Modeling Bike Sharing System using Built Environment Factors

Tien Dung Tran, Nicolas Ovtracht, Bruno Faivre D’arcier

To cite this version:

HAL Id: halshs-01474166
https://halshs.archives-ouvertes.fr/halshs-01474166
Submitted on 28 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Modeling bike sharing system using built environment factors

Tien Dung Tran*, Nicolas Ovtrachta, Bruno Faivre d’Arcier

*Laboratoire d’Economie des Transports, 14 avenue Berthelot, Lyon 69007, France

Abstract

This paper aims to present a modeling of bike sharing demand at station level in the city of Lyon. Robust linear regression models were used in order to predict the flows of each station. The data used in this project consists of over 6 million bike sharing trips recorded in 2011. The built environment variables used in the model are determined in a buffer zone of 300 meters around each bike sharing station. In order to estimate the bike sharing flow, we use the method of linear regression during the peak periods of a weekday. The results show that bike sharing is principally used for commuting purposes by long term subscribers while short term subscriber’s trips purposes are more varied. The combination between bike sharing and train seems to be an important inter-modality. An interesting finding is that student is an important user of bike sharing. We found that there were different types of bikesharing usage which are influenced by socio-economic factors depending on the period within the day and type of subscribers. The present findings could be useful for others cities which want to adopt a bikesharing system and also for a better planning and operation of existing systems. Further, the solutions to encourage the use of bikesharing will be various depending on type of subscribers. The approach in this paper can be useful for estimating car-sharing demand.

1. Introduction

In recent years, bike sharing has become more and more present in the world. Bike sharing can provide an alternative to traditional modes of transport or a complementary service for solving the “last mile problem” of getting from a public transportation stop to the final destination [1]. Furthermore, bike share systems may contribute to reduce pollution and automobile usage. In terms of benefits to users, the travel times of commuting trips can be potentially reduced by 10% [10]. Moreover, a bicycle-sharing system frees individuals from the need to secure their bicycles; bicycle theft is a common problem in urban regions [19]. In terms of health, the bike sharing has also positive impacts on male and older users [21]. Another advantage associated with this system is that the decision to make a trip by bicycle can be made in a short time frame [6]. As of September 2014, more than 700 cities in 57 countries host advanced bike-sharing programs, with a combined fleet of close to 800,000 bicycles [2]. Vélo’v was one of the first major public bike sharing systems in the world. It was installed in May 2005 with 2,000 bikes and 173 stations in Lyon, France. In 2014, the Vélo’v system has 343 stations with more than 4,000 bicycles and about 53,000 long-term subscribers. Vélo’v stations are mainly located in Lyon and Villeurbanne.

To ensure the success of bike sharing schemes, demand modeling plays an important role. Vélo’v in Lyon city is mature and successful bike sharing system that offers a unique opportunity for understanding the factors influencing its flows and usage.

In this paper, we use bike sharing trips data from minute-by-minute readings of bicycle flows at all 341 stations Vélo’v in 2011 given by JC Decaux - operator of bike sharing system of Lyon - to analyze the determinants of bicycle-sharing demand. The Vélo’v trips data is combined with built
environment attributes around station allowing us to examine the influence of these factors on bicycle sharing demand.

The main objective of the current paper is to quantify the influence of built environment factors on arrival and departure flows at bike sharing station level using a statistical linear regression method.

The rest of the paper is organized as follows. Section 2 provides a literature review of earlier researches and positions our research. Section 3 explains the data used in the modeling and the socio-economic variables around bike sharing stations. The statistical model and the results are discussed in section 4. Section 5 presents the visual representation of Vélo’s flows. Finally, section 6 concludes the paper with recommendations for future researches.

2. Literature review

There have been four generations of bike sharing since the introduction of the first bike sharing system in the 1960s in the Netherlands [4] [17]. Bike sharing has become more popular since the introduction of the 3rd generation. The third generation of bike sharing can be described by the automatic transaction kiosk at each station and identified bike sharing users. These systems have become relatively successful around the world. There are some bike sharing systems of fourth generation installed in Copenhagen and Madrid with improving docking stations, bike redistribution, integration with other transport modes [4] [17] and electrical bikes.

Vélo’s belongs to the 3rd generation of bike sharing systems. The Vélo’s system aggregated more than 6.2 million trips in the 2011 and more than 50,000 long term subscribers [13]. Vélo’s bike sharing system of Lyon city is installed in the city of Lyon and Villeurbaine which cover an area of 60 kilometer square. In 2014, there were 8.3 million bike sharing trips recorded with more than 59,000 long term subscribers [15].

In recent years, many researches have used traditional surveys in order to determine the factors that may promote the adoptions of bike sharing by urban populations [12] [13]. The automated data collected from docking stations constitutes a precious source of information to better understand the usage of bike sharing in the city.

A number of researches have determined factors affecting bike sharing usage and tried to predict bike sharing flow using different urban factors such as: population, job, bicycle lanes, proximity to public transport, bike sharing station density, altitude, retail shops, etc. [6] [16] [20]. These studies were conducted using daily, monthly or yearly aggregated data which can hide the variety of daily bike sharing usage [16] [20]. Hampshire studied the built environment on bike sharing usage using aggregated hourly arrival and departure rates at the sub-city district level in Barcelona and Seville, Spain [9]. They found that bike sharing station density, capacity of stations and number of points of interest are important factors to explain arrival and departure rates of bike sharing. However, in their study, the bike sharing flows studied are aggregated at the sub-city district level which was less pertinent than using bike sharing flows at station level.

There have been several studies conducted using data from the Vélo’s system. These studies use actual bike sharing flow data obtained from stations to determine the typology of bike sharing users or to analyze the characteristics of bike sharing usage. They contribute to the literature by studying user behavior in response to bike sharing system and examining the characteristics of this system. The average speed of bike sharing is 14 km per hour [11] and the average duration of bike sharing trip is about 15 minutes.

The current paper contributes to literature by determining the effect of type of subscribers and built environment attributes on bicycle arrival and departure flows at the station level using hourly bike sharing data. The estimated models will allow us to predict not only the demand of bike sharing (arrivals and departure flows) but also to better understand the influence of built environment to the bike sharing system. The results can be helpful for decision-makers to better manage bike sharing system and for cities who want to adopt a new bike sharing system.

3. Methodology and Data

For this study, the bike sharing trips are obtained from JC Decaux – operator of Lyon bike sharing system, for all stations during the year of 2011. Each trip gives us information about the departure and arrival station, the date and hour of check in and check out and the type of subscribers.

In terms of subscribers, we are going to analyze two types of bike sharing users: long-term subscribers who have an annual bike sharing subscription and short-term subscribers who have a one-day bike sharing subscription.

In order to calculate the flows of bike sharing, we aggregated the bike sharing trips per hour. All non-valid trips were eliminated. A non-valid trip is a trip less than 3 minutes or more than 3 hours. The data aggregated were calculated only for working days (from Monday to Friday and not during vacations). We eliminated also the bike sharing trips made during July and August because they are the months of vacations in France. Finally, 173 working days were counted for calculating bike sharing flows. The bike sharing flows are then divided by 173 and multiplied by 100 before using for the calculations in the models.

For estimating the bike sharing flows, we chose 2 peak periods during a working day: from 7 am to 9 am and from 5 pm to 8 pm.

3.1. The explicative variables

The hypothesis we use in this study is that the bike sharing usage of each station depends on the built environment around the station. In order to build the models, the independent variables have to be determined.

The data were calculated by the platform MOSART (Modeling and Simulation of Accessibility of Networks and Territories) [3]. We tested the different buffer zone sizes: 200m, 300m and 400m around bike sharing station. Finally, we decided to keep 300m buffer zone because the built environment variables are the most significant in the models.
and a 300 meter buffer zone is an appropriate walking distance between Vélo's stations [5].

The explicative variables used in our analysis can be categorized in five groups: public transport variable, socio-economic variable, topographic variable, bike sharing network variable and leisure variable. All the explicative variables were calculated in a buffer zone of 300 meters around each bike sharing station except the variable of bike sharing network density which was calculated in a buffer zone of 3,500 meters around each bike sharing station which corresponds to a 15 minute biking distance.

In terms of public transit variables, the number of metro, tramway and railway stations near a Vélo's station were generated to examine the influence of public transit on bike sharing flows. The variables of public transit were normalized by the number of passengers of each station per day for railway station and per year for metro and tramway station.

The socio-economic variables included four factors: (1) population, (2) number of jobs, (3) number of students in campus and (4) number of student residences near a bike sharing station. In our calculation about bike sharing users, the median age of bike sharing users in Lyon is 30 years old. It means that half of bike sharing users in Lyon were less than 30 years old in 2011. This element explains the choice of the two student variables for the models. The altitude of each station was calculated to examine the influence of topographic variable on bike sharing usage.

The length of bicycle facilities in the buffer zone was also calculated to capture the impact of placing Vélo's stations near bicycle facilities on the usage of the bike sharing system. The number of bike sharing stations in a 3,500 meter buffer zone around a Vélo's station and the capacity of each Vélo's station were computed to capture the effect of bike sharing network.

Leisure variables are also considered in our analysis. We also considered three types of points of interest near each station: (1) number of restaurants, (2) number of cinema, and (3) the presence of embankment road of Rhone River - the main sportive and leisure zone near a bike sharing station.

<table>
<thead>
<tr>
<th>Continuous variable</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>4</td>
<td>10977</td>
<td>4707.17</td>
<td>2481.25</td>
</tr>
<tr>
<td>Job</td>
<td>148</td>
<td>11828</td>
<td>2332.09</td>
<td>2114.43</td>
</tr>
<tr>
<td>Students in campus</td>
<td>0</td>
<td>25788</td>
<td>799.59</td>
<td>2892.43</td>
</tr>
<tr>
<td>Student residence</td>
<td>0</td>
<td>10</td>
<td>1.326</td>
<td>1.98</td>
</tr>
<tr>
<td>Railway station</td>
<td>0</td>
<td>20</td>
<td>0.26</td>
<td>2.02</td>
</tr>
<tr>
<td>Metro station</td>
<td>0</td>
<td>12</td>
<td>1.51</td>
<td>2.71</td>
</tr>
<tr>
<td>Tramway station</td>
<td>0</td>
<td>27</td>
<td>1.69</td>
<td>4.33</td>
</tr>
<tr>
<td>Altitude</td>
<td>164</td>
<td>289</td>
<td>180.84</td>
<td>28.04</td>
</tr>
<tr>
<td>Bicycle infrastructure</td>
<td>0</td>
<td>2835</td>
<td>1024.95</td>
<td>650.50</td>
</tr>
<tr>
<td>Station capacity</td>
<td>10</td>
<td>40</td>
<td>19.37</td>
<td>5.89</td>
</tr>
<tr>
<td>Network density</td>
<td>45</td>
<td>277</td>
<td>238.75</td>
<td>57.94</td>
</tr>
<tr>
<td>Cinema</td>
<td>0</td>
<td>4</td>
<td>0.25</td>
<td>0.68</td>
</tr>
<tr>
<td>Restaurant</td>
<td>0</td>
<td>28</td>
<td>3.06</td>
<td>5.34</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Categorical variable</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embankment road</td>
<td>8%</td>
</tr>
</tbody>
</table>

### Table 1. Descriptive summary of explicative variables

#### 4. Analysis and Discussion

#### 4.1. Robust linear regression models

In this study, we use robust linear regression model to estimate dependent variables such as arrival and departure flows. The using of robust linear regression method can help to limit the influence of the outliers and to reduce heteroscedastic errors. The arrival and departure flows at an hourly level for each station were used in the regression model.

Let \( i = 1, 2, \ldots, 341 \) be an index to represent each station. The dependent variable (arrival or departure flow) is modeled using a robust linear regression equation which has the following structure:

\[
Y_i = \beta X_i + \varepsilon
\]

where \( Y_i \) is the arrival or departure flow at the station \( i \) as dependent variable, \( X_i \) is a vector of explicative variables determined around bike sharing station \( i \). The model coefficients, \( \beta \), are what we have to estimate. The random error term, \( \varepsilon \), is assumed to have a normal distribution across the dataset.

#### 4.2. Results

In this section, the results of robust linear regression model estimation are discussed in order to understand the different effects of built environment and type of subscribers on the bike sharing usage in the city of Lyon. In order to have the final results, we considered many specifications. The statistically significant results for arrival and departure flows are presented in Table 2 and Table 3.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Long term users</th>
<th>Short term users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.8641</td>
<td>-0.0105</td>
</tr>
<tr>
<td>Altitude</td>
<td>-2.8990</td>
<td>-4.1288</td>
</tr>
<tr>
<td>Capacity</td>
<td>22.7220</td>
<td>6.7138</td>
</tr>
<tr>
<td>Network</td>
<td>1.7982</td>
<td>4.2554</td>
</tr>
<tr>
<td>Jobs</td>
<td>0.1316</td>
<td>11.1600</td>
</tr>
<tr>
<td>Student</td>
<td>0.0312</td>
<td>4.6965</td>
</tr>
<tr>
<td>Railway</td>
<td>77.0490</td>
<td>7.8939</td>
</tr>
<tr>
<td>R2</td>
<td>0.6900</td>
<td>0.635</td>
</tr>
</tbody>
</table>

#### Table 2. Model estimation results for morning peak period (7 am to 9 am)

#### 4.2.1. Public transport variables

In terms of public transport variables, we observe that railway station is the only variable that is significant in all the models of regression. It means that the combination between train and bike sharing seems to be the most important inter-modality of bike sharing. The users who combine bike sharing...
with train may be those who live in Lyon and work far from the city or inversely who live outside of Lyon city and work in the city.

The variables of metro and tramway stations are not significant in all the models. An explanation for the non-significant inter-modality between bike sharing and metro/tramway may be that the city of Lyon and Villeurbanne is about 60 kilometers square which is accessible from the center in about 20 minutes by bike that bike sharing user who lives in inner city does not need to combine with other modes of transport.

### Table 3. Model estimation results for afternoon peak period (5 pm to 8 pm)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Long term users</th>
<th>Short term users</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coefficients</td>
<td>t-stat</td>
</tr>
<tr>
<td>Intercepts</td>
<td>367.1500</td>
<td>1.4273</td>
</tr>
<tr>
<td>Altitude</td>
<td>-7.4967</td>
<td>-7.3199</td>
</tr>
<tr>
<td>Capacity</td>
<td>22.8160</td>
<td>4.5382</td>
</tr>
<tr>
<td>Network density</td>
<td>6.4859</td>
<td>10.0180</td>
</tr>
<tr>
<td>Population</td>
<td>0.0629</td>
<td>4.0785</td>
</tr>
<tr>
<td>Railway station</td>
<td>84.4820</td>
<td>6.0085</td>
</tr>
<tr>
<td>Student residence</td>
<td>58.2370</td>
<td>4.0557</td>
</tr>
<tr>
<td>Cinema</td>
<td>-</td>
<td>20.0550</td>
</tr>
<tr>
<td>Restaurant</td>
<td>-</td>
<td>5.1413</td>
</tr>
<tr>
<td>Embankment road</td>
<td>-</td>
<td>60.7550</td>
</tr>
<tr>
<td>R²</td>
<td>0.654</td>
<td>0.584</td>
</tr>
</tbody>
</table>

### 4.2.2. Socioeconomic variables

In terms of socioeconomic variables, the results shows that population and number of jobs have an important influence on the bike sharing usage of long term subscribers. The variables are significant in all the models of inbound and outbound flows. The influence of socioeconomic variables on bike sharing usage of short term subscribers is less important: the variables are only significant in the morning models.

If we have a look at type of subscribers, we observe a difference between the bike sharing usage of long term subscribers and short term subscribers. Long term subscribers’ usage tends to be more symmetric than short term subscribers’ usage between the morning and the afternoon. The explicative variables of inbound flow in the morning are also the explicative variables of outbound flow in the afternoon and the explicative variables of outbound flow in the morning are also the explicative variables of inbound flow in the afternoon. It means that the bike sharing trips made by long term subscribers are principally for commuting purposes.

The results show also that student is an important bike sharing user. We have used 2 variables: the number of students on the campus and the number of student residences. The presence of the variables in all the models shows us that student is an important bike sharing user. The coefficient associated with the number of students in university campus on a Vélo’s station’s arrival flow has, interestingly, the opposite sign in the morning and afternoon peak periods. The number of student residence has also a positive impact of the departure flows in the morning and the arrival flows in the afternoon. Bike sharing seems to be a mode of transportation well adopted by student thanks to the cheap price of subscription.

### 4.2.3. Topographic and bike sharing network variables

The variables of bike sharing network such as: bike sharing network density and capacity of station play important role in the generation of bike sharing flows. The variables are positively significant in all the models. It means that the increase of number of stations and the increase of station capacity have positive impact on the bike sharing flows. The altitude plays an obstacle role to bike sharing usage: this variable is negatively significant in all the models.

In terms of the variable on the bicycle infrastructure, we observe that this variable was not significant in all the models, it means that the bicycle infrastructure is not so important to bike sharing users during weekdays. A plausible explanation may be that the speed of car inside the city of Lyon is not so important because Lyon bike sharing users are familiar with biking on the street.

### 4.2.4. Leisure variables

In the other hand, we observe that bike sharing usage of long term subscribers seem not to be influenced by the leisure variables. The number of restaurants, the number of cinemas and the embankment road are not significant in any models of long term bike sharing usage.

On the other hand, the bike sharing usage of short term subscribers can be described by two words: occasional and leisure. In the morning, we can see that the bike sharing flows can be explained not only by the variables concerning the characteristics of bike sharing networks, the topography but also by the leisure variables. We observe that in the afternoon, the bike sharing usage of short term subscribers are principally explained by leisure variables such as: restaurant, cinema and embankment road along Rhone River (a sportive and leisure zone of Lyon along the bank of Rhone River).

The difference between bike sharing usage of long term subscribers and short term subscribers suggest that long term subscribers use bike sharing for commuting trips while short term users utilize bike sharing for occasional and leisure trips.

### 4.3. Limits of the study

In this study, the explicative variables used are collected in 2013 and 2014 while the dependent variables (bike sharing flows) were calculated in 2011. The meteorological variables cannot be able to taken into account in our traditional models in order to estimate their influence.
5. Geovisualization

In order to better understand the spatial and temporal variation of bicycle usage in the Vélo’v system, we represent the bicycle arrival and departure flow of every station visually using a geographic information system. For this purpose, the average flows of every station in every weekday in 2011 were considered. We mainly focus on the morning peak period (from 7 am to 9 am) and afternoon peak period (from 5 pm to 8 pm) in our visualization. The bike sharing flows of long term subscribers and short term subscribers are presented distinctly in order to understand the difference of bike sharing usage between them.

Firstly, we can see (Fig 1.) that in the morning bike sharing flows of long term users are very concentrated to the biggest railway station of Lyon and the university campus La Doua. An explanation for this trend is that bike sharing is combined with train to go to work both for those who live in Lyon city and work outside of the city and for those who do not live work in Lyon city but have a job in the city of Lyon.

In the afternoon (Fig 2.), the bike sharing stations in the city center are the stations the most frequented. A plausible explanation for the trend is that employees and students may find it easier to come home by bike than to go to work because of temporal constraint in the morning. Furthermore, people might also use the Vélo’v in the afternoon after work for different purposes such as going to restaurant, going to cinema or going to recreational areas that are mainly located in the city center.

The bike sharing flows of short term subscribers in the afternoon (Fig 3.) show a clear difference with the flows of long term subscribers. The main destinations of short term bike sharing users are the City Hall, Bellecour square, the Park “Tête d’Or”, the railway station La Part Dieu and the stations along the embankment road of Rhône River.

6. Conclusion

This study examined the factors influencing the usage flows of a bike sharing system in Lyon, France. It contributes to the literature by analyzing the effect of type of users and built environment attributes on arrival and departure flows of bike sharing at the station level, using data obtained from the Vélo’v system. The network density of bike sharing and the
station capacity are plausibly correlated to bicycle usage for every station. The altitude has a negative influence on the arrival and departure flow of bike sharing. The combination of bike sharing with train is the only significant inter-modality in the models.

Population and number of jobs are two main explicative factors for the bike sharing usage of long term users. Population positively affects outbound flows of bike sharing in the morning and inbound flows in the afternoon, while number of jobs positively affects inbound flows of bike sharing in the morning and outbound flows in the afternoon. Short term user’s bike sharing usage is not only explained, in the morning, population and number of jobs but also, in the afternoon, by leisure variables. The results show that during working day, long term users utilize bike sharing principally for commuting trips while short term users use bike sharing for occasional trips.

In the point of view of PSS (Product-Service Systems), this study can be useful for bike sharing operators, bike sharing users and for the environment. First, the results can contribute to the improvement of the quality of bike sharing system. The models of bike sharing demand allow us to estimate the inbound flows and outbound flows during the morning and afternoon peak periods. Understandings the bike sharing usage during these periods are helpful for minimizing the saturation risk of bike sharing stations and for reducing the cost of bike sharing redistribution [8].

Secondly, the more the availability at station level is improved, the more bike sharing demand can be satisfied. The bike sharing users have a system of transportation available 24h/24 which they can use when needed without having to buy a bicycle or being concern about the bike parking and bike maintenance. Further, bike sharing usage can contribute to time saving for the users because bike is the fastest mode of transportation in the French cities for a distance less than 5 km [14]. Thirdly, by promoting bike sharing usage, we can reduce the air pollution caused by the car.

The models presented in this study were estimated for working days. In future works, we are going to calibrate a model explaining the bike sharing demand during weekends that is different from the bike sharing usage on working days. The results of this study can be useful to predict the bike sharing flows at station level in order to improve the quality and the availability of service and for determining the position of new bike sharing station and sizing bike sharing station. The approach used in this study can be helpful for estimating car-sharing demand.

Acknowledgements

The research leading to these results has received data from the project ANR VéllInnov, JC Decaux and MOSART platform. We would like also to thank the Rhône-Alpes region for funding the PhD for one of the authors.

References