Providing quantified evidence to policy makers for promoting bike-sharing in heavily air-polluted cities: a mode choice model and policy simulation for Taiyuan-China

Weibo Li, Corresponding Author
University College London
UCL Energy Institute, Central House, 14 Upper Woburn Place, WC1H 0NN, London, UK
Tel: +44 (0) 75 871 96096; Email: weibo.li.10@ucl.ac.uk

Maria Kamargianni
University College London
UCL Energy Institute, Central House, 14 Upper Woburn Place, WC1H 0NN, London, UK
Tel: +44 (0) 20 310 85942 / 55942; Email: m.kamargianni@ucl.ac.uk

ABSTRACT
Developing countries are facing increasing challenges to make urban mobility sustainable and to tackle the continuously growing air pollution and congestion caused by the rapid increase in car ownership. As part of a broad strategy to achieve sustainable urban mobility, bike-sharing services could contribute to car usage decrease, especially for short-distance trips. However, most of the developing countries have limited quantified evidence regarding the factors affecting bike-sharing choice and this hinders policy makers from effectively promoting bike-sharing usage. The case study city is Taiyuan, which operates one of the most in demand bike-sharing schemes in China. This research investigates the factors affecting mode choice behavior with a focus on bike-sharing, and explores the effectiveness of different policy options aiming at increasing bike-sharing ridership. Nested logit and mixed nested logit models are developed using both stated preference and revealed preference data. Policy effectiveness is studied by examining modal split changes. The results reveal the significant negative impact of air pollution on bike-sharing choice. Nevertheless, improving air quality is found to be less effective in promoting bike-sharing ridership than improving bike-sharing service itself (e.g. through access time saving, travel cost saving); although it is more effective in suppressing private car usage.

Keywords: Bike-sharing; Mode choice; Air pollution; Mixed nested logit; Demand elasticity; SP/RP data

1. INTRODUCTION

Developing countries are facing increasing challenges to tackle the continuously growing air pollution and congestion caused by the rapid increase in car ownership. As part of a broad strategy to achieve sustainable urban mobility, bike-sharing services can help to reduce car usage, especially for short-distance trips. Research outcomes have shown that the benefits of bike-sharing are numerous; avoiding parking and maintenance troubles with private bikes, offering more convenient connection to public transport, reducing travel time and costs especially in city centers, improving body health, and opening up opportunities for more social and leisure experiences (DeMaio and Gifford, 2004; Jäppinen et al., 2013; Ricci, 2015).

Following the success in Europe and North America (DeMaio, 2009; Shaheen et al., 2010), bike-sharing schemes have been introduced in many cities in developing countries as well. However, although there are many mode choice studies for developed countries there is a lack of knowledge in the factors affecting bike-sharing choice in developing countries. This gap has significantly hindered policy making to effectively promote bike-sharing usage. More importantly, findings from developed countries may not be directly applied to developing countries as culture and local/geographical characteristics are significantly different (Maurer, 2012; Faghih-Imani et al., 2015; Kamargianni, 2015).

This research addresses the aforementioned gap by investigating the factors affecting mode choice behavior in heaviliy air-polluted cities in developing countries, while focusing on bike-sharing. It also explores the effectiveness of different policy options aiming at increasing bike-sharing ridership. Particular focus is placed on the impact of air pollution on mode choice, since such an effect has rarely been captured when the scope was largely limited to developed countries. Air pollution may play an important role in affecting mode choice behavior in developing countries, which usually have more severe air pollution levels over prolonged periods of time. Specifically, this study tests if an increase in air pollution level would depress the willingness to cycle and to what extent an improvement in air quality would increase bike-sharing demand.

Mode choice models are developed including nested logit and mixed nested logit (Hess et al., 2004; Ortúzar and Willumsen, 2011) to address inter-alternative correlation and panel effect. For models development, stated preference (SP) and revealed preference (RP) mode choice data is combined to obtain results with less behavioral bias (Hensher and Bradley, 1993; Ben-Akiva et al., 1994). Our case study city is Taiyuan (China), which currently operates one of the most in demand bike-sharing schemes in China. The models are compared across each other and the one with the best performance is selected to study policy impacts on modal split changes in the SP environment[1]. This research focuses on short-distance trips (within 2km), since it is the most prevalent bike-sharing traveling range in China (Gu Dong, 2016).

The paper is structured as follows. Section 2 reviews the current literature on factors affecting cycling and bike-sharing choices to draw insights and identify knowledge gaps. Section 3 presents the case study information and data sources. Section 4 explains the modeling framework and describes the model specifications in detail. Section 5 discusses on model estimation results, followed by a policy impact analysis in section 6. Section 7 concludes research findings and policy implications.

2. LITERATURE REVIEW

[1] This study does not forecast market demand in the real world.

Previous studies have identified several factors affecting bike-sharing choice as well as cycling choice. These factors can be classified into three categories: 1. Natural and built environmental conditions, 2. Trip and mode attributes, and 3. Socio-economic characteristics.

Natural and Built Environmental Conditions

Natural environmental conditions, such as weather, temperature, air-pollution, seem to heavily affect cycling choice. Some researchers incorporated different weather conditions (e.g. sunny, rain or snow) in their mode choice models (Daito and Chen, 2013; Kamargianni, 2015), while others also accounted for temperature impact (Parkin et al., 2008; Saneinejad et al., 2012; Motoaki and Daziano, 2015; De Chardon et al., 2017). In general, these studies came to similar conclusions; namely that adverse weather conditions and colder temperature would significantly discourage travelers from cycling. Many studies also analyzed the impact of topography. In particular, steeper roads would significantly discourage the choice of bicycle (Waldman, 1977; Rietveld and Daniel, 2004; Parkin et al., 2008; Mateo-Babiano et al., 2016; De Chardon et al., 2017), although Motoaki and Daziano (2015) argued that the impact of hills on the cycling route choice heavily depended on the fitness of cyclist. Additionally, the effect of air pollution has been studied, but, to our knowledge, among the great number of studies for developed countries, only Zahran et al. (2008) covered this effect via a cross-sectional analysis at the US county level and found pollution could decrease the number of cycling commuters on the road.

In relation to built environmental and land use impacts, cycling-related infrastructures have attracted significant attention in the existing literature. Many studies have focused upon the importance of increasing the number of cycle lanes and bike-sharing stations in promoting the use of cycling or bike-sharing, in terms of reduced travel time, increased safety and convenience (Akar and Clifton, 2009; Larsen and El-Geneidy, 2011; Hankey et al., 2012; Daito and Chen, 2013; Kamargianni and Polydoropoulou, 2013; Deenihan and Caulfield, 2015; Kamargianni, 2015; Maness et al., 2015; Wang et al., 2015; Mateo-Babiano et al., 2016; De Chardon et al., 2017). However, there were also papers that found an insignificant relationship between the number of cycling facilities and cycling choice (Rodríguez and Joo, 2004; Moudon et al., 2005; Xing et al., 2010). Some other relevant factors such as population density in community, the existence of university campuses and number of parks etc. were also studied (DeMaio and Gifford, 2004; Rodríguez and Joo, 2004; Barnes and Krizek, 2005; Moudon et al., 2005; Parkin et al., 2008; Maurer, 2012; Whalen et al., 2013; Kamargianni and Polydoropoulou, 2014).

Trip and Mode Attributes

Trip characteristics are also important factors that determine mode choices. Cycling has been found to be more associated with recreational-purpose trips (Moudon et al., 2005; Xing et al., 2010; Mateo-Babiano et al., 2016). Faghih-Imani et al. (2015) found that cycling trips occurred more during noon and evening periods for meal purposes, while most of the morning cycle trips were for commuting. Moreover, since bicycles move more slowly than motorized vehicles, there was overwhelming evidence showing the negative relationship between cycling choice and trip distance (Parkin et al., 2008; Zahran et al., 2008; Akar et al., 2013; Faghih-Imani et al., 2015; Wang et al., 2015). Xing et al. (2010) even argued that perceived trip distance had the largest influence compared to other variables. Meanwhile, some trip characteristics, such as travel time, travel cost, and comfort level, may be seen as factors affecting transport mode choice. Many researchers studied the impacts of the attributes associated with bike-sharing and versus those associated with alternative modes (such as car speed, parking availability, public transport cost, and service frequency) to evaluate mode choice decisions (Lin and Yang, 2011; Kamargianni and

Polydoropoulou, 2013; Whalen et al., 2013; Faghih-Imani et al., 2015; Fishman et al., 2015; Ahillen et al., 2016; De Chardon et al., 2017). Due to different sample characteristics and different measurements of mode attributes, the impact significance of each attribute more or less differs across studies.

Socio-economic Characteristics

Socio-economic characteristics have been widely studied, with age and gender emerging as among the most influential factors. Younger generations and males are usually keener to cycle (Shafizadeh and Niemeier, 1997; Rodríguez and Joo, 2004; Moudon et al., 2005; Parkin et al., 2008; Baker, 2009; Akar et al., 2013; Fishman et al., 2015; Ricci, 2015; Wang et al., 2015), whilst occupation and economic status may also play important roles in determining cycling choice. Xing et al. (2010) showed that travelers with lower income cycled more because those with higher income valued their time more highly and chose faster modes. Faghih-Imani et al. (2015) reached similar conclusions, arguing that the unemployed usually preferred cycling. However, some studies found that higher cycling rate could be associated with higher economic status (Parkin et al., 2008; Zahran et al., 2008; Fishman et al., 2015; Kamargianni, 2015) as a result of pursuing healthier lifestyles. In contrast, Baltes (1996) found that economic status and unemployment are both insignificant in determining cycling choice. Additionally, cycling was found to be a popular mobility choice among students (Baltes, 1996; Whalen et al., 2013; Wang et al., 2015). Vehicle ownership seems to be a more direct determinant of mode choice. In general, vehicle ownership could decrease the incentive or the need to cycle, either for educational (Rodríguez and Joo, 2004) or work purposes (Parkin et al., 2008). However, such an inverse relationship might be attributed to collinearity with other factors, that is those who do not own any vehicles and have to cycle could do so because of their disadvantaged income status that makes the purchase of a vehicle unaffordable, or travel distance may be too short to make it worthwhile (Baltes, 1996). Other socio-economic factors related to cycling choice include health status (Moudon et al., 2005) and educational level (Xing et al., 2010).

Another popular approach to study socio-economic characteristics (instead of assuming their direct effects on mode choice utilities) is exploring systematic taste heterogeneity (Amador et al., 2005; Cherchi and Ortúzar, 2011). More insightful results could be gained by also taking into account this effect. In the case of cycling, for instance, it reveals how different socio-economic groups would react to the impacts of natural and built environmental conditions as well as trip and mode attributes, e.g. female travelers were still reluctant to cycle even if in sunny days which in general could increase the attractiveness of cycling (Kamargianni, 2015).

Although many studies have been conducted on cycling and bike-sharing choices, gaps still exist. Firstly, there is a lack of mode choice studies in developing countries, particularly with respect to bike-sharing. The results in developed countries may have limited implications for developing countries since different local characteristics could lead to different results and conclusions. The existing literature has demonstrated such differentiations even when carried out within developed countries. Some studies also directly showed the context-specific nature of mode choice study through simultaneously studying multiple cases (Barnes and Krizek, 2005; Tang et al., 2011; Maurer, 2012; Faghih-Imani et al., 2015; Kamargianni, 2015). Secondly, there is a lack of literature focusing upon the impact of air pollution, which is generally not a significant concern in developed countries. However, it is essential to take into account such effects in the developing world where air pollution is a much more severe challenge. A recent study should be acknowledged (Campbell et al., 2016), in which the authors took into account air pollution's impact when using SP survey data from 623 participants and a multinomial logit model to study bike-sharing choice in Beijing. In our research, despite having a different scope and methodology as well as a larger sample, we extend further the findings on air pollution by revealing its effect on modal splits via a policy impact analysis.

3. CASE STUDY AND DATA

The case study city is Taiyuan, the capital city of Shanxi province in northern China. Taiyuan has more than 3 million citizens and operates one of the most in demand bike-sharing services in the country (Song, 2015). The service can be easily accessed via public transport card and cycle lanes are available on most streets. The city has sharp air pollution level variations making the impact on mode choice behavior worth exploring.

The data used in this paper originate from a paper-based questionnaire survey that collected both revealed and stated preference data. In terms of, RP data, the survey collected information about the socio-economic characteristics of the participants, while they were also asked to fill in their trip diary for one day. Due to resource constraints and the local cultural barriers, the use of GPS or smartphone based travel survey tools that could collect more advanced travel data was not feasible. As such, only essential travel information were provided in the trip diary (e.g. starting/end time of the trip, travel time, travel cost, mode used). In terms of SP experiments, the participants were presented with hypothetical situations for short-distance trips (less than 2km)[2], where they were asked to chose a transport mode.

Table 1 shows the SP experimental design for short-distance trips. In our design, there are six alternatives: 1. car, 2. electric bike, 3. bus, 4. car-sharing[3], 5. bike-sharing and 6. walk. Each of the alternatives possesses a number of mode specific attributes, joint with trip purpose, weather condition and air pollution level. The selection of these attributes were based on literature review, and their levels/values were derived from the pilot survey results (to produce the levels of travel times and travel costs, the averages of the perceived travel times and costs from the pilot trip diary survey were used as references).

The SP experiment followed the orthogonal main effects design (Hensher et al., 2005). Although this is not as advanced as several later proposed designs, such as D-optimal design and D-efficient design (Bliemer et al., 2009; Rose and Bliemer, 2009; Bliemer and Rose, 2010), this project adopted the traditional orthogonal design due to time, cost and availability of advanced data collection tools constraints. A summary of the different advantages and generations of these SP designs can be found in Ortúzar and Willumsen (2011). In total, 60 different scenarios were expertly generated for short-distance trips while satisfying the required degree of freedom in order to maintain orthogonality (Caussade et al., 2005; Hensher et al., 2005). The 60 scenarios were assigned to 30 blocks to further reduce the number of scenarios presented to each individual respondent. Eventually, each questionnaire contained 2 scenarios for short-distance trips and 1 out of every 30 respondents was given the same scenarios[4]. Appendix A gives an example of the two scenarios as presented to the respondents.

---

[2] There are two more types of scenarios in the SP experiment, medium-distance trips (2km to 5km) and long-distance trips (more than 5km), since the available alternatives and some attribute levels (i.e. travel time and travel cost) vary across distances. These two cases follow the same technical design as short-distance trips.

[3] Car-sharing was just about to enter Taiyuan at the time of the survey and there was imperfect knowledge regarding this concept among respondents. Thus, the concept was described in the beginning of the SP part to reduce understanding bias.

[4] Each participant responded to 2 short, 2 medium and 2 long-distance scenarios to limit the total number below 8, i.e. the threshold that most of the pilot survey participants would start to feel annoyed.

TABLE 1 SP Experimental Design, Short-distance Trip

Trip purpose: work/education, leisure, shopping.
Weather condition: sunny (-10°, -5°, 0°, 5°, 10°, 20°, 25°, 30°), snow (-10°, -5°, 0°), rain (5°, 10°, 20°, 25°, 30°).
Air pollution level: excellent, good, light pollution, medium pollution, heavy pollution, terrible pollution.

	Car	Electric bike	Bus	Car-sharing	Bike-sharing	Walk
Travel time	2, 3, 5, 7, 10 min.	5, 6, 7, 9 min.	5, 7, 10, 12, 15min.	2, 3, 5, 7, 10 min.	8, 10, 12min.	10, 15, 20, 25, 30min.
Travel cost*	￥1, 1.2, 1.4, 1.6, 1.8.		￥0.5, 1, 1.5, 2, 2.5.	￥0.8, 1, 1.5, 2, 3, 4, 5.	￥0, 0.5, 1.	
Parking space	Easy/hard to find parking					
Parking cost*	free, ￥2/h, ￥5/h, ￥8/h.					
Walking time to/from station			5min, 10min, 15min.	5min, 10min, 15min.	2, 5, 10 min.	
Bus Frequency			every 2min, 5min, 10min, 15min.			
Mobile app availability			Yes, no.	Yes, no.	Yes, no.	

* ￥1 ≈ $0.15

In collecting the data, the authors co-operated with Shanxi Transportation Research Institute, which provided 15 researchers assisting with the questionnaire distribution, questionnaire collection and incorporation of the data into electronic datasets. The questionnaire was distributed to 15,000 Taiyuan citizens during summer 2015 after a pilot survey in January 2015. Due to the population size of more than 3 million in the urban area, the concern on sample representativeness was addressed by calibrating the sample to Taiyuan census data. First, the sampled individuals were proportionally spread over the six districts in the urban area as per the population size in each district; and second, the gender distribution of sampled individuals in each district was set proportional to the population gender distribution in each district.

After preliminary data cleaning, 9,499 individuals provided valid data (see Appendix B for a comparison between the sample and the census data). Then, the SP mode choice data used for this paper, was further refined by keeping only observations that were rigorously consistent with the participants' RP trip diary information (i.e. if someone made SP choices in the short-distance scenarios but did not reveal any "within 2km" trips in the trip diary, these SP choices were excluded from the analysis). In the end, there are 4,769 individuals offering 9,028 valid observations for the short-distance trips SP experiment.

Table 2 shows the modal splits in these observations as well as a comparison to the modal splits in the RP trip diary. It is noteworthy that apart from car-sharing was not yet a mature option in Taiyuan at the time of the survey, private bike was deliberately excluded in the SP survey leading to another distinction between the two choice sets. This is due to private bike usage has dropped substantially after the city's huge success in bike-sharing and is expected to diminish further as bike-sharing continues to grow (Oortwijn, 2017; Poon, 2017). The statistics in Table 3 reveals a similar trend that bike possession rate is much lower than the other private modes in the sample.

Table 3 also presents other key descriptive statistics. Age and occupational status statistics indicate that adults with fixed jobs constitute the main group in the sample, indicating that the sample has successfully captured regular commuters whose mode choice behaviors are most considered in urban planning and policy-making. There is a high possession rate of public transport cards meaning that most of the sampled individuals can access both bus and bike-sharing services hassle-free. Almost all respondents are healthy enough to cycle, which ensures that bike-sharing is a feasible choice in a sufficient number of scenarios.

TABLE 2 Modal Splits in Short-distance Trips

SP data (9,028 obs.)		RP data (6,614 obs.)	
Bike-sharing	22%	Bike-sharing	18%
Walk	30%	Walk	31%
Electric bike	9%	Electric bike	12%
Bus	29%	Bus	26%
Car-sharing	2%	-	-
Car	8%	Car*	8%
-	-	Bike	5%

* In the RP data, it is also known that the 8% car trips consist of 6% car driver trips and 2% car passenger trips.

TABLE 3 Socio-economic Statistics of the Sample

		N=4,769
Gender	Male	51%
	Female	49%
Age	under 18	9%
	18-25	31%
	26-35	27%
	36-45	20%
	46-59	11%
	60 or above	2%
Marital status	Single	47%
	Married	53%
Educational level	High school or below	29%
	College	32%
	Undergraduate	34%
	Graduate and above	5%
Occupational status	Fixed job	68%
	Student	24%
	Retired	2%
	Self-employed or unemployed	6%
Public transport card	Percentage of possession	74%
Cycling capability	Health enough to cycle	94%
Household monthly income (after tax)*	Under ￥3000	34%
	￥3000 -￥6000	36%
	￥6000 -￥9000	16%
	￥9000 -￥15000	9%
	￥15000 -￥30000	4%
	Over ￥30000	1%
Household car	Percentage of possession	46%
Household electric bike	Percentage of possession	42%
Household bike	Percentage of possession	17%

* ￥1 ≈ $0.15

4. MODELING FRAMEWORK AND MODELS SPECIFICATION

To estimate the mode choice models we utilize the SP dataset and the combined SP and RP dataset. This approach is followed, because the scenarios are hypothetical and the choices made could be inconsistent to the behavior in reality. Thus, combing SP data with RP data as a way to reduce such bias has become a popular practice in choice modeling (Hensher and Bradley, 1993; Ben-Akiva et al., 1994; Bradley and Daly, 1997; Bhat and Sardesai, 2006; Cherchi and Ortúzar, 2011; Lavasani et al., 2017). This study takes advantage of having access to both data types and pools together SP and RP mode choice data based on distance criteria (within 2km, see Table 2).

In terms of modelling, nested logit (NL) models using SP and both SP and RP mode choice data are developed (as base models) to account for any potential correlation among the alternatives. Due to the panel structure of SP data (i.e. repeated choice observations from a single respondent), mixed nested logit (mixed NL) models are further developed to capture the correlation across choice observations. Mixed logit is a flexible model structure that can approximate any random utility model (McFadden and Train, 2000; Hensher and Greene, 2003). A mixture of multinomial logit can simultaneously address the aforementioned inter-alternative correlation and panel effect by adding error components. However, arguments have arisen supporting the use of a mixture of nested logit in order to avoid any potential confounding effects when introducing more than one type of error component (Hess et al., 2004; Ortúzar and Willumsen, 2011). Hence, we follow the mixed NL approach to develop the mode choice models for this study. The mathematical equations used to specify the model are provided below (Eq.(1) – Eq.(7)) (for more information see: Hess et al., 2004; Ortúzar and Willumsen, 2011).

The utility function for an alternative i ($i \in C_n$) chosen by an individual n ($n = 1,...,N$) at the t^{th} ($t = 1,...,T$) number of SP scenario is given by:

$$U_{int} = \sum_{k=1}^{K} \beta_k X_{intk} + \sigma_i \eta_{in} + \varepsilon_{int} \qquad (1)$$

while the measurable part of the utility is defined as:

$$V_{int} = \sum_{k=1}^{K} \beta_k X_{intk} + \sigma_i \eta_{in} \qquad (2)$$

where C_n is the choice set, U is the utility associated with a mode choice, X is the vector of explanatory variables, and the normally distributed error component η with zero mean captures the panel effect. The estimated parameters are β_k and σ. V is the measurable utility and ε is the unobserved term i.i.d. Extreme Value and independent from η.

The choice probability functions are:

Choice of a nest (upper level):

$$P_{M_s nt} = \frac{e^{\lambda_s IV_{snt}}}{\sum_{z=1}^{Z} e^{\lambda_z IV_{znt}}} \qquad (3)$$

Choice of an alternative inside a nest (lower level):

$$P_{int|M_s} = \frac{e^{V_{int}/\lambda_s}}{\sum_{j \in M_s} e^{V_{jnt}/\lambda_s}} \quad (4)$$

General choice of an alternative:

$$P_{int} = P_{M_s nt} P_{int|M_s} \quad (5)$$

where P is choice probability, M_s represents the nest s ($s = 1,...,z$), IV is the expected maximum utility for the choice of alternatives inside a nest, λ is the scale parameter measuring the different variances across nests.

The general choice probability function is integrated over η, gives (now P_{int} is fully denoted as the conditional probability $P_{nt}(i_t \mid X_{int}, \beta_k, \eta_{in}, C_n)$):

$$L_n(i \mid X_{in}, \beta_k, \sigma_i, C_n) = \int \prod_{t=1}^{T} P_{nt}(i_t \mid X_{int}, \beta_k, \eta_{in}, C_n) f(\eta_{in}) d\eta_{in} \quad (6)$$

Log-likelihood function that needs to be maximized:

$$LL(\beta, \eta) = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} InL_n(i \mid X_{in}, \beta_k, \sigma_i, C_n) \quad (7)$$

where y_{in} takes the value of 1 if an individual n chooses an alternative i and 0 otherwise.

Several models have been estimated to identify the correct explanatory variables and their appropriate forms. For each variable, we measured its impact on all mode choice utilities and identified the one which parameter value is closest to zero for normalization. Variables that displayed highly insignificant effects on mode choice utilities were dropped out to avoid type I errors[5]. These include snowy weather, car parking space availability and bus frequency etc. A linear relationship was adopted to measure the impact of temperature as it showed much higher significance than a curvilinear relationship (i.e. extreme and moderate temperature). Socio-economic factors were tested in two ways: 1.by assuming their direct effects on mode choice utilities, and 2. by interacting with other attributes (i.e. systematic taste heterogeneity). The results showed that model fitness improved significantly with the latter manner. To capture systematic taste heterogeneity, the sub-categories of the socio-economic variables were merged into two general groups (i.e. low and high) to more explicitly reveal their impacts. For inter-alternative correlation, many possibilities were tested including bike-sharing and electric bike as two wheeled vehicle, bike-sharing and walk as active mode, bike-sharing and car-sharing as newly emerged sharing economy, car and car-sharing as comfortable automobile, bus and car-sharing as shared automobile. Eventually, only bus and car-sharing were found to have significant correlation. Table 4 presents the variables included in the final models and the ways they were measured.

Regrading, the NL and mixed NL models using the combined SP and RP, the RP trip diary data was utilised to estimate the parameter values on the following variables: “Rain”, “Commute”, “Travel cost”, “Travel time” and all the socio-economic factors. “Air pollution”, “Temperature”, “Parking cost”, “Access time” and “App availability” were not captured in the RP data and such as

[5] Incorrect rejection of a true null hypothesis

we cannot estimate these paremeters. Meanwhile, the values of "Air pollution" and "Temperature" displayed little variations across the observed RP trips and were therefore considered as redundant. It is because the trip diary survey was conducted only in summer days and the case study city Taiyuan has very stable pollution and temperature levels in this season. Different scaling factors (to correct variance difference) were adopted in the model estimation[6].

Finally, three availability conditions were included in the mode choice models: 1. Car is available to households that own a car, 2. Electric bike is available to households that own an electric bike, and 3. Cycling is available to those who are able to cycle given their state of health. The availability conditions can increase model validity by helping to explain the circumstances within which someone does not choose a particular mode due to the fact that the mode is not an available option. Possession of a driving license was not considered an availability condition since the choice of car or car-sharing could be made by drivers as well as passengers; possession of public transport card was also excluded as travelers would still access bus or bike-sharing service by paying cash or borrowing others' card.

TABLE 4 Explanatory Variables and Measurements

Variable	Measurement
Air pollution	air quality index (AQI) by taking the average value of each level (25 for excellent level '0-50', 75 for good level '51-100', 125 for light pollution '101-150', 175 for medium pollution '151-200', 250 for heavy pollution '201-300', 400 for terrible pollution 'above 300')
Rain	1 if weather is rainy, 0 if otherwise
Temperature	temperature in °C
Commute	1 if trip purpose is commute (i.e. work/education), 0 if otherwise
Travel cost	in RMB
Parking cost	in RMB/hour
Travel time	in min
Access time	in min, walking time to stations/parking spots
App availability	1 if a smart phone application is available, 0 otherwise
Male	1 if gender is male, 0 if female
Lower age	1 if age is "under 18" or "18-25" or "26-35", 0 if "36-45" or "46-59" or "60 or above"
Lower income*	1 if household monthly income is "under ￥3000" or "￥3000-￥6000" or "￥6000-￥9000", 0 if "￥9000-￥15000" or "￥15000-￥30000" or "over ￥30000"
Lower education	1 if educational level is "high school or below" or "college", 0 if "undergraduate" or "graduate and above"

* ￥1 ≈ $0.15

5. MODEL ESTIMATION RESULTS

To estimate the NL and mixed NL models, PythonBiogeme (Bierlaire, 2016) was used. Table 5 shows the findings on the SP data and Table 6 shows the findings on the pooled data. We first compare across these modeling outputs and then discuss the factors affecting the choice of bike-sharing and other mode choices in general.

5.1 Models performance and comparison

The first model is a NL model based on the use of SP data. Bus and car-sharing are found to share some common unobserved attributes under the so-called nest "shared automobile". The

[6] In this study SP data is the primary data source and the RP utilities were scaled relative to it (Hensher and Bradley, 1993).

output μ value 2.24, complies with the specification requirement of nested logit as it is greater than 1, where $\mu = 1/\lambda$ [7] (Hess et al., 2004; Ortúzar and Willumsen, 2011). There is no other significant correlation being detected among the rest alternatives. Panel effect is revealed next using a mixed NL model and appears to be significant on all mode choices (car-sharing is normalized). The nesting parameter μ shrinks as expected (Hess et al., 2004) since the mixed NL model decomposes the error term further than the NL model. The model performance increases by capturing the panel effect given the significant improvements in likelihood ratio test and adjusted rho-bar squared.

When the RP data is added, the model performance increases further compared to the two models based on only SP data. Meanwhile, panel effect is estimated simultaneously in the RP data as there are also repeated observations from an individual in the RP trip diary. Nests are tested on the RP mode choices as well although they did not turn out significant as in the SP case. Overall, the mixed NL model based on combined SP and RP data shows the best performance and will therefore be used next to study the factors' impacts on mode choices.

5.2 Model estimation results

5.2.1 Model estimation results: Bike-sharing

Regarding natural environmental conditions, firstly, air pollution is found to have significant negative effect on bike-sharing choice. Due to the possible concern on health damage an increase in air pollution level would discourage travelers from using bike-sharing. Next, the impacts of weather and temperature are shown to be similar to those found in earlier studies. A rainy weather can significantly decrease the demand for bike-sharing and a warmer weather can increase the probability to use bike-sharing.

The impacts of trip and mode attributes are revealed next. When conducting commute trips (for work or education) bike-sharing is a less preferable option. In other words, as the most literature shows, bike-sharing is more likely to be used for leisure purposes. As for travel cost and travel time, bike-sharing choice is, as expected negatively correlated with the former and however positively correlated with the latter. A discussion on this finding is given in the next subsection (5.2.2). Access time to bike-sharing parking spots is negatively associated with its choice which means longer walking distance will discourage people from using the service. It is also found a negative coefficient on bike-sharing app availability. Such a result is nevertheless in line with the fact that the existing bike-sharing app in Taiyuan is not popular at all among the registered bike-sharing users as shown in the operator's latest report (Taiyuan Public Transport Holdings, 2016). The bike-sharing docking stations in Taiyuan is quite dense (there is a docking station every 500m on average) and probably this has made a smartphone app (e.g. provide real-time information on bike availability) rather redundant.

Finally, the choice of bike-sharing is not significantly associated with any key socio-economic characteristics (gender, age, household income and education level) although their effects are analyzed in the way of systematic taste heterogeneity (results not included in the final models due to high insignificance). Such a finding is in fact similar to the results of the aforementioned Beijing study (Campbell et al., 2016) in which the authors showed bike-sharing users could emerge across the social spectrum with no significant preference from any particular groups of people.

---

[7] λ was defined earlier in Eq. 3 and Eq. 4.

5.2.2 Model estimation results: Rest of the modes

Apart from bike-sharing, air pollution also has significant negative impact on walk, electric bike and bus choices. Car-sharing is the only mode that displays positive correlation between its utility and higher air pollution level (in fact car choice shows a positive relationship too, but it is normalized to base when specifying the model). The impact of adverse weather is consistent with air pollution, such that rain will discourage the choices of electric bike and walk while increasing the attractiveness of car and car-sharing. As for temperature, another mode choice besides bike-sharing that is preferred under warmer weather is walking, whereas car and car-sharing are more likely to be chosen when temperature falls.

In terms of trip purpose, walking is a significantly preferred mode for short-distance commute trips. A more interesting result is found on private car choice. In Table 5, people's stated choices imply that they do not like to use cars for commuting; however, when their actual behavior is incorporated (combined SP and RP data), private car choice turns out to be positively associated with commute trips (Table 6). Regarding, the rest of the modes (electric bike, bus and car-sharing) no significant correlation has been found between their choices and trip purposes.

An increase in travel cost will decrease the utility of all mode choices, although such an impact on bus choice and car choice is insignificant as shown by the mixed NL model in Table 5 and 6. However, for travel time, its effect is positively associated with all mode choice utilities except for walk. Hess et al. (2005) offered a comprehensive explanation for such a phenomenon and positive travel time coefficients would simply indicate the existence of conjoint activities[8] and travel-experience factors[9] (Salomon and Mokhtarian, 1998) that people perceive when making mode choice decisions. In microeconomic term, the marginal opportunity cost of travel time would be offset or even overwhelmed by the marginal benefit of travel time associated with a mode choice. As a result, the willingness to pay for travel time saving is not possible to derive in this case since the "travel time" variable captures not only the effect of travel time, but also the effect of any conjoint activities and travel-experience factors.

The willingness to pay for access time savings can be estimated using the ratio of marginal utilities of access time over travel cost. The access time variable on the choices of bike-sharing, car-sharing and bus all display negative signs meaning that longer walking journeys to the stations or parking spots can reduce the utilities associated with these choices. In the case of short-distance trip, the estimated willingness to pay values are ￥0.12, ￥0.16 and ￥1.02 per minute for bike-sharing, car-sharing and bus respectively. Future studies, especially in the context of China, are welcome to compare to the results. At last, the remaining mode attributes have the expected signs of impact: bus app availability (positive), car-sharing app availability (positive) and car parking cost (negative).

Systematic taste heterogeneity is firstly captured in the NL models with its significant impact being found on the choices of bus, car and walk (no other systematic taste heterogeneity is detected as significant apart from those presented). Recall that bus usage is negatively correlated with air pollution, the positive coefficients on the two interacted terms (air pollution and lower age group, air pollution and lower income group) suggest that younger and less wealthy people would still use bus service even if air quality becomes worse. On the contrary, the group of male travelers is found to prefer bus less than female travelers, while air pollution would further push the male group away from using the service. For the taste heterogeneity on trip purpose, in the SP only model (Table 5), the lower income group do not prefer neither car nor walk for commuting, no

[8] That is the negative marginal utility of a travel-time increase is compensated by the gains in utility resulting from simultaneously conducted activities.
[9] Such as the comfort, pleasure or the positive social perception associated with traveling by a particular mode.

matter the mode itself is actually a preferable option (walk) or a less preferable option (car) for commute journeys. In the pooled dataset (Table 6), the lower income group still dislikes car and walk for commute purpose even though car is now positively associated with commuting as we showed earlier. Nevertheless, these results have become slightly different when panel effect is incorporated; the t-statistics measuring systematic taste heterogeneity decrease in the mixed NL models and some values then become insignificant (Table 5 and 6).

TABLE 5 Model Estimation Results Using SP Data

	NL		Mixed NL	
	Coef.	**t-stat**	**Coef.**	**t-stat**
$\alpha_{bikeshare}$	0.97	1.88	2.85	3.62
α_{walk}	2.23	7.71	4.02	7.72
α_{ebike}	0.23	0.57	0.80	1.10
$\alpha_{carshare}$	- 17.80	- 4.20	- 0.03	- 0.06
α_{car}	0.98	2.21	1.07	1.28
Natural environmental conditions				
Air pollution-bikeshare	- 0.0032	- 4.66	- 0.0081	- 6.20
Air pollution-walk	- 0.0049	- 9.20	- 0.0111	- 9.48
Air pollution-ebike	- 0.0028	- 4.59	- 0.0078	- 6.40
Air pollution-bus	- 0.0041	- 4.63	- 0.0062	- 4.27
Air pollution-carshare	0.0213	3.36	0.0011	0.74*
Rain-bikeshare	- 0.51	- 2.63	- 0.64	- 2.54
Rain-walk	- 1.10	- 8.15	- 1.74	- 8.89
Rain-ebike	- 0.74	- 4.39	- 0.73	- 2.92
Rain-carshare	5.37	3.51	1.46	4.72
Rain-car	0.16	0.84*	1.06	3.40
Temperature-bikeshare	0.01	3.23	0.02	3.89
Temperature-walk	0.01	2.38	0.01	2.60
Temperature-carshare	- 0.24	- 4.45	- 0.08	- 7.09
Temperature-car	- 0.02	- 4.23	- 0.05	- 6.17
Trip and mode attributes				
Commute-bikeshare	- 0.76	- 7.22	- 1.23	- 7.61
Commute-walk	0.25	2.96	0.22	1.31*
Commute-car	- 0.23	- 1.43*	- 0.79	- 2.66
Travel cost-bikeshare	- 0.69	- 6.17	- 0.78	- 4.92
Travel cost-bus	- 0.41	- 3.45	- 0.08	- 0.62*
Travel cost-carshare	- 2.05	- 3.37	- 0.27	- 2.63
Travel cost-car	- 0.29	- 0.74*	- 0.90	- 1.22*
Parking cost-car	- 0.06	- 2.78	- 0.09	- 2.26

Travel time-bikeshare	0.27	5.03	0.38	4.37
Travel time-walk	- 0.03	- 2.60	- 0.05	- 2.17
Travel time-ebike	0.24	4.28	0.38	3.82
Travel time-bus	0.12	6.93	0.18	7.72
Travel time-carshare	0.18	1.07*	0.13	3.39
Travel time-car	0.04	0.81*	0.01	0.14*
Access time-bikeshare	- 0.17	- 8.17	- 0.24	- 7.45
Access time-bus	- 0.11	- 6.60	- 0.24	- 8.10
Access time-carshare	- 0.17	- 0.94*	- 0.08	- 1.98
App availability-bikeshare	- 0.87	- 9.58	- 1.11	- 8.10
App availability-bus	0.12	1.28*	0.70	5.44
App availability-carshare	2.14	3.30	0.24	1.40*
Systematic taste heterogeneity				
Air pollution * Male-bus	- 0.0017	- 4.94	- 0.0018	- 3.67
Air pollution * Lower age-bus	0.0024	6.29	0.0020	3.75
Air pollution * Lower income-bus	0.0013	2.31	0.0013	1.61*
Commute * Lower income-car	- 0.33	- 2.67	- 0.53	- 1.99
Commute * Lower education-walk	- 0.18	- 3.18	- 0.18	- 1.31*
Inter-alternative correlation & Panel effect				
$\mu_{sharedmotor}$	2.24	7.30#	1.84	6.75#
$\sigma_{bikeshare}$	-	-	0.84	4.60#
σ_{walk}	-	-	3.28	23.23#
σ_{ebike}	-	-	2.58	13.25#
σ_{bus}	-	-	1.78	15.39#
σ_{car}	-	-	3.27	12.66#
Number of observations	9028		9028	
Initial log-likelihood	- 14122.8		- 14122.8	
Final log-likelihood	- 12188.0		- 11079.7	
Likelihood ratio test	3869.5		6086.1	
Adjusted rho-bar squared	0.13		0.21	

* parameter values not meeting the 95% significance level

t-test against base value of 1

TABLE 6 Model Estimation Results Using Combined SP and RP Data

	NL		Mixed NL	
	Coef.	**t-stat**	**Coef.**	**t-stat**
$\alpha_{bikeshare}$ (SP)	1.64	8.62	1.89	10.19
α_{walk} (SP)	1.82	8.57	1.91	9.43

α_{ebike} (SP)	0.33	1.97	0.75	4.79
$\alpha_{carshare}$ (SP)	- 21.9	- 3.81	- 1.66	- 2.59
α_{car} (SP)	0.11	0.61	0.50	2.79
$\alpha_{bikeshare}$ (RP)	- 0.04	- 0.42	0.24	2.88
α_{bike} (RP)	- 0.43	- 3.65	0.39	5.01
α_{walk} (RP)	- 0.03	- 0.29	0.45	5.72
α_{ebike} (RP)	- 0.03	- 0.32	0.43	5.61
$\alpha_{cardriver}$ (RP)	- 0.72	- 5.26	0.16	2.08
$\alpha_{carpassenger}$ (RP)	- 1.29	- 7.11	- 0.05	- 0.56
Natural environmental conditions				
Air pollution-bikeshare (SP)	- 0.0048	- 8.89	- 0.0045	- 8.29
Air pollution-walk (SP)	- 0.0046	- 9.24	- 0.0045	- 9.17
Air pollution-ebike (SP)	- 0.0029	- 5.01	- 0.0022	- 3.93
Air pollution-bus (SP)	- 0.0052	- 6.06	- 0.0020	- 2.65
Air pollution-carshare (SP)	0.0274	3.27	0.0023	1.96
Rain-bikeshare (SP & RP)	- 0.15	- 6.37	- 0.10	- 3.89
Rain-walk (SP & RP)	- 0.48	- 4.41	- 0.62	- 6.99
Rain-ebike (SP & RP)	- 0.26	- 1.71*	- 0.40	- 2.77
Rain-carshare (SP)	8.60	3.91	1.26	4.11
Rain-car (SP & RP)	0.88	8.37	0.41	8.32
Temperature-bikeshare (SP)	0.01	2.19	0.01	3.16
Temperature-walk (SP)	0.01	1.67*	0.01	4.12
Temperature-carshare (SP)	- 0.27	- 4.45	- 0.05	- 4.95
Temperature-car (SP)	- 0.03	- 6.04	- 0.02	- 4.37
Trip and mode attributes				
Commute-bikeshare (SP & RP)	- 0.12	- 5.36	- 0.18	- 10.27
Commute-walk (SP & RP)	0.05	2.83	0.06	7.90
Commute-car (SP & RP)	0.30	6.66	0.03	2.48
Travel cost-bikeshare (SP & RP)	- 0.61	- 6.69	- 0.72	- 8.33
Travel cost-bus (SP & RP)	- 0.15	- 1.42*	- 0.10	- 0.10*
Travel cost-carshare (SP)	- 1.66	- 3.40	- 0.30	- 3.16
Travel cost-car (SP & RP)	- 0.12	- 2.11	- 0.04	- 1.22*
Parking cost-car (SP)	- 0.04	- 2.17	- 0.03	- 1.66*
Travel time-bikeshare (SP & RP)	0.06	6.60	0.04	5.75
Travel time-bike (RP)	0.11	7.93	0.05	6.16
Travel time-walk (SP & RP)	- 0.02	- 6.58	- 0.01	- 5.56
Travel time-ebike (SP & RP)	0.14	6.94	0.09	5.83

Travel time-bus (SP & RP)	0.08	7.85	0.05	6.04
Travel time-carshare (SP)	0.36	2.04	0.07	2.01
Travel time-car (SP & RP)	0.09	5.50	0.07	6.26
Access time-bikeshare (SP)	- 0.09	- 5.09	- 0.09	- 4.58
Access time-bus (SP)	- 0.08	- 5.05	- 0.10	- 6.78
Access time-carshare (SP)	- 0.07	- 0.35*	- 0.05	- 1.57*
App availability-bikeshare (SP)	- 0.66	- 8.49	- 0.66	- 8.14
App availability-bus (SP)	0.07	0.82*	0.33	4.51
App availability-carshare (SP)	2.38	3.08	0.27	1.96
Systematic taste heterogeneity				
Air pollution * Male-bus (SP)	- 0.0016	- 4.84	- 0.0010	- 3.23
Air pollution * Lower age-bus (SP)	0.0025	6.53	0.0010	2.89
Air pollution * Lower income-bus (SP)	0.0014	2.40	0.0005	0.94*
Commute * Lower income-car (SP & RP)	- 0.41	- 7.10	- 0.01	- 0.01*
Commute * Lower education-walk (SP & RP)	- 0.17	- 6.59	- 0.02	- 3.47
Inter-alternative correlation & Panel effect				
$\mu_{sharedmotor}$ (SP)	2.21	4.91#	1.68	4.89#
$\sigma_{bikeshare}$ (SP & RP)	-	-	1.51	10.88#
σ_{walk} (SP & RP)	-	-	1.05	7.04#
σ_{ebike} (SP & RP)	-	-	1.31	12.32#
σ_{bus} (SP & RP)	-	-	1.74	14.01#
σ_{car} (SP & RP)	-	-	1.15	7.20#
Scaling factor (RP)	4.83	7.93#	5.96	9.53#
Number of observations	15642		15642	
Initial log-likelihood	- 24788.3		- 24788.3	
Final log-likelihood	- 21010.1		- 16994.7	
Likelihood ratio test	7556.4		15587.1	
Adjusted rho-bar squared	0.15		0.31	

* parameter values not meeting the 95% significance level

t-test against base value of 1

6. POLICY IMPACT ANALYSIS

A number of scenarios are proposed to help explore the effectiveness of different policy options aiming at increasing bike-sharing ridership.The model estimation results of the mixed NL model based on combined SP and RP data are used for simulation. The simulation method is sample enumeration.

A key objective is to find out to what extent an improvement in air quality would promote bike-sharing usage. To begin with, 20% air quality increase is set as a mid-term target in our policy scenarios in accordance with the air pollution reduction target in China (Zhang, 2017). Specifically, the central government has set a 2012-2017 five-year plan to decrease the air pollution levels in the country's top 3 city clusters (i.e. Beijing-Tianjin-Hebei cluster, the Yangtze cluster centered by Shanghai and the Pearl cluster centered by Guangzhou) by 25%, 20% and 15% respectively. As a result, the median target (20%) is selected as the reference for this study. Next, a 50% air quality increase is proposed as a long-term target. It is based on the fact that coal burning accounts for 50%-70% of air pollution in the above mentioned 3 clusters (Wang, 2014). Thus, a 50% air quality increase is set to represent an optimistic "coal free era" in the long-term.

To generate broader insights, measures for bike-sharing service improvement are also proposed. As per the model estimation results, reductions in travel cost and access time are introduced and joint with air quality improvement to create more scenarios for analysis. Table 7 shows the simulation results and the key insights are identified as follows:

- Firstly, better air quality can indeed improve the demand for bike-sharing (Baseline to M1 and L1); meanwhile the demand for walking also rises whereas private car usage drops. However, by comparing to the rest of scenarios (M2-M5 and L2-L5), it is easily noticed that air quality improvement is less effective than bike-sharing service improvement (e.g. access time saving, travel cost saving) in promoting bike-sharing ridership.

- Secondly, a saving in access time to bike-sharing parking spots appears to be more effective than a saving in bike-sharing travel cost in short-distance trips. In M4 and M5 (or L4 and L5) when access time reduction starts to intervene, bike-sharing ridership rises more significantly than M2 and M3 (or L2 and L3). The elasticity analysis in Table 8 reflects the same fact that the probability to choose bike-sharing is more elastic to a change in access time (-0.274) than a change in travel cost (-0.118).

- Finally, by looking through M2-M5 and L2-L5 (i.e. measures focusing on bike-sharing service improvement), it is seen that the increases in bike-sharing demand largely come from the shrinking demand for walking and bus rather than private car. The cross elasticity values also reveal the same trend (Table 8). Such a discovery leads to an interesting choice in policy making: the improvement of bike-sharing service (e.g. access time saving, travel cost saving) is more effective than air quality improvement in promoting bike-sharing usage; however, the latter is on the other hand more useful in suppressing private car demand as the figures show. Hence, since all policy measures come with costs it should be policy makers' discretion to prioritize target and make use of the two options.

TABLE 7 Scenarios and Modal Splits

Scenarios		
Mid-term	M1	20% air quality increase
	M2	20% air quality increase + 20% bike-sharing travel cost saving
	M3	20% air quality increase + 50% bike-sharing travel cost saving
	M4	20% air quality increase + 50% bike-sharing travel cost saving + 20% bike-sharing access time saving
	M5	20% air quality increase + 50% bike-sharing travel cost saving + 50% bike-sharing access time saving
Long	L1	50% air quality increase
	L2	50% air quality increase + 20% bike-sharing travel cost saving

	L3	50% air quality increase + 50% bike-sharing travel cost saving					
	L4	50% air quality increase + 50% bike-sharing travel cost saving + 20% bike-sharing access time saving					
	L5	50% air quality increase + 50% bike-sharing travel cost saving + 50% bike-sharing access time saving					
Modal Splits							
		Bike-sharing	Walk	Electric bike	Bus	Car-sharing	Car
	Baseline	21.5%	30.2%	9.2%	28.8%	2.4%	7.9%
Mid-term	M1	22.0%	30.9%	9.1%	28.7%	1.9%	7.4%
	M2	22.6%	30.7%	9.0%	28.5%	1.9%	7.4%
	M3	23.4%	30.4%	8.9%	28.1%	1.8%	7.4%
	M4	24.7%	29.8%	8.8%	27.7%	1.8%	7.2%
	M5	26.7%	28.9%	8.6%	27.0%	1.8%	7.0%
Long-term	L1	22.7%	31.7%	8.8%	28.7%	1.4%	6.7%
	L2	23.2%	31.5%	8.8%	28.5%	1.4%	6.6%
	L3	24.1%	31.2%	8.7%	28.1%	1.4%	6.5%
	L4	25.4%	30.6%	8.6%	27.6%	1.4%	6.4%
	L5	27.4%	29.7%	8.3%	26.9%	1.3%	6.3%

TABLE 8 Direct and Cross Point Elasticity

Choice probability of	Bike-sharing travel cost	Bike-sharing access time
Bike-sharing (direct)	- 0.118	- 0.274
Walk (cross)	0.038	0.084
Bus (cross)	0.035	0.072
Car (cross)	0.034	0.066

7. CONCLUSIONS

This study investigated the factors affecting mode choice behavior in Taiyuan (China) with a focus on bike-sharing choice. Based on the combined SP and RP short-distance trip data, NL and Mixed NL models were developed to study the impacts of natural environmental conditions, trip and mode attributes as well as systematic taste heterogeneity on mode choices. In the end, the potential impacts of a number of policy options on modal split changes were analyzed.

The mixed NL model well addressed the inter-alternative correlation between bus and car-sharing as well as the panel effect caused by repeated choice observations. The incorporation of RP data into SP data significantly increased the model performance and the credibility of model estimation results. The signs of coefficients are in general consistent between the SP alone models and the models using combined SP and RP data. Several key insights were generated for bike-sharing choice. People would be more likely to use the service if air quality was better; the service users also favored warmer weather and disliked rain; bike-sharing appeared to be a more popular choice in leisure trips rather than commute trips; lower travel cost and shorter access time to parking spots would encourage its ridership. Moreover, by comparing the results to the existing findings in developed countries, a significant difference was revealed with respect to socio-economic factors. Bike-sharing choice was often significantly associated with particular socio-economic groups as shown in the literature. In this research by examining through systematic taste heterogeneity, none of the socio-economic groups significantly interacted with any factors affecting bike-sharing choice. The finding was however in line with the earlier study in Beijing (Campbell et al., 2016), in which the results also showed the users of bike-sharing service could arise anywhere from the social spectrum.

The policy impact analysis offered more intuitive information to policy makers. In short-distance trips, improving bike-sharing service itself (e.g. access time saving, travel cost saving) would be more effective than improving air quality for promoting bike-sharing usage. To take one step further, access time saving was found to be more effective than travel cost saving. Nevertheless, if suppressing private car usage was also a policy target, then air quality improvement could be reconsidered since it was more effective than bike-sharing service improvement which was more likely to bring down the demand for walking and bus rather than private car.

Overall, this study is one of the first works that explores air pollution's impact on mode choice behavior as well as factors affecting bike-sharing choice in a developing country. The findings could benefit policy making by revealing the effectiveness of different policy options, although how to deliver the proposed policy options in reality remains as a challenge to policy makers and such an issue is beyond the scope of this work. Cities with close characteristics to Taiyuan could benefit the most from the results and the insights. Researchers from developing countries could also make use of the methodologies in this research to study similar issues in their own cases; especially in cities that have overt local and geographical differences to Taiyuan.

ACKNOWLEDGEMENTS

We sincerely appreciate the support from Shanxi Transportation Research Institute in their funding and advice provided during the data collection. We want to thank the following individuals who made the most significant contributions in the data recording task: Mr Li Peiyu from Shanxi Experimental Secondary School, Mr Hou Juntao from Peking University and Ms Zhao Helan. We would like also to thank the two reviewers of this paper. Finally, we express our gratitude and special thanks to the editor, Prof Juan de Dios Ortúzar, for providing us valuable comments to improve this work.

REFERENCES

Ahillen, M. Mateo-Babiano, D. Corcoran, J. (2016) Dynamics of bike sharing in Washington, DC and Brisbane, Australia: Implications for policy and planning. International journal of sustainable transportation, 10(5), pp.441-454.

Akar, G. Clifton, K. (2009) Influence of Individual Perceptions and Bicycle Infrastructure on Decision to Bike. Transportation Research Record: Journal of the Transportation Research Board, 2140, pp.165-172.

Akar, G., Fischer, N. Namgung, M. (2013) Bicycling Choice and Gender Case Study: The Ohio State University. International Journal of Sustainable Transportation, 7(5), pp.347-365.

Amador, F.J. González, R.M. Ortúzar, J. de D. (2005) Preference heterogeneity and willingness to pay for travel time savings. Transportation, 32(6), pp.627-647.

Baker, L. (2009) How to get more bicyclists on the road; to boost urban cycling, figure out what women want. Scientific American Magazine.

Baltes, M. (1996) Factors Influencing Nondiscretionary Work Trips by Bicycle Determined from 1990 U.S. Census Metropolitan Statistical Area Data. Transportation Research Record: Journal of the Transportation Research Board, 1538, pp.96-101.

Barnes, G. Krizek, K. (2005) Estimating Bicycling Demand. Transportation Research Record: Journal of the Transportation Research Board, 1939, pp.45-51.

Ben-Akiva, M. Bradley, M. Morikawa, T. Benjamin, J. Novak, T. Oppewal, H. Rao, V. (1994) Combining revealed and stated preferences data. Marketing Letters, 5(4), pp.335-349.

Bhat, C. Sardesai, R. (2006) The impact of stop-making and travel time reliability on commute mode choice. Transportation Research Part B: Methodological, 40(9), pp.709-730.

Bierlaire, M. (2016) PythonBiogeme: a short introduction. Report TRANSP-OR 160706, Series on Biogeme. Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Bliemer, M. Rose, J. (2010) Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Transportation Research Part B: Methodological, 44(6), pp.720-734.

Bliemer, M. Rose, J. Hensher, D. (2009) Efficient stated choice experiments for estimating nested logit models. Transportation Research Part B: Methodological, 43(1), pp.19-35.

Bradley, M. Daly, A. (1997) Estimation of logit choice models using mixed stated preference and revealed preference information. Understanding travel behaviour in an era of change, pp.209-232.

Campbell, A.A. Cherry, C.R. Ryerson, M.S. Yang, X. (2016) Factors influencing the choice of shared bicycles and shared electric bikes in Beijing. Transportation research part C: emerging technologies, 67, pp.399-414.

Caussade, S. Ortúzar, J. de D. Rizzi, L.I. Hensher, D.A. (2005) Assessing the influence of design dimensions on stated choice experiment estimates. Transportation research part B: Methodological, 39(7), pp.621-640.

Cherchi, E. Ortúzar, J. de D. (2011) On the use of mixed RP/SP models in prediction: Accounting for systematic and random taste heterogeneity. Transportation science, 45(1), pp.98-108.

Daito, N. Chen, Z. (2013) Demand of Bike-sharing Travels: Evidence from Washington, DC. In Transportation Research Board 92nd Annual Meeting, No.13-3869.

De Chardon, C.M. Caruso, G. Thomas, I. (2017) Bicycle sharing system 'success' determinants. Transportation Research Part A: Policy and Practice, 100, pp.202-214.

Deenihan, G. Caulfield, B. (2015) Do tourists value different levels of cycling infrastructure?. Tourism Management, 46, pp.92-101.

DeMaio, P. (2009) Bike-sharing: History, Impact, Models of Provision, and Future. Journal of Public Transportation, 12(4), pp.41-56.

DeMaio, P. Gifford, J. (2004) Will Smart Bikes Succeed as Public Transportation in the United States? Journal of Public Transportation, 7(2), pp.1-15.

Faghih-Imani, A., Hampshire, R., Marla, L. Eluru, N. (2015) An Empirical Analysis of Bike-sharing Usage and Rebalancing: Evidence from Barcelona and Seville. SSRN Electronic Journal.

Fishman, E. Washington, S. Haworth, N. Watson, A. (2015) Factors influencing bike share membership: an analysis of Melbourne and Brisbane. Transportation research part A: policy and practice, 71, pp.17-30.

Gu Dong. (2016) Bike-sharing report: riding data. Available at: http://www.ichuangye.cn/index.php/article-edit-2569-21.html

Hankey, S., Lindsey, G., Wang, X., Borah, J., Hoff, K., Utecht, B. Xu, Z. (2012) Estimating use of non-motorized infrastructure: Models of bicycle and pedestrian traffic in Minneapolis, MN. Landscape and Urban Planning, 107(3), pp.307-316.

Hensher, D. Bradley, M. (1993) Using stated response choice data to enrich revealed preference discrete choice models. Marketing Letters, 4(2), pp.139-151.

Hensher, D. Greene, W. (2003) The mixed logit model: the state of practice. Transportation, 30(2), pp.133-176.

Hensher, D. Rose, J. Greene, W. (2005) Applied Choice Analysis. Cambridge: Cambridge University Press.

Hess, S. Bierlaire, M. Polak, J. (2004) Development and application of a mixed cross-nested logit model. In 21th European Transport Conference (No. TRANSP-OR-CONF-2006-044).

Hess, S. Bierlaire, M. Polak, J. (2005) Estimation of value of travel-time savings using mixed logit models. Transportation Research Part A: Policy and Practice, 39(2), pp.221-236.

Jäppinen, S. Toivonen, T. Salonen, M. (2013) Modeling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach. Applied Geography, 43, pp.13-24.

Larsen, J. El-Geneidy, A. (2011) A travel behavior analysis of urban cycling facilities in Montréal, Canada. Transportation Research Part D: Transport and Environment, 16(2), pp.172-177.

Lavasani, M. Hossan, M. Asgari, H. Jin, X. (2017) Examining methodological issues on combined RP and SP data. Transportation Research Procedia, 25, pp.2335-2348.

Lin, J. Yang, T. (2011) Strategic design of public bicycle sharing systems with service level constraints. Transportation Research Part E: Logistics and Transportation Review, 47(2), pp.284-294.

Kamargianni, M. (2015) Investigating next generation's cycling ridership to promote sustainable mobility in different types of cities. Research in Transportation Economics, 53, pp.45-55.

Kamargianni, M. Polydoropoulou, A. (2013) Hybrid Choice Model to Investigate Effects of Teenagers' Attitudes towards Walking and Cycling on Mode Choice Behavior. Transportation Research Record: Journal of the Transportation Research Board, 2382, pp.151-161.

Kamargianni, M. Polydoropoulou, A. (2014) Generation's Y Travel Behavior and Perceptions Towards Walkability Constraints. Transportation Research Record: Journal of the Transportation Research Board, 2430, pp.59-71.

Maness, M. Cirillo, C. Dugundji, E. (2015) Generalized behavioral framework for choice models of social influence: Behavioral and data concerns in travel behavior. Journal of Transport Geography, 46, pp.137-150.

Mateo-Babiano, I. Bean, R. Corcoran, J. Pojani, D. (2016) How does our natural and built environment affect the use of bicycle sharing?. Transportation Research Part A: Policy and Practice, 94, pp.295-307.

Maurer, L. (2012) Feasibility study for a bicycle sharing program in Sacramento, California. In Transportation Research Board 91st Annual Meeting, No.12-4431.

McFadden, D. Train, K. (2000) Mixed MNL models for discrete response. Journal of applied Econometrics, pp.447-470.

Motoaki, Y. Daziano, R. (2015) A hybrid-choice latent-class model for the analysis of the effects of weather on cycling demand. Transportation Research Part A: Policy and Practice, 75, pp.217-230.

Moudon, A., Lee, C., Cheadle, A., Collier, C., Johnson, D., Schmid, T. Weather, R. (2005) Cycling and the built environment, a US perspective. Transportation Research Part D: Transport and Environment, 10(3), pp.245-261.

Oortwijn, J. (2017) With $700 Million Chinese Bike Sharing Firm ofo Targets Europe. Available at: http://www.bike-eu.com/home/nieuws/2017/7/with-700-million-chinese-bike-sharing-firm-ofo-targets-europe-10130515

Ortúzar, J. de D. Willumsen, L.G. (2011) Modelling transport. New Jersey: Wiley, 4th Edition.

Parkin, J., Wardman, M. Page, M. (2008) Estimation of the determinants of bicycle mode share for the journey to work using census data. Transportation, 35(1), pp.93-109.

Poon, L. (2017) Rebuilding the 'Kingdom of Bicycles'. Available at: https://www.citylab.com/transportation/2017/04/bikeshare-save-biking-in-china/521181

Ricci, M. (2015) Bike-sharing: A review of evidence on impact and processes of implementation and operation. Research in Transportation Business & Management.

Rietveld, P. Daniel, V. (2004) Determinants of bicycle use: do municipal policies matter?. Transportation Research Part A: Policy and Practice, 38(7), pp.531-550.

Rodriguez, D. Joo, J. (2004) The relationship between non-motorized mode choice and the local physical environment. Transportation Research Part D: Transport and Environment, 9(2), pp.151-173.

Rose, J. Bliemer, M. (2009) Constructing efficient stated choice experimental designs. Transport Reviews, 29(5), pp.587-617.

Salomon, I. Mokhtarian, P.L. (1998) What happens when mobility-inclined market segments face accessibility-enhancing policies?. Transportation Research Part D: Transport and Environment, 3(3), pp.129-140.

Saneinejad, S. Roorda, M. Kennedy, C. (2012) Modeling the impact of weather conditions on active transportation travel behavior. Transportation Research Part D: Transport and Environment, 17(2), pp.129-137.

Shafizadeh, K. Niemeier, D. (1997) Bicycle Journey-to-Work: Travel Behavior Characteristics and Spatial Attributes. Transportation Research Record: Journal of the

Transportation Research Board, 1578, pp.84-90.

Shaheen, S., Guzman, S. Zhang, H. (2010) Bikesharing in Europe, the Americas, and Asia. Transportation Research Record: Journal of the Transportation Research Board, 2143, pp.159-167.

Song, L. (2015) Taiyuan bike-sharing model. Available at: http://www.chinanews.com/sh/2015/05-13/7273904.shtml

Taiyuan Public Transport Holdings. (2016) Bike-sharing service data. Annex 6: The statistics of App download. Taiyuan: China. Available at: http://www.tybus.com

Tang, Y., Pan, H., Shen, Q. (2011) Bike-Sharing Systems in Beijing, Shanghai, and Hangzhou and Their Impact on Travel Behavior. In 90th Annual Meeting of the Transportation Research Board, Washington, DC.

Waldman, J. (1977) Cycling in Towns. Quantitative Investigation, No. LTR Working Paper 3 Monograph.

Wang, Y.M. (2014) China can reduce carbon and air pollution without harming economy, study finds. Available at: http://newclimateeconomy.net/content/release-china-can-reduce-carbon-and-air-pollution-without-harming-economy-study-finds

Wang, C., Akar, G. Guldmann, J. (2015) Do your neighbours affect your bicycling choice? A spatial probit model for bicycling to The Ohio State University. Journal of Transport Geography, 42, pp.122-130.

Whalen, K., Páez, A. Carrasco, J. (2013) Mode choice of university students commuting to school and the role of active travel. Journal of Transport Geography, 31, pp.132-142.

Xing, Y., Handy, S. Mokhtarian, P. (2010) Factors associated with proportions and miles of bicycling for transportation and recreation in six small US cities. Transportation Research Part D: Transport and Environment, 15(2), pp.73-81.

Zahran, S., Brody, S., Maghelal, P., Prelog, A. Lacy, M. (2008) Cycling and walking: Explaining the spatial distribution of healthy modes of transportation in the United States. Transportation Research Part D: Transport and Environment, 13(7), pp.462-470.

Zhang, C. (2017) Can China meet its 2017 air quality goals?. Available at: https://www.chinadialogue.net/article/show/single/en/9574-Can-China-meet-its-2-17-air-quality-goals-

APPENDIX A: An example of the two short-distance scenarios as seen by a respondent (translated from Chinese)

Scenario 1: Travel within 2km, to work/education, sunny day, 10° C, with light pollution

	Car	**E-bike**	**Bus**	**Car share**	**Bike share**	**Walk**
	Travel 3 min	Ride 5 min	Travel 5 min	Travel 7 min	Ride 8 min	Walk 20 min
	Fuel ￥1.2		Ticket ￥1	Cost ￥3	Cost ￥0	
	Easy to park car					
	Parking ￥5/h					
			Walk 5 min to station	Walk 5 min to station	Walk 2 min to station	
			Every 2 min			
			With app	With app	With app	
Your choice (please tick)						

Scenario 2: Travel within 2km, to shopping, snowy day, -10° C, with excellent air quality

	Car	**E-bike**	**Bus**	**Car share**	**Bike share**	**Walk**
	Travel 7 min	Ride 5 min	Travel 12 min	Travel 7 min	Ride 8 min	Walk 15 min
	Fuel ￥1.6		Ticket ￥1	Cost ￥1	Cost ￥1	
	Hard to park car					
	Parking ￥5/h					
			Walk 5 min to station	Walk 10 min to station	Walk 5 min to station	
			Every 2 min			
			With app	With app	Without app	
Your choice (please tick)						

APPENDIX B: Sample Data versus Census Data after the 2-stage Calibration (census data source: Shanxi Statistical Yearbook 2014, available at: China Statistics Press, http://csp.stats.gov.cn/)

Districts of Taiyuan	**Sample**			**Census**		
	Population	Male	Female	Population	Male	Female
						In: number of people
Xiaodian	2,293	1,192	1,101	820,004	429,098	390,906
Wanbailin	2,091	1,066	1,025	765,956	390,413	375,543
Xinghualing	1,794	879	915	653,854	321,154	332,700
Yingze	1,632	816	816	601,109	299,120	301,989
Jiancaoping	1,127	741	386	424,294	205,182	219,112
Jinyuan	562	238	324	225,849	115,219	110,630
Total	9,499			3,491,066		
						In: percentage
Xiaodian	24%	52%	48%	23%	52%	48%
Wanbailin	22%	51%	49%	22%	51%	49%
Xinghualing	19%	49%	51%	19%	49%	51%
Yingze	17%	50%	50%	17%	50%	50%
Jiancaoping	12%	66%	34%	12%	48%	52%
Jinyuan	6%	42%	58%	7%	51%	49%
Total	100%			100%		

Note: after dropping out invalid questionnaires (from 15,000 to 9,499), the sample data remains consistent with the census data except for the gender distribution in the least two populated districts “Jiancaoping” and “Jinyuan”.