and urban problems such as traffic congestion and air pollution increasing, awareness of the need for sustainable and green transport has risen. Among many alternative travel modes bicycling stands out as a low-cost and zero-emission mode of transport. It has been found to reduce air pollution and congestion (Fishman 2015) and increased physical activity which has a positive impact on health (Shaheen, Cohen, and Martin 2013). Higher bicycle usage has been related to higher quality of life in and a better image of cities (Pucher, Dill, and Handy 2010).

In Germany, policy makers have recognized cycling's positive impact on environment and society. Since 2002, the German government pursues the "Nationaler Radverkehrsplan" which aims at increasing the percentage of cycling traffic amongst all transportation modes to 15 % in 2020 as an important part of their overall mobility strategy and energy- and environment protection concept (BMVB 2012).

Governments and municipalities are therefore implementing measures to promote bicycle usage in cities. One option is the introduction of station-based public bicycle sharing systems which provide a flexible and low cost transportation opportunity. They offer the benefits of cycling in general combined with its own specific advantages to cities: By providing better access to public transport stations, bike sharing has been found to increase public transport usage through mode share (Shaheen, Cohen, and Martin 2013). Additionally, low-cost access to mobility enables socially excluded segments of society to have access to mobility (Grieco and Urry 2011). Also, bike sharing can lead to an increased usage of private bicycles since it leads to more visibility and normalization of the image of bicycling (Hayes et al. 2011).

It is therefore not surprising that the number of implemented public bicycle sharing systems (PBS) has grown tremendously in the last decades. Many urban cities in Europe, America and Asia have some form of bicycle sharing scheme installed nowadays (Fishman 2015). One has to differ between free-floating and station-based PBS, the former allowing the positioning of bikes anywhere in the city where as the latter needs bikes be returned to stations designated to the PBS.

A successful bike sharing system is not only characterized by a reliable functioning fleet and good pricing but also well placed and easily accessible stations. Moreover, good bike infrastructure and a land use policy supporting the complementarity of areas stations are placed in have been found to be essential (Levy, Golani, and Ben-Elia 2017; Ren et al. 2014).

All these factors are decided by city planners who require a data-centric understanding of how cycles are used in a city in order to anticipate the effectiveness of their interventions and make more successful future policy decisions. According to Levy, Golani, and

Ben-Elia (2017), urban planning has only in recent years shifted away from a car centric approach where cycle usage was only incorporated as a constant substracted from motorized mobility. There is therefore still a long way to sound data-driven decision-making in developing cycle friendly cities.

Bike sharing systems offer an easily accessible large amount of data about bike rentals of users in a city and can therefore generate valuable insight into intra-urban movements of the population. That is why research in transport science on predicting bike share demand and evaluating influential factors in station-based systems has been booming in recent years (e.g. Cheng et al. 2019; El-Assi, Salah Mahmoud, and Nurul Habib 2017; Faghih-Imani et al. 2014; Hyland et al. 2018; Levy, Golani, and Ben-Elia 2017). Many studies evaluated and identified an extensive number of influential factors such as demography, land use, built environment, transport infrastructure, temporal variables, topology and weather using a wide range of methods.

All of these papers, however, solely focused on station-based analysis, predicting what factors lead one station to produce or attract trips. They ignore potential additional information that can be derived from analyzing the interaction of origin and destination of a trip in the context of a city's structure. This is especially important for evaluating the impact of aspects municipalities have influence on (e.g. land use, built environment or transport infrastructure (Pucher, Dill, and Handy 2010)). Land use, for example, can have a different impact on demand depending on where cyclists travel from. Bike infrastructure is especially relevant since it serves as the connection between stations.

This so called spatial interaction modeling is widely used in other areas of transport modeling, for example to predict commuting flows between regions (e.g. Ren et al. (2019), Tsiotas et al. (2019)) and has been found to have strong explanatory power.

Additionally, former research on factors driving bike sharing demand so far has been of purely exploratory nature not making use of theoretic foundations.

Also, this research has completely ignored a social perspective on bike sharing mobility. A successful bike sharing system offers access to everyone. Previous literature in sociology found for example that access to modes of transport (transport disadvantage) is associated with social exclusion due to low access to goods and services which difficults such people's participation in society (Grieco and Urry 2011, 21). Very often, very specific groups are affected by transport disadvantage, namely migrants, the elderly or those with low income (Grieco and Urry 2011). Also, previous studies have found that cyclists are often white, male and young and of higher education (Fishman 2015). This highlights the presence of social economic structures in the spatial context which can not be ignored.

This paper will attempt to derive actionable advice for city planners to design a successful and inclusive bike sharing system. To address issues in previous research it will do so by analyzing spatial interaction patterns in an example bike sharing system and evaluate on that basis what factors drive bike sharing demand. The analysis will focus on factors that can be influenced by transport and urban planners. It will try to fill the theoretical gap by developing a theory of macro- and micro-level mechanisms that drive cycling demand as well as the methodological gap by developing a spatial interaction model for bike sharing tested on data from the public bike sharing system "StadtRAD" in Hamburg.
This is also especially relevant since research on bike sharing in Germany has been very limited so far (exception is Duran-Rodas, Chaniotakis, and Antoniou (2019)) and previous research has found it difficult to compare findings of bike share analysis across cities or countries due to different topological, social and meteorological properties of cities (Mattson and Godavarthy 2017). Last, this paper will showcase how a social perspective can be incorporated in such analyses.

The remaining paper is structured as follows: After a literature review of previous bike sharing research a theoretical foundation of spatial interaction through bike sharing is developed. This is followed by a description of the study region, the data and variables that will be used in the model as well as an outline of the method used to test hypotheses. I conclude with reporting results, a discussion of its implications for social science perspectives as well as a critical examination of shortcomings of my approach.

2 Literature Review

Many studies have focused on estimating travel demand in bike share systems. They have made use of more and more publicly available data sets containing origin-destination pairs of bicycle trips between bike share stations.

Multiple studies have been conducted for bike sharing systems in China (e.g. Wu et al. (2018), Lin et al. (2019)), the US (e.g. Noland, Smart, and Guo (2018)), Canada (Faghih-Imani et al. 2014), Australia (Mateo-Babiano et al. 2016), France (e.g. Borgnat et al. (2012)) and Spain (Hampshire and Lavanya 2012). Only Duran-Rodas, Chaniotakis, and Antoniou (2019) analyzed German bike sharing networks.

Most of previous research has focused on estimating station-level trip generation or attraction by aggregating the number of outgoing or incoming trips to one station on a hourly, daily or monthly basis (e.g. Duran-Rodas, Chaniotakis, and Antoniou (2019), El-Assi, Salah Mahmoud, and Nurul Habib (2017), Faghih-Imani et al. (2014), Hyland et al. (2018), Levy, Golani, and Ben-Elia (2017), Scott and Ciuro (2019)).

Influential factors evaluated in these studies reach from weather and temperature characteristics, temporal variables such as weekday, season and time of day, population or employment in station area or distance to points of interest such as restaurants, universities or shopping centers. Other factors include transport infrastructure, competing transport modes (e.g. distance to next subway stop) or the type of land use in a station's area. Since the applied methods were station focused, researchers usually defined a service area for each station, which equals a certain radius of a walkable distance around a station, and then computed variables within these areas, for example the percentage of bike paths of all roads within 400 m of a station.

For variables relevant to municipal policy making, namely bike and transport infrastructure as well as land use, literature mostly came to unambiguous findings.

Previous studies mainly found a positive impact of more and better bicycle infrastructure on bike sharing usage. For example, higher street density and connectivity as well as proximity of bike lanes were associated with increased bike sharing usage (e.g. Lin et al. (2019), Pucher, Dill, and Handy (2010)). El-Assi, Salah Mahmoud, and Nurul Habib (2017), Faghih-Imani et al. (2014) found that improved bike lanes and paths increased bike sharing ridership.

Concerning land use, bike stations located in residential or commercial areas increased bike share usage depending on time of day (mainly commuting travel in morning and evening peak hours) (Lin et al. 2019). Noland, Smart, and Guo (2018) found an increased usage between stations with different land use properties (e.g. start in residential and end in commercial area) and also a temporal dependency.

Methods to evaluate the impact of these variables on trip generation or attraction mirror the complex nature of bike sharing trip data. Models where estimated using simple linear regressions (Faghih-Imani et al. 2014; Mateo-Babiano et al. 2016; Zhang et al. 2018) or, in more advanced studies, poisson or negative binomial models due to the count nature of trip occurrence (Wang et al. 2016). Additionally, different kinds of models where used to account for a large number of zero trips between certain stations by introducing zero-inflation to negative binomial models (Wang et al. 2016). Others incorporated temporal and spatial auto correlation by using lagged spatial variables or spatial weight matrices (Zhang et al. 2017). Another string of research applied multilevel models to account for individual trips being nested into stations or customers (Hyland et al. 2018; Scott and Ciuro 2019). Zaltz Austwick et al. (2013) and Yao et al. (2019) conducted network analysis, their methods where, however, restricted to a very basic and descriptive exploration of bike sharing systems' network structure.

An important concept in transport modeling is spatial interaction. It denotes a concept applied in a variety of research fields and refers to the movement of objects (e.g. persons or goods) between an origin and destination point (Masucci et al. 2013). It has been applied to model migration (Li et al. 2017) and trade flows (Baltagi, Egger, and Pfaffermayr 2003) for example and found its way to transport modeling where it was applied to research commuting (Stefanouli and Polyzos 2017; Ren et al. 2014 for example) or urban mobility (Khan et al. 2017).

One widely used model to predict spatial interaction between two (in the transport case: geographical) units is the gravity model. In its simplest form it assumes that the number of interactions between specific units depends proportionally on attraction factors (also called mass factors) of origin and destination as well as inversely proportional on the distance between them (Patuelli and Arbia 2016). It needs to be noted that the distance function is a general deterrence function and does not necessarily have to be of geographical dimension (Patuelli and Arbia 2016).

The gravity model has been applied in the transport domain, for example to estimate overall traffic, commuting or tourism flows between regions (Stefanouli and Polyzos (2017), Yang, Li, and Li (2019) or Zhang, Cheng, and Jin (2019) to mention the latest works). Ren et al. (2019) and Demissie, Phithakkitnukoon, and Kattan (2019) find that the gravity model is also applicable to intra-urban travel characterized by shorter distances.

Past studies have criticized the gravity model for being too simplistic and lacking a sound theoretical guidance (Ren et al. 2014). The division into one start and one destination zone might omit important additional catchment centers and additional significant variables might be omitted (Stefanouli and Polyzos 2017). Lenormand, Bassolas, and Ramasco (2016) showed that the gravity model performs better in simulating commuting flows than the intervening opportunities model and Stefanouli and Polyzos (2017) and Masucci et al. (2013) find that the gravity model can compete with the radiation model given additional parameters. That is why the gravity model still exceeds most other models proposed, mainly due to its flexibility and simple parameter estimation and calibration (Ren et al. 2019; Stefanouli and Polyzos 2017).

Besides the classical mass value (population) and distance (travel time or geographical distance), the gravity model for transport mobility can be extended by using additional influential mass and distance or cost variables. On an inter-regional level this has been done by using GDP at origin and destination for example (see Stefanouli and Polyzos (2017) and Tsiotas et al. (2019)). On the intra-zonal level Ren et al. (2019) incorporated land use at start and destination into their models.

Concerning modeling bike sharing flows, no study has made use of the full potential of the gravity model. A first step in the direction were made by El-Assi, Salah Mahmoud, and Nurul Habib (2017) and Mateo-Babiano et al. (2016) who incorporated population and land use characteristics of start and end station into their regressions. Their studies, however, did not consider a distance factor between stations.

Noland, Smart, and Guo (2018) took it one step further and were the first to apply a form of spatial interaction modeling to PBS in New York. They considered land use and population at start and end station, as well as the distance between station pairs. Noland et al., however, use only a rudimentary measure for bike infrastructure quality and distance between stations and focused their analysis on differences between casual and subscriber users.

All studies making use of bike sharing data are of pure exploratory nature so far not incorporating a theoretical base for their analyses. Y. Wang et al. (2018) are the first to incorporate theory into researching bike sharing usage by introducing a theory of perceived value for bike sharing. They test their theory using a survey and find strong support of a relationship between a higher perceived value of bike sharing and increased usage of such.

Concerning a social and accessibility perspective on bike sharing, research is scarce due to issues with available data. Wang et al. (2016) found that there are generational differences between bike sharing users, their analysis is, however, solely based on regular subscribers of a bike sharing system and not occasional users. Sociological research in transport has focused on analyzing through surveys how access to transport and social exclusion are related. Social exclusion has been found to be driven by lack of access to a functioning transportation system for communities of low-income or migrant background amongst others (Grieco and Urry 2011, 22). Transport policy and provision often reflects socio-spatial patterns (Grieco and Urry 2011, 211). Groups that are especially affected by transport disadvantage are ethnic minorities, the elderly population as well as those with low income. Land-use policy can influence this as well by locating new developments at more accessible locations (Grieco and Urry 2011, 223). Additionally, a few studies have tried to evaluate how socioeconomic background relates to bike usage in general. Welsch, Conrad, and Wittowsky (2018) found that migrant background is negatively correlated to cycle usage. Also, cycling is more widespread among users who are younger and male. Transport therefore plays an important role in reinforcing patterns of social exclusion and disadvantage. Further insights on how bike sharing is influenced by this mechanisms are therefore essential.

This literature review lets conclude that even though widely researched, modeling bike sharing demand is still lacking a methodologically well-specified spatial interaction perspective, a theoretical base as well as a socioeconomic perspective. Since other research areas in modelling transport flows show promising results and potential of the gravity model I will apply an extended version of the gravity model to an example data set of bike sharing trip data in order to evaluate influential factors on trip occurrence. Additionally, I will apply the analysis to a German bike sharing system which has only been done by one study so far using a different approach. Moreover, I will develop a theory based on the theory of perceived value as suggested by Y. Wang et al. (2018) to explain why certain factors influence bike sharing usage in the way they do. Lastly I will incorporate citizenship in the analysis to showcase a social science application of this model approach and to include a socioeconomic perspective on the topic.

3 Theory and Hypotheses

Spatial intra-urban interaction is based on an individual's decision to move between an origin and a destination. This can be based on the activity-mobility theory by Axhausen and Gärling (1992) which states that the movement of people is based on changing places to pursue different social activities. Ren et al. (2019) derive the concept of places as physical space and geographic entities which are connected with each other via aggregated flows. Flows serve as “concrete manifestation of spatial interaction that reflects the interrelationships among places” (Ren et al. 2019). The volume of such flows is also known as interaction intensity (Taylor and Catalano 2002). In the bike sharing context, places refer to PBS stations and their surroundings whereas flows are movements of persons from one rental station to the other.

The activity-mobility theory, however, solely explains that people move between physical places in general. What drives their decision to travel to a specific place using a specific mode of transport has been found to be both driven by macro-level external factors as well as an individual’s traits on the micro-level (Scheiner and Holz-Rau 2007). The next section will elaborate on a theoretic foundation.

3.1 Macro - level drivers of intra-urban travel

The macro-level driver of spatial mobility is the urban form which defines the physical characteristics of a city (Scheiner and Holz-Rau 2007). These range from its shape and configuration also to nonphysical characteristics such as population density.

Population density or size provides the “source” of travel. The larger the population, the higher the trip volume, since there are potentially more travelers. When applied to bike sharing, one can also expect an increased travel flow between two stations if the population of origin and destination is higher. This leads to the following hypothesis.

H1: The higher the population at origin or destination, the higher bike sharing usage between these stations.

As activity-mobility theory states people travel to pursue social activities. These can take place in different areas of the city that serve different purposes. In the intra-urban case this can be professional for example, e.g. commuting between work and residence or for leisure. For interaction to take place, the land use of origin and destination need to be complementary (Ren et al. 2019). This leads to H2:

H2: Different land use types at start or destination lead to different travel volumes.

3.2 Individual - level drivers of intra-urban travel

A first theoretic approach on what drives an individual’s mode choice is made by Y. Wang et al. (2018) who use the theory of perceived-value to explain individual-level travel mode choices for bike sharing. Said theory goes back to Zeithaml (1988) who stated perceived value of a consumer as their assessment of the utility of a product or service. Previous research validates the assumption of increased usage of a product with increased perceived value for the case of public transport (Lai and Chen 2011). Y. Wang et al. (2018) also find support for their theory of perceived-value for bike sharing. Scholars differentiate between multiple dimensions of perceived value of which the following are relevant for the bike sharing case: (1) functional value (2) emotional value (3) social value (4) conditional value.

Functional value is defined as the perceived utility of a product or service based on its specific attributes. This translates to the convenience and safety of using a shared bike to conduct a journey (Fishman 2015; Y. Wang et al. 2018).

From a transport planning perspective, perceived comfort and safety of biking and bike sharing in a city can be actively increased by municipalities by providing well-connected and high quality bike infrastructure such as bicycle paths or lanes which leads to hypothesis 3.

H3: The better bicycle infrastructure between origin and destination, the higher bike sharing usage between these stations.

Referring back to the activity-mobility theory where people are bound to move between places to pursue certain social activities the transport mode of choice functions as a connection to said activities. The perceived functional value of using a shared bike should therefore increase if it enables the individual to reach these destinations comfortably. This leads to 4:

H4: The closer the station to points of social activity the higher bike sharing usage at these stations.

The *conditional value* translates into the value of a product or service given a certain situation or circumstance. In the bike sharing case Y. Wang et al. (2018) state for example an increased distance between origin and destination can negatively influence the perceived value of using bike sharing.

H5: The shorter the distance between origin and destination, the higher bike sharing usage between these stations.

Social and emotional value also play an important role assigning value to using a shared bike. Previous research has found that lifestyle, for example, shapes attitudes towards cycling (Scheiner and Holz-Rau 2007). A positive attitude towards environmentally friendly travel is such a trait that would increase the perceived emotional value of using a shared bike. Analyzing this, however, must be left to further research, since this paper's goal lies in evaluating factors that city planners can influence.

4 Data and Study Area

4.1 Study Area

Hamburg is with roughly 1.9 million inhabitants the second largest city in Germany. It shows unique geographic structures and properties due to its position at the largest port in Germany. The port is surrounded by important touristic sites and shopping opportunities which altogether form the center of the city. The city of Hamburg finances StadtRAD in order to improve close-distance mobility and to improve cycling as a main mode of transport (BMVB 2012). It is one of the most successful station-based bike sharing systems in Germany with 207 stations, about 3 million trips every year and over 2500 bikes. Until 2022, 150 additional stations and 2000 bikes are to be added ("Radverkehrsentwicklung 2020 - hamburg.de" 2020). As can be seen in Figure 1, a large portion of stations is concentrated in the city center and stations do spread out to outer areas of the city as well.

4.2 Data

The analysis is based on historical daily trips between stations (dependent variable) of StadtRAD Hamburg as well as on data from open street map and the statistical bureau of

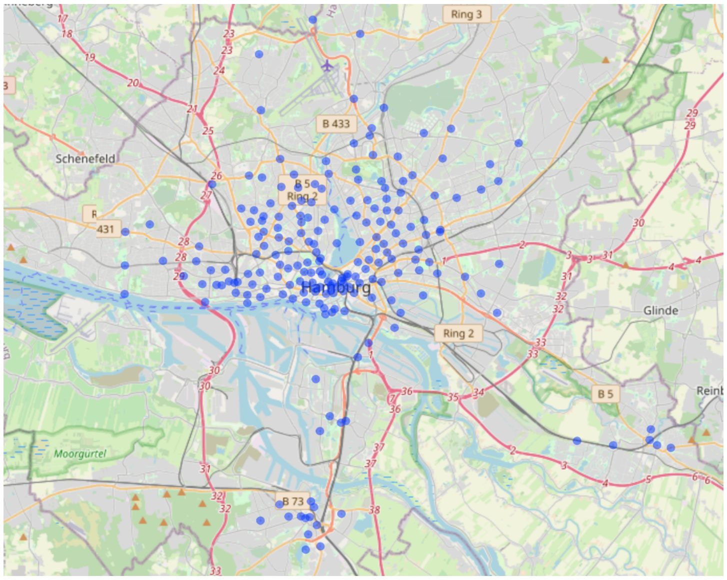

Figure 1: Map of all stations of StadtRAD Hamburg across the city.

Hamburg for independent variables (e.g. land use, population or points of interest around stations).

Historical ridership data from the Hamburg bike sharing system “StadtRAD” is part of a large data set published and made accessible by the Deutsche Bahn AG, owner and maintainer of “Call a Bike”, a large bike sharing system in five German cities (“Call A Bike Reportingdaten, Deutsche Bahn AG” 2020). The overall data set for Hamburg consists of 212 stations that recorded about eight million OD trips between 2014 to 2017. The data set contains time, start and stop station and their coordinates as well as a unique user-id per trip. For the analysis, the data set will be aggregated to contain counts of trips per origin-destination station pair per day. Additionally, the analysis will focus on one week in September 2016 where there was steady weather (no rain and temperatures between 18-21 degrees during the day) for all seven days to exclude weather effects. After excluding unrealistic trip durations (less than three minutes and more than five hours) and faulty stations as well as adding all combinations not yielding any trips the final data set consisted of 297,052 observations of 207 stations.

Data for additional parameters describing OD-pair attributes such as population, land use and bike infrastructure will be obtained from the openstreetmap-API (“Geofabrik Download Server” 2020), the “Transparenzportal Hamburg” (“Transparenzportal Hamburg” 2020) and the Census of Germany from 2011 (“Zensus 2011 - Ergebnisse zum Download” 2020).

5 Variables

This section describes the variables used in the analysis of this paper. It consists of a dependent variable and explanatory variables on trip and station level.

Table 1: Summary statistics of all variables used in the analysis. Note: "O" stands for origin station, "D" for destination station.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Trip count	297,052	0.255	0.936	0	0	0	40
Route time (scaled)	297,052	0.000	1.000	−1.612	−0.748	0.584	4.017
O - population (scaled)	297,052	−0.000	1.000	−1.478	−0.851	0.740	3.051
D - population (scaled)	297,052	−0.000	1.000	−1.478	−0.851	0.740	3.051
O - POIs close transport (scaled)	297,052	−0.000	1.000	−1.486	−0.735	0.391	5.832
O - POIs close education (scaled)	297,052	0.000	1.000	−0.542	−0.542	0.239	7.268
O - POIs close health (scaled)	297,052	−0.000	1.000	−0.391	−0.391	0.027	10.915
O - POIs close leisure (scaled)	297,052	−0.000	1.000	−0.942	−0.716	0.340	3.585
D - POIs close leisure (scaled)	297,052	−0.000	1.000	−0.942	−0.716	0.340	3.585
D - POIs close transport (scaled)	297,052	0.000	1.000	−1.486	−0.735	0.391	5.832
D - POIs close education (scaled)	297,052	0.000	1.000	−0.542	−0.542	0.239	7.268
D - POIs close health (scaled)	297,052	0.000	1.000	−0.391	−0.391	0.027	10.915
O - distance to closest station (scaled)	297,052	0.000	1.000	−1.181	−0.633	0.215	4.603
D - distance to closest station (scaled)	297,052	−0.000	1.000	−1.181	−0.633	0.215	4.603
O - distance to city center (scaled)	297,052	−0.000	1.000	−1.384	−0.713	0.396	3.620
D - distance to city center (scaled)	297,052	−0.000	1.000	−1.384	−0.713	0.396	3.620
Within 30 minutes	297,052	0.321	0.467	0	0	1	1
Loop trip	297,052	0.005	0.070	0	0	0	1
O - landuse residential (scaled)	297,052	−0.000	1.000	−1.173	−1.010	0.801	2.678
D - landuse residential (scaled)	297,052	0.000	1.000	−1.173	−1.010	0.801	2.678
O - landuse commercial (scaled)	297,052	−0.000	1.000	−0.436	−0.436	−0.041	6.349
D - landuse commercial (scaled)	297,052	−0.000	1.000	−0.436	−0.436	−0.041	6.349
O - landuse retail (scaled)	297,052	0.000	1.000	−0.342	−0.342	−0.342	6.643
D - landuse retail (scaled)	297,052	−0.000	1.000	−0.342	−0.342	−0.342	6.643
O - landuse industrial (scaled)	297,052	−0.000	1.000	−0.187	−0.187	−0.187	8.381
D - landuse industrial (scaled)	297,052	−0.000	1.000	−0.187	−0.187	−0.187	8.381
Percentage cycle way (scaled)	297,052	0.000	0.998	−1	−0.8	0.4	24
O - citizenship not German (scaled)	297,052	0.000	1.000	−1.226	−0.590	0.199	5.122
D - citizenship not German (scaled)	297,052	0.000	1.000	−1.226	−0.590	0.199	5.122

5.1 Dependent Variable

The dependent variable of the model is a count measure of all trips that have taken place between two stations per day. It was obtained by aggregated ridership data by day and OD-station-pair. As can be seen in Table 1, 87 percent of all observations contain zero trips which has to be considered in later analysis. There were a maximum of 40 trips a day between two station pairs.

5.2 Explanatory Variables: Station Level

Station-level variables where defined on service area level for each bike share station. A service area is the walkable radius around a station since stations will only be frequented by customers if they are within a walkable distance. Previous literature uses a distance between 250 and 500 meters around a station, newer research agrees on 400 meters being a realistic radius (Duran-Rodas, Chaniotakis, and Antoniou 2019; Noland, Smart, and Guo 2018).

To examine the influence of land use on trip attraction and production a polygon data frame from open street map was obtained for Hamburg. Using the buffer of 400 m explained above, each land use type's area intersecting with this buffer zone was calculated and from this separate variables for each land use type containing percentage values where obtained.

The population in a stations buffer zone was calculated using population data on the 100x100 meter level from the German census in 2011. Each grid cell contained the absolute number of persons registered there and a deviation measure in case the number of inhabitants had to be concealed due to data privacy reasons. I then aggregated the population counts in all cells inside or partially inside a station's buffer zone to account for circular service areas square grid cells. Due to some deviance in cases with small populations this variable is not 100 percent accurate but should be sufficient for this cause since population number was only concealed in cases of less than three inhabitants in a cell.

I similarly obtained the share of inhabitants with non-German citizenship within a station's service area. I first computed the sum of all non-Germans and then calculated the percentage. Here, the deviance is larger than with the population data since smaller numbers occur more often. A more detailed distinction between different citizenship countries was available but led to a deviation too large which would have introduced too much bias.

To examine how points of interest influence trip generation and attraction, I obtained different points of interest from 10 categories from open street map (see Table 4 in Appendix for a detailed list). These include for example shopping centers, universities, tourist attractions, metro, railway and bus stops, restaurants and night clubs. I categorized them into leisure and culture, transport, health and education. Since previous studies found evidence that BSS usage was higher for stations with the higher number of points in the vicinity (Faghih-Imani et al. 2014) I calculated the number of points for each category within 400 meters of each station.

Additionally, I included a distance measure to the closest bike sharing station in order to account for competing stations and bike share network density.

To control for different trip counts on weekends and weekdays I derived this categorical measure from the trip data set which contained the start and end date of the trip. I only consider start date, however, since there is only a very small number of cases where rentals occurred over night. Including both start and end date of a rental would have introduced unnecessary additional cases for overnight rentals which is not the focus of this analysis.

I additionally introduced a variable measuring the distance to the city center since the PBS' structure in Hamburg is clearly dependent on that. There are more and denser stations in the center than on the outskirts.

5.3 Explanatory Variables: trip level

The analysis uses two variables on trip level: The distance between two stations as well as the quality of bike infrastructure between them.

For distance between origin and destination station I used the time in minutes for the shortest path between them. For intra-urban mobility, Ren et al. (2019) used simply the logarithm of the euclidean distance between origin and destination. Previous studies have found, however, that the average travel time reflects the cost of a journey best for inter-regional commuting (Tsiotas et al. 2019). Noland, Smart, and Guo (2018) made therefore use of the shortest path distance using the street network in New York in order to estimate the effect of increased distance on bicycle trip generation. Since the StadtRAD data set does not contain GPS-traced routes I obtained the route between two stations as the shortest path for bikes and the corresponding time it takes to cycle it using the Cyclestreets API ("CycleStreetsAPI" 2020). The Cyclestreets API is based on a bike-centric routing engine. Not knowing the actual route bike share users took with their bike does of course pose some limitations on this measure. I therefore compared the distance measure for each station pair with the average actual rental distance and found that the shortest path distance (in minutes) did not deviate substantially. Using euclidean distance might therefore be sufficient, from a theoretic point of view, however, time in minutes is a much more realistic measure. Due to the small deviation and the theoretic reasoning I therefore decided to use route time.

To examine whether the pricing model of the PBS influences trip count I calculate a dummy variable based on the distance measure above which measure whether the destination station is reachable within 30 minutes. At StadtRAD a trip is for free for the first half hour.

As a measure of quality of bike infrastructure I calculate the percentage of bike paths on a shortest path between two stations. I did so by overlaying a data set of cycle ways from open street map with the calculated shortest routes obtained from the Cyclestreets API. As the route time measure above this again is limited by not knowing for sure which route cyclists took. I rely here on the comparably small deviation of rental times from shortest route times.

All explanatory non-dummy variables showed strongly left-skewed distributions due to a high number of zeros in many cases. These variables were therefore scaled by substracting their mean from each value and dividing by the standard deviation. This scaling approach leads to more stable regression models and facilitates interpretation of coefficients [Garay2011].

6 Model

6.1 Methodology

In order to derive influential factors on bike sharing usage for an inclusive and improved design of a bike sharing system I will use a spatial interaction model of bike sharing flows. The gravity model is widely used to model such travel volumes (see literature review for details). In this section I will therefore develop an extended gravity model for bike sharing.

6.1.1 The "Traditional" Gravity Model

The simple gravity model is based on the assumption that the geographical movement between two places depends solely on the product of the attraction force of the destination and the "mass" of the origin as well as the total travel distance (or cost) between these places (Masucci et al. 2013). The theory of perceived value supports this assumption. This relationship can be formally depicted as:

$$t_{ij} = k\frac{m_i^{\beta_1} m_j^{\beta_2}}{d_{ij}^{\gamma}}$$

where t_{ij} denotes the number of interactions between an origin i and destination j, k denotes a normalization factor, m_i the mass value for the origin and m_j for the attraction force of the destination and d_{ij} some distance measurement between origin and destination. The exponents β_1, β_2 and γ are the effect coefficients which will be determined by model estimation, usually using multiple regression analysis. γ denotes the variable influence of the distance factor on the overall outcome.

Taking the logarithm of both sides of the equation of the gravity model and adding an error term, the model takes the linear form which can then be easily estimated using multivariate regression analysis. This can be formally depicted as

$$ln(t_{ij}) = ln(k) + \beta_1 ln(m_i) + \beta_2 ln(m_j) + \beta_3 ln(d_{ij}) + \epsilon_{ij}$$

where t_{ij} denotes the number of trips between an origin station i and destination station j, k denotes a normalization factor, m_i the mass value for the origin and m_j for the attraction force of the destination and d_{ij} some distance measurement between origin and destination. The β-factors are the coefficients that will be estimated by the regression model. ϵ_{ij} denotes the error term.

In traditional gravity models of transport, e.g. for commuting flows, the population at origin and destination city serves as attraction factor (for example Stefanouli and Polyzos (2017), Tsiotas et al. (2019) or Yang, Li, and Li (2019)). The cost or deterrence function of the gravity model describes an inversely proportionally decreasing travel volume with increased distance.

The basic bike sharing gravity model will therefore consist of population as mass values as well as a distance function which incorporates geographic distance (depicted as travel time). Formally notated that is,

$$ln(t_{ij}) = ln(k) + \beta_1 ln(_i) + \beta_2 ln(p_j) + \beta_3 ln(d_{ij}) + \epsilon_{ij}$$

where t_{ij} denotes the number of trips between an origin station i and destination station j, k denotes a normalization factor, p_i population for the origin and p_j for the population of the destination and d_{ij} the route time in minutes between origin and destination. The β-factors are the coefficients that will be estimated by the regression model. ϵ_{ij} denotes the error term.

6.2 Extended The Gravity Model

As previous research has shown, gravity models can be extended with additional parameters that influence the flow volume between places via mass or attraction factors of places and the deterrence function between them.

In order to derive a full gravity model for station-based bike sharing and to test the hypotheses formulated in the theory section of this paper, the following parameters are introduced to the model:

Land use at origin and destination station as well as distance variables to points of interest, the city center and competing modes of transport are added as additional mass values to stations. The deterrence function is extended by a factor of bike infrastructure quality which moderates the distance-decay-function used in the simple gravity model.

Lastly, controlling variables for weekends versus weekdays and loop trip are introduced since they can be differently affected by mass and distance values in the gravity model.

Including all these variables leads to an extended gravity model specified as follows:

$$ln(t_{ij}) = ln(K) + \beta^{'} ln(M_i) + \beta^{'} ln(M_j) + \beta^{'} ln(D_{ij}) + \epsilon_{ij}$$

where t_{ij} denotes the number of trips between an origin station i and destination station j, K denotes a normalization factor, M_i and M_j the matrix of mass values at station level, D_{ij} the matrix of variables on trip-level between station i and j and $\beta^{'}$ denotes the matrices of the regression coefficients. ϵ_{ij} denotes the error term.

Due to the count nature of the dependent variable and the present overdispersion in a poisson model, a negative binomial model will be estimated. Given the large amount of zeros in the data that highly surmount the expected number of zeros in such a data set, a zero-inflated model negative binomial (ZINB) approach will be conducted. The zero-inflated approach assumes that zeros in a data set occur due to two different causes: True zeros result from a certain combination of influential variables such as different population levels, land use, vicinity of points of interest or competing transport modes and so on. Excess zeros, however, are caused by non-man made factors such as a distance too large to be cycled (Zhang et al. 2017). The zero-inflated negative binomial model accounts for both processes by estimating two separate models for the generation of true zeros and excess zeros (Garay et al. 2011). The distribution of excess zeros hereby follows a logit distribution, the model estimates the probability for the logit process to generate zeros while true zeros are modeled to follow a negative binomial distribution. Both models are estimated simultaneously using maximum likelihood estimation. Since large distances can both be the cause of excess and true zeros, I will be using a mixed model so that both the count and the probability model can produce zeros. The count model incorporates the gravity model as described above while the probability model will be estimated using only distance as an independent variable.

7 Results

To report model performance and fit, multiple measures were calculated and compared. For negative binomial models Pseudo R-squared measures are more appropriate (Garay et al. 2011), I will therefore report McFadden's and Cox and Uhler's Pseudo R-squared. The AIC and BIC penalize the model for additional parameters. Additionally, the BIC controls for the number of observations. As can be seen in Table 2 all additional variables added did indeed improve model fit since the AIC and BIC decrease mostly and the pseudo R-squared increase.

The result of a VIF-test as well as the examination of Pearson correlation coefficients among all independent variables did not indicate multicollinearity between independent variables.

I estimated negative binomial models because they showed less overdispersion than poisson and zero-inflated-poisson (as can bee seen in Table 5 in the appendix). I additonally estimated a zero-inflated model since it showed an improved AIC and BIC (see Table 5). The models are still slightly overdispersed which needs to be considered when interpreting results since this leads to more signifcant relationships in the data than there truly are.

For the probability model, the distance component is not significant. This might indicate that excess zeros are not created after all but are resulting from certain combinations of parameters in the model.

To evaluate how different factors influence bikesharing usage five models were estimated. Model 1 is a basic gravity model not incorporating additional mass values or deterrence factors apart from population and distance. Both factors are in line with theory and stay at similar values across models. A higher population indicates increased bikeshare usage whereas an increased distance strongly decreases it. The null-hypothesis 1 and 5 can therefore not be rejected. This strongly supports the gravity model approach as well, since this simple gravity model already shows an acceptable pseudo R-squared of 0.23. All other variables added in the following models only add a small improvement to model 1.

For the remaining models control variables for a station being reachable within the free of cost 30 minutes, loop trips, distance to other modes of transport and other bike share stations and to city center and weekends were introduced. Their coefficients change only slightly across models. There are less trips on weekends and less loop trips which agrees with the descriptive analysis of the data. Reachable within 30 minutes has a slight positive effect. Since it was slightly correlated with the distance measure this needs to be interpreted with care. The distance to city center has a negative effect on bikeshare usage in model 1, this disappears, however, once land use is added to the model. Reason for this could be the negative correlation between residential landuse and distance to city center. Other bikeshare stations within the vicinity of a station decrease bike usage which supports previous findings of competition between stations. Other transport modes closeby increase bike share usage which agrees with previous research that attributed this effect to increased usage through mode share (Shaheen, Cohen, and Martin 2013).

Model 2-4 evaluate the factors that are focus of this paper which are number of points of interests of different categories and percentage of landuse categories in a station's service area as well as the share of bike paths among shortest bike routes between origin and destination station. I accumulativly added first the points of interest (POIs), then landuse and then bike infrastructure quality since robustness checks indicated that these combinations had most explanatory power. The sign of the effects of POIs and land use do not change across models and also magnitudes of coefficients only change slightly when the additional variables are added. For all categories of POIs, the more of them are in close vicinity of a station the higher the bike usage at that station in both directions (origin stations and destination stations). This supports the activity-mobility theory and hypothesis four.

The results of land use do not quite respond to hypothesis 2. All categories have a negative effect which can have multiple reasons. From a theoretical point of view it is possible that land use simply does not have an effect on daily aggregates of bike share usage. Previous research considered temporal variations during the day for their studies, here I only used daily aggregates (Ren et al. 2014; Demissie, Phithakkitnukoon, and Kattan 2019). Another possible explanation could lie in variable misspecifications. For example, the percentage of residential landuse was positively correlated with distance to city centre since the outer areas are usually more inhabitated than the city center. It is therefore possible that the landuse variables simply depict a relationship of in city centre or not in city centre. Additionally, Hamburgs has quite a unique structure in its core since it is situated at a very large port and river as well as a large lake (the Alster). Many stations are in the city centre and close to these areas of water which is why the percentage of commercial landuse is quite often smaller than for example residential landuse in areas further away from the center, since a large portion of a station's service area is attributed to water. As can be seen in Figure 2, the percentage of areas with no landuse increase with smaller distance to city center.

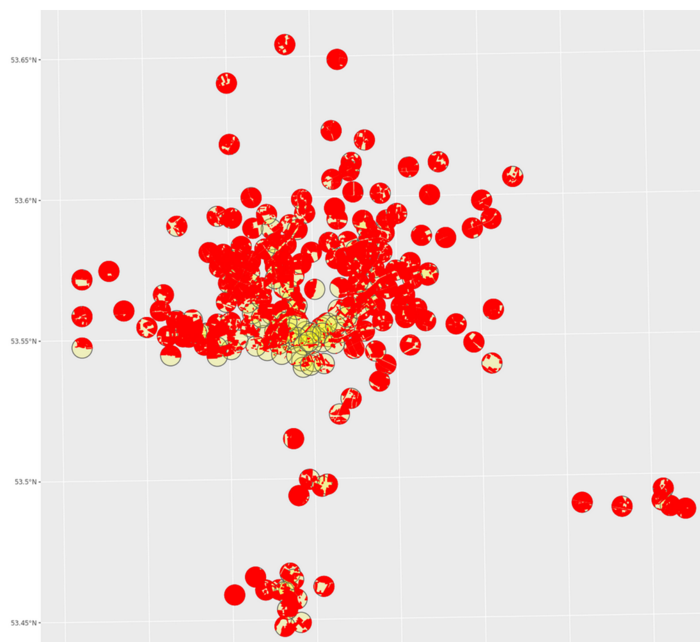

Figure 2: Map of intersecting landuse in 400 m service area of all stations across the city. Red designates areas with intersecting landuse, yellow no intersection (e.g. no landuse or water). One can clearly recognize how areas with a low percentage of land use are located closer to the city center.

As theory suggests, the percentage of cycle ways on a shortest route between stations positively impacts bike share usage. This is a first indication that better bike infrastructure does indeed increase bicycle or at least bikesharing usage which is in line with previous findings (El-Assi, Salah Mahmoud, and Nurul Habib 2017; K. Wang, Akar, and Chen 2018). One needs to keep in mind, however, that one does not know the true route a bikeshare user has taken with her bike between two stations.

Model 5 incorporates a socio economic factor into this analysis, namely the percentage of non-German citizens in the service area of a station. This is a showcase of how a social perspective can be incorporated in bikeshare demand modeling. For both origin and destination stations a higher percentage of non-German population decreases bikeshare usage. This finding is similar to Welsch, Conrad, and Wittowsky (2018) who found that persons with migrant background cycle less. What is interesting is that areas with higher percentage of non-Germans are located closer to the city center than other residential areas (see Figure 3). One would therefore have to examine further whether the decreased usage of shared cycles therefore results from a higher proximity to important places for daily activities or if it indeed results from a more cultural cause.

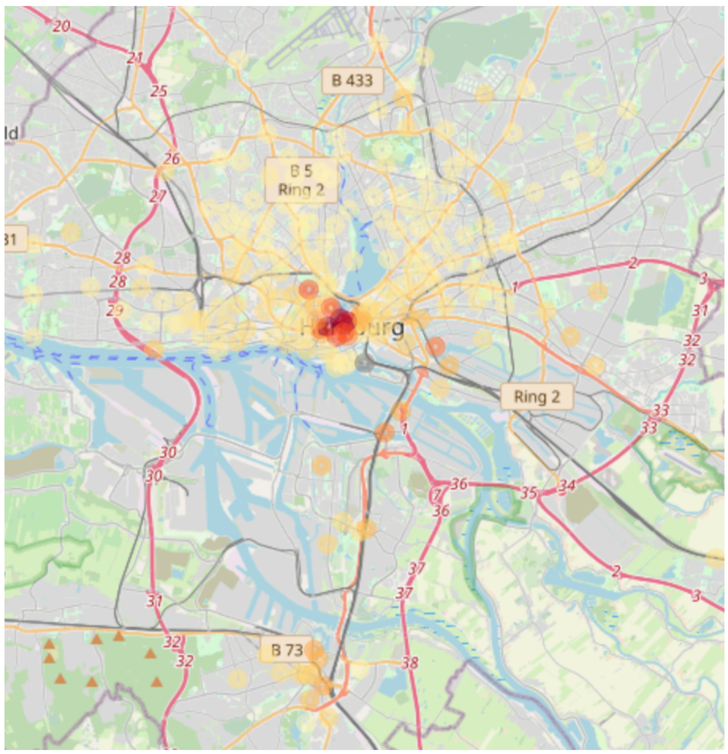

Figure 3: Map of percentage of non-German citizens in 400 m service area of all stations across the city. Red means high percentage, yellow low percentage. One can clearly recognize how areas with a high percentage are located closer to the city center.

To summarise, all the null-hypotheses of all hypotheses apart from H2 which concerns land use could be rejected. All variables that measured factors that can be influenced by city planners and are therefore relevant for the design of a successful bike share system had explanatory power and were in line with the theory suggested. In the next section I will discuss implications of these findings as well as their shortcomings.

Table 2: Zero-inflated negative binomial models.

	Dependent variable:				
	trip count				
	(1)	(2)	(3)	(4)	(5)
Route time (scaled)	-2.245^{***}	-2.912^{***}	-2.919^{***}	-2.932^{***}	-2.936^{***}
	(0.030)	(0.022)	(0.022)	(0.022)	(0.022)
O - population (scaled)	0.071^{***}	0.086^{***}	0.172^{***}	0.176^{***}	0.148^{***}
	(0.005)	(0.006)	(0.009)	(0.009)	(0.009)
D - population (scaled)	0.071^{***}	0.083^{***}	0.159^{***}	0.164^{***}	0.137^{***}
	(0.005)	(0.006)	(0.009)	(0.009)	(0.009)
O - POIs close transport (scaled)		0.026^{***}	0.022^{***}	0.018^{***}	0.012^{*}
		(0.006)	(0.006)	(0.006)	(0.007)
O - POIs close education (scaled)		0.032^{***}	0.028^{***}	0.030^{***}	0.028^{***}
		(0.005)	(0.005)	(0.005)	(0.005)
O - POIs close health (scaled)		0.048^{***}	0.039^{***}	0.042^{***}	0.041^{***}
		(0.005)	(0.005)	(0.005)	(0.005)
O - POIs close leisure (scaled)		0.079^{***}	0.031^{***}	0.030^{***}	0.077^{***}
		(0.007)	(0.008)	(0.008)	(0.010)
D - POIs close leisure (scaled)		0.098^{***}	0.054^{***}	0.054^{***}	0.097^{***}
		(0.007)	(0.008)	(0.008)	(0.009)
D - POIs close transport (scaled)		0.032^{***}	0.029^{***}	0.031^{***}	0.025^{***}
		(0.006)	(0.006)	(0.006)	(0.006)
D - POIs close education (scaled)		0.028^{***}	0.024^{***}	0.027^{***}	0.025^{***}
		(0.005)	(0.005)	(0.005)	(0.005)
D - POIs close health (scaled)		0.052^{***}	0.044^{***}	0.047^{***}	0.045^{***}
		(0.005)	(0.005)	(0.005)	(0.005)
D - distance to closest station (scaled)		0.139^{***}	0.173^{***}	0.171^{***}	0.171^{***}
		(0.010)	(0.011)	(0.011)	(0.011)
D - distance to closest station (scaled)		0.134^{***}	0.166^{***}	0.160^{***}	0.159^{***}
		(0.010)	(0.011)	(0.011)	(0.011)
O - distance to city center (scaled)		-0.056^{***}	-0.016	-0.012	-0.018
		(0.014)	(0.014)	(0.014)	(0.014)
O - distance to city center (scaled)		-0.023^{*}	0.012	0.019	0.017
		(0.014)	(0.014)	(0.014)	(0.014)
Route time < 30 min		0.063^{***}	0.058^{***}	0.056^{***}	0.049^{**}
		(0.021)	(0.021)	(0.021)	(0.021)
Loop trip		-0.229^{***}	-0.223^{***}	-0.251^{***}	-0.240^{***}
		(0.040)	(0.039)	(0.039)	(0.039)
Weekend		0.038^{***}	0.030^{**}	0.029^{**}	0.028^{**}
		(0.012)	(0.012)	(0.012)	(0.012)

O - landuse residential (scaled)			−0.164***	−0.161***	−0.151***
			(0.010)	(0.010)	(0.010)
D - landuse residential (scaled)			−0.148***	−0.144***	−0.135***
			(0.010)	(0.010)	(0.010)
O - landuse commercial (scaled)			−0.056***	−0.054***	−0.050***
			(0.006)	(0.006)	(0.006)
D - landuse commercial (scaled)			−0.051***	−0.047***	−0.043***
			(0.006)	(0.006)	(0.006)
O - landuse retail (scaled)			−0.066***	−0.067***	−0.064***
			(0.009)	(0.009)	(0.009)
D - landuse retail (scaled)			−0.060***	−0.064***	−0.062***
			(0.009)	(0.009)	(0.009)
O - landuse industrial (scaled)			−0.066***	−0.065***	−0.054***
			(0.007)	(0.007)	(0.007)
D - landuse industrial (scaled)			−0.062***	−0.061***	−0.051***
			(0.007)	(0.007)	(0.007)
O - citizenship not German (scaled)					−0.066***
					(0.007)
D - citizenship not German (scaled)					−0.061***
					(0.007)
Percentage cycle way (scaled)				0.073***	0.078***
				(0.006)	(0.006)
Constant	−2.558***	−3.561***	−3.563***	−3.553***	−3.558***
	(0.039)	(0.015)	(0.015)	(0.015)	(0.015)
Probability model					
Route time	-5.345−	−5.250	−5.198	−5.173	−5.175
Constant	−16.852	−16.852	−17.343	−17.692	−17.585
	(0.067)	(198.891)	(260.611)	(311.294)	(297.516)
AIC	255144.0	254430.9	253713.2	253552.0	253373.5
BIC	255218.2	254664.2	254031.3	253880.7	253723.3
McFadden's Pseudo R^2	0.232	0.234	0.236	0.237	0.237
Cox and Uhler's Pseudo R^2	0.228	0.230	0.232	0.233	0.233
Dispersion factor	1.939	1.926	1.858	1.805	1.818
Observations	297,052	297,052	297,052	297,052	297,052
Note:	*p<0.1; **p<0.05; ***p<0.01				

8 Conclusion and Discussion

To plan cities of the future, successful bike sharing systems are key. Success denotes here not only a high usage but also accessibility to all groups of society. City planners have been lacking a sound data-driven approach for decision making in the past decades and research has only recently started to make use of large amounts of data available from bike sharing systems. So far, a theoretic foundation is lacking in this field. This paper has been able to address these issues by developing a spatial interaction model of bike sharing which is for the first time based on theoretic assumptions. It also contributes a socioeconomic perspective on bike sharing for the first time.

The analysis was based on the theory of perceived value and modeled a zero-inflated negative binomial regression model based on the gravity model. The results indicate support for the theory of perceived functional and conditional value since a higher percentage of bike paths between stations, a higher number of important points of interest increased usage whereas an increased distance decreased usage. Further research should examine how social and emotional values influence adoption of bike sharing.

The gravity model has shown to have strong explanatory power on its own as well, since, in its simplest form with only three factors included, it already performs very close to more complex models.

Now what implications do the findings have on how urban planners can design a successful and inclusive bike sharing scheme?

Good bike infrastructure is indeed a must. The higher the percentage of bike paths along a potential route taken between stations the higher bike share usage. Moreover, increased distance between stations is strongly demotivating bike share usage. Placing additional stations as well as a development policy of placing important points of interests more spread over the city could increase how much bike sharing is used. And lastly, since higher population increases bike share demand, stations should not only be placed in the city center but also in more populated residential areas. Concerning inclusivity, further research here is definitely necessary. The approach of this paper showed a lack of cycling in areas with a higher migrant population which is in line with former research (Welsch, Conrad, and Wittowsky 2018). Municipalities should concentrate therefore on promoting cycling in such communities since this lack of bicycle usage does not seem to result from a lack of available stations. Due to unavailability of additional data, a further examination of this mechanism is at this point probably still subject to surveys. This analysis has shown nevertheless that socioeconomic factors can be incorporated into such modeling attempts to a certain point. Given the serious consequences an unequal access to transport can have on a community's participation in society, it is vital that this topic is further examined. To this day, the socioeconomic structures in the spatial context are still too widely ignored in urban planning (Welsch, Conrad, and Wittowsky 2018).

Given the contributions this paper makes it does have some shortcomings that will be discussed in the following. It firstly has to be noted that bike sharing usage is an incredibly complex phenomenon that is influenced by many factors that are themselves intercorrelated.

The availability of vast amounts of data about bike sharing mobility flows in cities help to gain valuable insight into cyclist's movements across the city. It depicts, however, only a fraction of the actual bicycle mobility. Further research is therefore necessary in order to evaluate in what ways private bicycle use differs from shared bicycles. Additionally, station-based bike sharing poses a special situation since users are bound to specific geographic start and end points.

Moreover, bike sharing data sets are subject to endogeneity bias which states that the error term is related to parameters in the model. Faghih-Imani and Eluru (2016) found that bike sharing scheme planners tend to place stations in geographical areas where they expect much usage to occur and that bike sharing usage is correlated to the size of the network. PBS usage is therefore not a true reflection of mobility patterns in the city but dependent on the design of the system. Causality relationships are therefore very difficult to establish. This also implies that one cannot include areas that have no stations at all. These might share common properties that are excluded from analysis since they simply do not appear in the data. Additionally, a self-selection bias of bike sharing users is likely to be observed which is why bike sharing travel flows are not entirely representative of the true population's mobility patterns (Faghih-Imani and Eluru 2016). Schauder and Foley (2015), for example, examined how differing health leads individuals to make different transport mode choices and Walker, Ehlers, and Banerjee (2011) found a relationship between transport mode choice and social influences and taste preference.

Since there is no additional socioeconomic ridership information available one can derive information solely from the station service area. This paper has shown that there can still be some very useful insights generated from this. It poses, however, some limitations on establishing causality, since the researcher can never be sure of the true socioeconomic status of bike share users. These issues show that the bike sharing data available nowadays is not sufficient to make claims about causality for actual bike usage in cities. There are still important issues, especially concerning a social perspective on cycling and bike sharing that cannot be solely addressed by mining big data from bike sharing systems. Traditional research approaches such as surveys still have many advantages here that should be incorporated in future research. Nevertheless, the advantages of incorporating analysis results of large data remains very useful since it still generates helpful insights at a comparably low cost.

Methodologically, the complex nature of bike sharing data incorporates some factors whose examination was beyond the scope of this paper. For example, previous studies have noted that spatial auto correlation (Cheng et al. 2019) as well as a multilevel structure (Hyland et al. 2018; Scott and Ciuro 2019) might be present in the data. Additionally, other distance decay functions than the linear one used in this model have been suggested in other gravity model approaches (Halás, Klapka, and Kladivo 2014).

For further research it would be also be interesting to examine mechanisms of a spatial interaction model for bike sharing when including a temporal component. Bike sharing usage has been found to vary during the course of a day with peaks in morning and evening hours for weekday commutes and a midday peek on weekends (Hyland et al. 2018; Fishman 2015). It was sufficient in this scenario to look at aggregates on day-level, still it is interesting to examine what mechanisms lie beneath a temporal variation in trips. The effect of land use could be very different and deliver more answers than it did in this paper. Moreover, there are important additional factors that could be considered in further research, namely properties of the system itself. These are factors that could influence the functional value of using a shared bike as Y. Wang et al. (2018) have suggested. This includes information on bike availability at stations, the condition of bikes used and pricing schemes.

To conclude, this paper made multiple contributions by developing a full spatial interaction model based on theory for the first time. It was applied to an example bike sharing system in Germany, a country where research on bike sharing systems is still scarce. Findings of the analysis were mostly consistent with previous research which allowed it to derive guidance for urban planners on how to design a successful and inclusive bike sharing system. One needs to keep in mind, however, that anonymous data of that size poses some serious shortcomings on deriving insights generalizable to the whole population of a city, different cities or cycling in general. Thus, in order to design a successful bike sharing system, one cannot solely rely on results from an aggregated analysis but should also consider other approaches such as surveys or detailed case studies.

9 References

Axhausen, Kay, and Tommy Gärling. 1992. "Activity-based approaches to travel analysis Conceptual frameworks, models and research problems." *Transport Reviews* 12 (4): 323–41.

Baltagi, Badi, Peter Egger, and Michael Pfaffermayr. 2003. "A generalized design for bilateral trade flow models." *Economics Letters* 80 (3): 2019.

BMVB. 2012. "Nationaler Radverkehrsplan 2020: den Radverkehr gemeinsam weiterentwickeln," 82S. https://trid.trb.org/view/1253654.

Borgnat, Pierre, Céline Robarrdet, Patrice Abry, Patrick Flandrin, and Jean-Baptiste Rouquier. 2012. "A Dynamical Network View of Lyon's Vélo' v Shared Bicycle System." In *Dynamics on and of Complex Networks: Applications to Time-Varying Dynamical Systems*, 2:159–200.

"Call A Bike Reportingdaten, Deutsche Bahn AG." 2020. Accessed November 16. https://data.deutschebahn.com/dataset/data-call-a-bike.

Cheng, Lyu, Wang Yunshan, Xia Yan, Liua Zhiyuan, and Wang Wei. 2019. "Spatial Analysis of the Relationship between Urban Built Environment and Public Bicycle Ridership." In *CICTP*, 5948–60.

"CycleStreetsAPI." 2020. Accessed June 12. https://www.cyclestreets.net/.

Demissie, Merkebe Getachew, Santi Phithakkitnukoon, and Lina Kattan. 2019. "Trip Distribution Modeling Using Mobile Phone Data: Emphasis on Intra-Zonal Trips." *IEEE Transactions on Intelligent Transportation Systems* 20 (7): 2605–17.

Duran-Rodas, David, Emmanouil Chaniotakis, and Constantinos Antoniou. 2019. "Built Environment Factors Affecting Bike Sharing Ridership: Data-Driven Approach for Multiple Cities." *Transportation Research Record*.

El-Assi, Wafic, Mohamed Salah Mahmoud, and Khandker Nurul Habib. 2017. "Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto." *Transportation* 44 (3). Springer US: 589–613.

Faghih-Imani, Ahmadreza, and Naveen Eluru. 2016. "Determining the Role of Bicycle Sharing System Infrastructure Installation Decision on Usage: Case Study of Montreal BIXI System Ahmadreza." *Transportation Research Part A: Policy and Practice* 94: 685–98.

Faghih-Imani, Ahmadreza, Naveen Eluru, Ahmed M. El-Geneidy, Michael Rabbat, and Usama Haq. 2014. "How does land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (BIXI) in Montreal." *Transport Geography* 41: 306–14.

Fishman, Elliot. 2015. "Bikeshare: A Review of Recent Literature." *Transport Reviews* 36 (1). Taylor & Francis: 92–113.

Garay, Aldo M., Elizabeth M. Hashimoto, Edwin M. M. Ortega, and Víctor H. Lachos. 2011. "On estimation and influence diagnostics for zero-inflated negative binomial regression models." *Computational Statistics and Data Analysis* 55 (3): 1304–18. doi:10.1016/j.csda.2010.09.019.

"Geofabrik Download Server." 2020. Accessed February 28. https://download.geofabrik.de/europe/germany.html.

Grieco, Margaret, and John Urry. 2011. *Mobilities: new perspectives on transport and society*. Edited by Margaret Grieco and John Urry. Ashgate Publishing Limited.

Halás, Marián, Pavel Klapka, and Petr Kladivo. 2014. "Distance-decay functions for daily travel-to-work flows." *Journal of Transport Geography* 35: 107–19.

Hampshire, Robert C, and Marla Lavanya. 2012. "An Analysis of Bike Sharing Usage: Explaining Trip Generation and Attraction from Observed Demand." *TRB Annual Meeting*, 1–17.

Hayes, Simon, Christel Frühauf Martin, Janett Büttner, Hendrik Mlasowsky, and Tim Birkholz. 2011. "Optimising Bike Sharing in European Cities - Ein Handbuch."

Hyland, Michael, Zihan Hong, Helen Karla Ramalho de Farias Pinto, and Ying Chen. 2018. "Hybrid cluster-regression approach to model bikeshare station usage." *Transportation Research Part A: Policy and Practice* 115. Elsevier: 71–89. http://dx.doi.org/10.1016/j.tra.2017.11.009.

Khan, Saeed Farooq, Neil Bergmann, Raja Jurdak, Branislav Kusy, and Mark Cameron. 2017. "Mobility in cities: Comparative analysis of mobility models using Geo-tagged tweets in Australia." *2017 IEEE 2nd International Conference on Big Data Analysis, ICBDA 2017*, 816–22. doi:10.1109/ICBDA.2017.8078751.

Lai, Wen Tai, and Ching Fu Chen. 2011. "Behavioral intentions of public transit passengers-The roles of service quality, perceived value, satisfaction and involvement." *Transport Policy* 18 (2). Elsevier: 318–25. doi:10.1016/j.tranpol.2010.09.003.

Lenormand, Maxime, Aleix Bassolas, and José J. Ramasco. 2016. "Systematic comparison of trip distribution laws and models." *Journal of Transport Geography* 51: 158–69. doi:10.1016/j.jtrangeo.2015.12.008.

Levy, Nadav, Chen Golani, and Eran Ben-Elia. 2017. "An exploratory study of spatial patterns of cycling in Tel Aviv using passively generated bike-sharing data generated bike-sharing data Abstract Keywords." *Big Data and Transport Geography*, 1–24.

Li, Fangzhou, Zhiming Feng, Peng Li, and Zhen You. 2017. "Measuring directional urban spatial interaction in China: A migration perspective." *PLoS ONE* 12 (1): 1–19. doi:10.1371/journal.pone.0171107.

Lin, Pengfei, Jiancheng Weng, Quan Liang, Dimitrios Alivanistos, and Siyong Ma. 2019. "Impact of Weather Conditions and Built Environment on Public Bikesharing Trips in Beijing." *Networks and Spatial Economics*. Networks; Spatial Economics.

Masucci, A. Paolo, Joan Serras, Anders Johansson, and Michael Batty. 2013. "Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows." *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 88 (2): 1–8. http://arxiv.org/abs/1206.5735.

Mateo-Babiano, Iderlina, Richard Bean, Jonathan Corcoran, and Dorina Pojani. 2016. "How does our natural and built environment affect the use of bicycle sharing?" *Transportation Research Part A: Policy and Practice* 94: 295–307.

Mattson, Jeremy, and Ranjit Godavarthy. 2017. "Bike share in Fargo, North Dakota: Keys to success and factors affecting ridership." *Sustainable Cities and Society* 34 (July): 174–82.

Noland, Robert B., Michael J. Smart, and Ziye Guo. 2018. "Bikesharing Trip Patterns in New York City: Associations with Land Use, Subways, and Bicycle Lanes." *International Journal of Sustainable Transportation* 13 (9). Taylor & Francis: 664–74. https://doi.org/10.1080/15568318.2018.1501520.

Patuelli, Roberto, and Giuseppe Arbia. 2016. *Spatial econometric interaction modelling: Where spatial econometrics and spatial interaction modelling meet.* 9783319301945.

Pucher, John, Jennifer Dill, and Susan Handy. 2010. “Infrastructure, programs, and policies to increase bicycling: An international review.” *Preventive Medicine* 50 (SUPPL.). Elsevier Inc.: S106–S125. http://dx.doi.org/10.1016/j.ypmed.2009.07.028.

“Radverkehrsentwicklung 2020 - hamburg.de.” 2020. Accessed April 15. https://www.hamburg.de/bwvi/medien/13460228/radverkehrsentwicklung-2020/.

Ren, Mengyao, Yaoyu Lin, Meihan Jin, Zhongyuan Duan, Yongxi Gong, and Yu Liu. 2019. “Examining the effect of land-use function complementarity on intra-urban spatial interactions using metro smart card records.” *Transportation*, no. 0123456789. Springer US. doi:10.1007/s11116-019-09977-7.

Ren, Yihui, Mária Ercsey-Ravasz, Pu Wang, Marta C. González, and Zoltán Toroczkai. 2014. “Predicting commuter flows in spatial networks using a radiation model based on temporal ranges.” *Nature Communications* 5. doi:10.1038/ncomms6347.

Schauder, Stephanie A, and Mark C Foley. 2015. “The relationship between active transportation and health.” *Journal of Transport & Health*. Elsevier, 1–7. doi:10.1016/j.jth.2015.06.006.

Scheiner, Joachim, and Christian Holz-Rau. 2007. “Travel mode choice: Affected by objective or subjective determinants?” *Transportation* 34 (4): 487–511. doi:10.1007/s11116-007-9112-1.

Scott, Darren M., and Celenna Ciuro. 2019. “What factors influence bike share ridership? An investigation of Hamilton, Ontario’s bike share hubs.” *Travel Behaviour and Society* 16 (July 2018). Elsevier: 50–58. https://doi.org/10.1016/j.tbs.2019.04.003.

Shaheen, Susan, Adam Cohen, and Elliot Martin. 2013. “Public bikesharing in North America.” *Transportation Research Record*, no. 2387: 83–92.

Stefanouli, Maria, and Serafeim Polyzos. 2017. “Gravity vs radiation model: two approaches on commuting in Greece.” *Transportation Research Procedia* 24. Elsevier B.V.: 65–72. http://dx.doi.org/10.1016/j.trpro.2017.05.069.

Taylor, P J, and G Catalano. 2002. “Measurement of the World City Network.” *Urban Studies* 39 (13): 2367–76.

“Transparenzportal Hamburg.” 2020. Accessed April 15. http://transparenz.hamburg.de/.

Tsiotas, Dimitrios, George Aspridis, Ioannis Gavardinas, Labros Sdrolias, and Dagmar Škodová-Parmová. 2019. “Gravity modeling in social science: the case of the commuting phenomenon in Greece.” *Evolutionary and Institutional Economics Review* 16 (1). Springer Japan: 139–58. doi:10.1007/s40844-018-0120-y.

Walker, Joa, Emily Ehlers, and Ipsita Banerjee. 2011. “Correcting for endogeneity in behavioral choice models with social influence variables” 45: 362–74.

Wang, Kailai, Gulsah Akar, and Yu Jen Chen. 2018. “Bike sharing differences among Millennials, Gen Xers, and Baby Boomers: Lessons learnt from New York City’s bike share.” *Transportation Research Part A: Policy and Practice* 116 (June). Elsevier: 1–14. https://doi.org/10.1016/j.tra.2018.06.001.

Wang, Xize, Greg Lindsey, Jessica E. Schoner, and Andrew Harrison. 2016. “Modeling bike share station activity: Effects of nearby businesses and jobs on trips to and from stations.” *Journal of Urban Planning and Development* 142 (1): 1–9.

Wang, Yacan, Matthew A. Douglas, Benjamin Thomas Hazen, and Martin Dresner. 2018. “Be green and clearly be seen: How consumer values and attitudes affect adoption of bicycle sharing." *Transportation Research Part F: Traffic Psychology and Behaviour* 58. Elsevier Ltd: 730–42. doi:10.1016/j.trf.2018.06.043.

Welsch, Janina, Kerstin Conrad, and Dirk Wittowsky. 2018. "Exploring immigrants travel behaviour : empirical findings from Offenbach am Main , Germany." *Transportation* 45 (3). Springer US: 733–50. doi:10.1007/s11116-016-9748-9.

Wu, Jiansheng, D Ph, Luyi Wang, Weifeng Li, and D Ph. 2018. "Usage Patterns and Impact Factors of Public Bicycle Systems : Comparison between City Center and Suburban District in Shenzhen." *Journal ofUrban Planning and Development* 144 (3): 1–17.

Yang, Yang, Dong Li, and Xiang (Robert) Li. 2019. "Public Transport Connectivity and Intercity Tourist Flows." *Journal of Travel Research* 58 (1): 25–41.

Yao, Yi, Yifang Zhang, Lixin Tian, Nianxing Zhou, Zhilin Li, and Minggang Wang. 2019. "Analysis of network structure of urban bike-sharing system: A case study based on real-time data of a public bicycle system." *Sustainability (Switzerland)* 11 (19): 12–19.

Zaltz Austwick, Martin, Oliver O'Brien, Emanuele Strano, and Matheus Viana. 2013. "The Structure of Spatial Networks and Communities in Bicycle Sharing Systems." *PLoS ONE* 8 (9). doi:10.1371/journal.pone.0074685.

Zeithaml, Valarie A. 1988. "Consumer Perceptions of Price, Qualitye." *Journal of Marketing* 52 (3): 2–22. doi:10.2307/1251446.

"Zensus 2011 - Ergebnisse zum Download." 2020. Accessed March 15. https://www.zensus2011.de/DE/Home/Aktuelles/DemografischeGrunddaten.html.

Zhang, Lianfa, Jianquan Cheng, and Cheng Jin. 2019. "Spatial interaction modeling of OD flow data: Comparing geographically weighted negative binomial regression (GWNBR) and OLS (GWOLSR)." *ISPRS International Journal of Geo-Information* 8 (5).

Zhang, Shuichao, Zhuping Zhou, Haiming Hao, and Jibiao Zhou. 2018. "Prediction model of demand for public bicycle rental based on land use." *Advances in Mechanical Engineering* 10 (12): 1–8.

Zhang, Ying, Tom Thomas, Mark Brussel, and Martin van Maarseveen. 2017. "Exploring the impact of built environment factors on the use of public bikes at bike stations: Case study in Zhongshan, China." *Journal of Transport Geography* 58. Elsevier Ltd: 59–70. http://dx.doi.org/10.1016/j.jtrangeo.2016.11.014.

Appendix

	transport	education	health	leisure	leisure.1	leisure.2	leisure.3
1	railway_station	uni	pharmacy	nightclubs	sports	restaurant	shopping
2	railway_halt	school	hospital	restaurants	sports_centre	fast_food	supermarkets
3	tram_stop	kindergarten	doctor	library	swimming_pool	cafe	malls
4	bus_stop	college	dentist	arts_centre	tennis	pub	kiosk
5	bus_station			market_place	golf	bar	bakery
6	airport			theatre	stadium	biergarten	bank
7	ferry_station			nightclub	cinema	park playground	

Table 3: List of all point of interest categories used in aggregated POI-variables.

	Model	AIC	BIC	R2McF	R2CU	disp
1	Poisson	283708.42	284015.87	0.36	0.42	2.03
2	Poisson (zero_inflated)	263846.83	264175.48	0.25	0.25	5.04
3	Negative Binomial	253729.31	254036.76	0.24	0.23	1.79
4	Negative Binomial (zero-inflated)	253552.03	253880.68	0.24	0.23	1.81

Table 4: Model performance indicators for poisson and negative binomial, with zero-inflation and wihtout.