Data-Driven Investigation of Spatial Patterns in the Usage of Shared Micro Mobility Vehicles

Masterarbeit

zur Erlangung des akademischen Grades „Master of Science (M. Sc.)“ im Studiengang Wirtschaftswissenschaft der Wirtschaftswissenschaftlichen Fakultät der Leibniz Universität

vorgelegt von

Name: Bäßmann Vorname: Felix Niklas

[redacted] [redacted]

Prüfer: Prof. Dr. M. H. Breitner

Hannover, den 30.09.2021

Contents

FIGURE INDEX 2

TABLE INDEX 4

INDEX OF ABBREVIATIONS...5

SYMBOL INDEX 6

ABSTRACT 7

1. INTRODUCTION INTO URBAN MOBILITY RESEARCH 7

1.1. MOTIVES FOR INVESTIGATING MOBILITY DATA 7
1.2. LITERATURE REVIEW ON THE USAGE OF MOBILITY DATA IN DIFFERENT SETTINGS 8
1.3. LITERATURE ON TRIP DISTRIBUTION LAWS 10

2. CASE STUDY ON TRIP DISTRIBUTION LAWS 12

2.1. FORMAL INTRODUCTION OF TRIP DISTRIBUTION LAWS 12
2.2. MODEL CHOICE FOR PREDICTING ORIGINS AND DESTINATIONS OF TRIPS...13
2.3. APPLYING THE GRAVITY AND RADIATION MODEL 14

3. IDENTIFYING IMPACT VARIABLES ON E-SCOOTER MOVEMENTS 21

3.1. DESCRIPTION OF DATA INVESTIGATION APPROACH 21
3.2. EXPLORING PUBLIC TRANSPORT DATA...22
3.2.1. DATA SOURCE FOR PUBLIC TRANSPORT DATA 22
3.2.2. CLASSIFICATION OF STOPS...23
3.2.3. COUNTING ARRIVALS AND DEPARTURES OF PUBLIC TRANSPORT (DENSITY)...26
3.2.4. CLASSIFICATION OF TRANSPORT (DIVERSITY) 28
3.3. MATCHING E-SCOOTER TRIPS TO PUBLIC TRANSPORT LOCATIONS 31
3.3.1. DATA SOURCING FOR E-SCOOTER TRIPS 31
3.3.2. DESCRIPTIVE ANALYSIS OF E-SCOOTERS TRIPS AT PTLS 34
3.3.3. CORRELATION ANALYSIS BETWEEN E-SCOOTERS AND DENSITY AND DIVERSITY AT PTLS 36
3.4. EXPANDING THE MODEL THROUGH BUILT ENVIRONMENT, DEMOGRAPHIC AND SOCIETAL DETERMINANTS..39
3.4.1. IDENTIFYING POTENTIAL IMPROVEMENTS...39

3.4.2. DETERMINANTS WITH COARSE SPATIAL RESOLUTION 40
3.4.3. DETERMINANTS WITH HIGH SPATIAL RESOLUTION 42
3.5. EXPANDING THE MODEL BY SOCIAL MEDIA INPUT 46
3.5.1. INTRODUCTION AND EVALUATION OF SOCIAL MEDIA 46
3.5.2. APPLICATION OF SOCIAL MEDIA ON SIGHTS 48
3.5.3. APPLICATION OF SOCIAL MEDIA ON SUBDISTRICTS 49
3.5.4. CRITICAL EVALUATION ON SIGHTS AND INSTAGRAM ENRICHMENT 52
3.6. INCORPORATING THE PARAMETERS INTO THE LINEAR MODEL 53
3.7. UTILIZING MACHINE LEARNING TO ESTIMATE E-SCOOTER MOVEMENTS 55
3.7.1. MACHINE LEARNING THEORY 55
3.7.2. DESIGNING A DEEP LEARNING NETWORK TO ESTIMATE E-SCOOTER MOVEMENTS 57
3.7.3. FITTING THE NETWORK TO ESTIMATE E-SCOOTER MOVEMENTS 60
3.7.4. COMPARING MACHINE LEARNING MODELS 61
3.7.5. SPATIAL DESCRIPTION OF DLN PREDICTIONS 64
3.7.6. INTERPRETATION AND CRITICISM 65

4. APPLY FINDINGS ON E-SCOOTER MOVEMENTS ON ORIGIN AND DESTINATION ESTIMATION .. 67

4.1. DEFINING A BASELINE FOR COMPARISON WITH PREVIOUS FINDINGS 67
4.2. ORIGIN-DESTINATION ESTIMATION WITH NEW PARAMETERS 69

5. DISCUSSION OF RESULTS 74

6. LIMITATIONS 76

7. SUMMARY 77

LITERATURE INDEX 79

APPENDIX 87

Figure Index

Figure 1: Berlin and the Tier operating area as a black-lined polygon. 15
Figure 2: Grid for Berlin with population data and Tier operating area. 16
Figure 3: Selection of origins for e-scooter trips, close to or at (from top left to top right): S+U Potsdamer Platz, Husemannstr. (Berlin), U Ernst-Reuter-Platz, Graefestr. (Berlin). White “O” signals the starting cell. 19

ABSTRACT

Improvements in the availability, quality and quantity of mobility data led to a comeback of trip distribution laws and the exploration of determinants of mobility. While there is broad literature on long and medium distance trips and available models to describe these patterns, not so much research exists for short distance e-scooter sharing trips. While there is some on demand of these rental vehicles, explanations for trip patterns are scarce. This thesis aims to fill this gap by investigating patterns and quantifiable input variables which identify locations with high usage of e-scooter sharing services and use the generated information to explore their impact on the origin and destination of the trips taken by e-scooter users as well as their probabilities. To do so, linear and nonlinear regression as well as machine learning techniques are used to build descriptive and inferential models. The findings show how e-scooter usage interacts with the built environment and it can also help to better understand intra-city Micromobility, especially from e-scooters. The insights can be used to build user-orientated infrastructure and increase fleet efficiency.

Key words: Micromobility, E-Scooter, O-D Trips, Machine Learning, Urban Planning

1. Introduction into Urban Mobility Research

1.1. Motives for Investigating Mobility Data

With the increased digitization of all services, users leave behind data traces about their behaviour. This is also true for mobility services and journeys undertaken by individuals. This led to increased research in the investigation and modelling of origins and destinations of travellers. This momentum is mainly driven by improvements in the underlying data quality and quantity. During most of the 20^{th} century surveys, aggregated and sampled data was used. These sources have nowadays been replaced by data about trips from millions of individuals with a spatial resolution of a few meters. This allows for new attempts to make generalized statements and improve theories and models.

By undertaking this research, the understanding of mobility patterns, especially origins and destinations and traffic flow can be improved and thus, leading to more efficient land usage in cities. Global migration patterns cause cities to grow larger and the space in human agglomerations to become more desirable, therefore efficient usage will be key, considering that the share of people living in cities being expected to grow by 700 million between 2021 and 2030 (deStatis, 2018) to 60 percent of the global population (UN, 2018). Taking for example the four German cities with more than one million inhabitants (Berlin, Hamburg, Munich and Cologne), traffic infrastructure accounts for 21.55 up to 26.73 percent of the built-up area with the rest consisting of residential space (Statistische Ämter des Bundes und der

Länder, 2019). Therefore, this thesis aims to add to the efficient planning of traffic flows based on mobility data which pursues the goal to achieve higher efficiency in urban land use.

1.2. Literature Review on the Usage of Mobility Data in Different Settings

There are various different paths of working with intracity mobility data. Saberi et al. (2017) built a network graph to analyse origin-destination demand networks and additionally did a literature review on previous work and data sources for intra urban mobility. Amongst others they name Liang et al. (2012) who used GPS data from taxis in Beijing to estimate the spatial scaling of mobility. This type of source data was also used by Gao et al. (2013) to evaluate travel times in the city over time for more efficient trip planning. A methodological approach on fundamental concepts regarding the incorporation of social media data in urban computation was done by Silva et al. (2019). With regard to public transport Hasan et al. (2013) investigate the first, second and third highest ranking trip destination from 262 smart card users in London, which is used to check-in and pay for public transport fares at every origin and destination of a journey.

Besides this research in established means of urban transport, during the 21st century new mobility services have emerged. For short journeys the segment of Micromobility providers has seen a lot of interest from investors (McKinsey, 2019). This segment consists mainly of bike and e-scooter sharing services (Roland Berger, 2020). Nowadays these come usually as dockless sharing method, which refers to the method that a rented vehicle can be picked up and left almost anywhere inside the business area of the service provider (Tier, 2021). In the past sharing schemes with fixed docking locations were used for renting and returning the vehicle (stadtRAD Hamburg, 2021). With dockless systems the questions arise on where, when and how much vehicles should be deployed in an area.

With regard to bike sharing research into these variables has been carried out. A variety of studies exists which investigate determinants of customer demand, usage patterns and deployment strategies. Mooney et al. (2019) for example find higher demand and higher availability of dockless bikes in areas with higher income and better education.

Trana et al. (2015) find altitude, capacity, network density, population, railway stations, student residences, cinemas and restaurants to have an impact on the frequency of movements at docking stations for bike sharing in Lyon. Shen et al. (2018) investigated 19 different possible impact factors on dockless bike sharing for grid cells in Singapore. This was based on various underlying ideas, i.e. weather conditions, population or public transport variables. With regard to spatial parameters, the highest significance codes could be found for the availability of dockless bikes in the cells that had the biggest sum of length of cycling paths

and accessibility of bike racks, the share of land use of commercial and industrial buildings, the distance to Mass Rapid Transport stations as well as the distance to the central business district of the city. A similar investigation was conducted by Faghih-Imani et al. (2014) on 25 parameters in Montreal with respect to docking stations. They investigated the infrastructure in terms of the density of bike sharing stations around one station, major or minor roads close by as well as characteristics of the build environment like metro stations, restaurants, universities and job offerings in a range of 250 meters around the bike sharing station.
Yang et al. (2019) studied the impact of the opening of a new metro line in Nanchang, China on the trip patterns and distribution of dockless patterns and found a significant shift in origins and destinations of journeys. Li et al. (2020) investigated the average time a bike of a dockless sharing scheme remained unused between two bookings. As spatial resolution they used the census areas of Shanghai. They used Ordinary Least Squares (OLS) regression as well as Geographically Weighted Regression (GWR) to estimate the impact of population, public transport and points of interest in the census area on the dependent variable. For public transport they estimated an impact area of the subway and bus stops and calculated the relative share of impact areas to the seize of the census cell as well as using data on the number of travellers at each station. For population they used the density of this metric and for points of interest the relative share of restaurants, daily life services, residence and commercial facilities against each other. They found the highest significance on share of the subway impact area followed by the share of restaurants and daily life services. Furthermore, the rider numbers at subway stations and population density were also found to have significance levels below 0.1.

For e-scooters the amount of research is not as dense as for bike sharing. One starting point is the proposal from (Zhaoa, et al., 2021) on how to scrape and manipulate the data from Application Programming Interfaces (APIs) of e-scooter sharing providers to extract the underlying journeys executed by available e-scooters. There is some literature on temporal usage patterns like Bai and Jiao (2020) who analysed the patterns of e-scooter usage for Austin and Minneapolis where they identified high usage around downtown and the university area but with different temporal patterns. Furthermore, they could not make constant findings in both cities about land usage and green spaces and therefore highlighted to take spatial uniqueness into account when investigating mobility patterns. Bai et al. (2021) support the literature review by pointing out that only few studies provide insights on what activities drive travel demand for e-scooters. In their work to fill this gap they identify from top to bottom daily dining and drinking, shopping and recreational activities as highly correlated with e-scooter demand.

A different approach is undertaken by Lee et al. (2021) who built a complex model on predicting revenue for a potential e-scooter scheme in Manhattan and make assumptions on trips which are likely to be substituted by e-scooter journeys.
With regard to trips undertaken by e-scooter users and the routes they choose there is some related work that focus on the characteristics of these paths. For example, Zuniga-Garcia et al. (2021) found that e-scooter users use mainly roadways (33%), sidewalks (18%) and bike lanes (11%) with a 38% share of uncharacterised paths. Zhang et al. (2021) compliment these findings by conducting an investigation on what streets e-scooter riders prefer. They find users willing to take longer routes when they can ride on bikeways (59% longer paths), multi-use paths (29%), one-way roads (21%) or tertiary roads (15%).

1.3. Literature on Trip Distribution Laws

While one could argue, that the previously introduced articles offer an indication on where trips start by metrics like time to book (Li, et al., 2020) or high numbers of trip activity (Shen, et al., 2018), there is not too much research in high spatial resolution about the determinants on which basis the users pick a destination or the determinants describing where it will be located based on a fixed origin.

There are different theories on how destinations of journeys are spread across an area. These models are summed up under the name of trip distribution laws. The research into trip distribution laws can be simplified into two subgroups which try to describe mobility patterns of people across geographical regions, independently from the distance that is covered by the journeys. On the one hand there is mainly the Gravity Model and the Radiation Model for the party arguing that travel destinations are based on the population numbers and costs associated with the distance of the end points. The Gravity Model which was for a long time the leading framework in predicting population flows was hence applied on a variety of fields (Simini, et al., 2012). For example, it was used to describe the spread of Influenza (Li, et al., 2011) or global trade patterns and potentials (Batra, 2006). Extending the original notation Beiró et al. (2016) added social media traces to improve the prediction capabilities at which destinations people will arrive in. In an intracity context (Mazzoli, et al., 2019) applied the model on $1km \times 1km$ grid cells in London, Paris and other cities to estimate commuter flows. In comparison the original proposal of the Radiation Model was validated against data from commuting, intra-day mobility, call patterns and trade (Simini, et al., 2012).

On the other hand, used models are the Intervening Opportunities Model and the Rank-Based Model. These models state that the destination choice is driven by the options and opportunities that are available in between the start and at the possible destination. Therefore, if aggregated i.e. over all agents that travel, the mean distance travelled per journey will be shorter if the density of opportunities and alternative options is higher (Noulas, et al., 2012, p. 1). The argumentation of opportunity models comes with an understanding based on intuition, i.e. depending on where the closest supermarket is that suits the needs of the individual, it will drive longer or shorter and the higher the density of possible supermarkets, the easier the person will find what it needs close-by. The ideas of intervening opportunities and how they affect human migration and mobility for example, were tested initially in intracity migration, e.g. how likely it is for people to move from one cell to another based on the rent prices in the new area and comparable rent prices in cells closer to the previous place of living (Stouffer, 1940). Additionally, related work focused on intercity migration patterns (Galle & Taeuber, 1966). For traffic estimation like work and home commuting, inputs like job opportunities retail employment and population have been found to be good determinants, also called features (Clark & Peters, 1965). For the rank-based model, three studies could be found investigating the explanatory power of this trip distribution law: Noulas et al. (2012), Chen, Gao & Xiong (2017) as well as Santani & Gatica-Perez (2013). Each of these used data from location based social media networks to estimate the probability that after checking in into a certain location on the social network another one of the locations is visited afterwards and highlighted online. In these studies, the locations are always places like shops, restaurant or other consumer orientated businesses and contrary to the other trip distribution laws not bound to a geographic area but to a geographic point.

However, while some of these studies and models focus on urban mobility only very few of these are investigating Micromobility patterns. During the literature review one study could be found that has been published recently investigating the explanatory power of the Gravity Model on the origin-destination matrices of Bike Sharing services for a spatial resolution from $5km \times 5km$ to $500m \times 500m$ (Li, et al., 2021). With regard to e-scooters no study could be found to systemically investigate origin-destination (O-D) patterns and the determinants to describe and predict the destinations given a certain origin.

In order to extend the existing research on trip distribution laws and to complement the literature on the spatial determinants that affect e-scooter usage and the users' selection of origin and destinations, will this thesis first examine which features generate high and low numbers of e-scooter arrivals and departures in a fine special granularity and then use these findings to make statements on the origin and destination of customers as well as investigating the determinants that affect the probability that a certain route is chosen.

more emphasis on the design of the grid cells. A grid cell size in dependence of the distance to the origin could lead to finer results and investigation opportunities close to the start while reducing the number of zero observations further away from the origin cell.

A major limitation of the thesis are missing results from other municipalities, with regard to the already trained model as well as training a new model on the same features in a different city. Such potentially gained results would proof or disproof the statements beyond the case study of Berlin and would give answers on the generalisation capabilities of the model.

Another open research topic are the poor fitted O-D trajectories. There might be the potential to detect patterns in these routes where absolute values of residuals are high, i.e. if e-scooter are used relatively more often when public transport or other means of traffic offer poor connections to the surrounding area.

7. Summary

This thesis adds to the existing literature in applying trip distribution laws to e-scooter O-D trajectories as well as estimating determinants of traffic by means of OLS and machine learning algorithms which gives new insights about input features and further examination options for decision makers and researchers. Using trip distribution laws to describe e-scooter routes has to the best of the authors knowledge not been carried out before. Additionally, this new research was then expanded by public transport as well as built environment data.

The Gravity Model is found to have less explanatory power than comparable studies using dockless bike sharing data. Substituting population by public transport count, sights, cafés or hotels increased the performance of the Gravity Model. Out of all e-scooter trips in the data set about half started or ended at a public transport location. These movements are higher where there is high frequency in public transport. Therefore, it could be found that public transport is a determinant for e-scooter Micromobility. These findings suggest that e-scooters are complimentary to public transport as they are used for the first and last mile of a traveller's journey inside the city of Berlin. The findings show that this pattern holds up for origin and destination descriptions of journeys as well. It is concluded that public transport describes popular locations that create traffic.

Features of the built environment describing the popularity and population of given areas have a modest impact improving the estimates from public transport. Contrary to existing literature job opportunities as measured by office buildings marked in OSM and bicycle infrastructure were found to have no correlation with e-scooter usage and routes taken.

In general terms, even with the rapid increase of available data, Mobility Models today cannot bridge the gap between low explanatory power but broad application possibilities and high correlation coefficients on models that only have narrow application areas where they are trained on, which is as well a result of this thesis. In future work the generalisation capabilities must be further investigated, especially in other spatial areas and cities with different population distributions. With more data there is the potential to add additional dimensions like temporal distributions and weather conditions.