Article

Scissors Difference of Socioeconomics, Travel and Space Consumption Behavior of Rural and Urban Households and Its Impact on Modeling Accuracy and Data Requirements

Ming Zhong ^{1,2,3} , Qi Tang ^{1,2,3}, Xiaofeng Ma ^{1,2,3,*} and John Douglas Hunt ⁴

¹ Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan 430063, China; mzhong@whut.edu.cn (M.Z.); donaldmickey@whut.edu.cn (Q.T.)

² Engineering Research Center of Transportation Safety, Ministry of Education, Wuhan 430063, China

³ National Engineering Research Center for Water Transportation Safety, Wuhan 430063, China

⁴ Department of Civil Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada; jdhunt@ucalgary.ca

* Correspondence: maxiaofeng@whut.edu.cn; Tel.: +86-027-86582280

Received: 16 July 2019; Accepted: 20 September 2019; Published: 8 October 2019

Abstract: It is believed that the “scissors difference” of socioeconomics between rural and urban households in typical municipalities of China is significant. This may result in differences in their behavior and has important implications for urban land use and transportation planning policies, as well as related modeling accuracy and data requirements. However, detailed analyses regarding such “scissors differences” between rural and urban groups in China have not been done before. In this study, travel survey data collected from the City of Wuhan in 2008 is used to study if rural and urban households are statistically different in terms of household income, household size, space consumption, highest household mobility and travel distance. A set of statistical tests, such as the Kolmogorov–Smirnov test, Mann–Whitney U test and Kruskal–Wallis H test, are applied to the study data. The study results show that the “scissors difference” is found to be statistically significant in terms of household size (HS), household income (HI), building area (BA) consumed and household mobility (except for travel distance) between rural and urban households. Conversely, analyses applied to travel distance of urban and rural household subgroups (categorized by HS and HI) reveal that the urban and rural counterparts show almost exactly opposite behavior. The study results also suggest that such differences should be explicitly considered in relevant modeling exercises by separately setting up urban and rural household groups, but the number of household groups used should be determined based on a balance between modeling accuracy and data required/modeling workload.

Keywords: “scissors difference” of socioeconomics; urban land use; statistical tests; household groups

1. Introduction

The term “scissors difference” was initially proposed to describe the trend of increasing price gaps between industrial products and agricultural products [1]. Since its initial use, the term has been widely applied to describe similar situations in economic research [2,3]. In China, a “scissors difference” policy had been in use since the founding of the People’s Republic of China in 1949, up to approximately the last decade, which involved using the wealth collected from rural areas to subsidize the development of urban areas, i.e., agriculture support industry. Under this policy, urban areas received relatively higher subsidies, while rural areas received lower subsidies [4]. Such a policy has resulted in an unequal exchange of wealth and an unbalanced and uneven distribution of social and economic attributes between urban and rural households, such as salary differences (and therefore

household income), housing consumption and travel behavior. This paper aims to explore this topic and quantify this “scissors difference” between urban and rural households.

Developing countries generally face the challenge of significant urban–rural gaps during their process of industrialization and urbanization [5]. The theory of urban bias holds that urban–rural gaps are exogenous [6] and that development policies implemented by the governments of developing countries are done in response to economic and social issues that mostly favor cities rather than their rural counterparts [7].

Kuznets [8] notes that, as countries develop, income inequality initially increases, peaks and then finally decreases. In developed countries, there is little difference in annual income inequality and income mobility between the rural and urban sectors of the economy [9]. However, in developing countries, a significant income disparity exists between urban and rural areas [10–13]. For example, changes in China’s urban–rural income gap has followed a U-curve pattern, i.e., in the course of economic development, the income gap first decreases then increases [14]. Many researchers have suggested that the level of urbanization has a significantly negative effect in regulating the relationship between the demographic structure of urban and rural areas, and is a key cause of the urban–rural income gap. In regions with low urbanization levels, the increasingly aging population of the area decreases the total social demands and increases the unemployment rate in rural areas, thereby further expanding the urban–rural income gap [15]. Su et al. [16] also found that patterns of interaction between urbanization and the urban–rural income gap varied according to different regions in China. Empirical results showed that urbanization was a “Granger-cause” of an urban–rural income gap, mainly in the eastern region of China. Other researchers found that public expenditure expansion [17], and infrastructure development [18] has a positive influence on reducing the rural–urban income inequality, but that financial development, economic growth and consumer prices aggravate rural–urban income inequality in the long run [19,20].

Urbanization is also accompanied by increasing motorization and longer commuting trips. A spatially unequal distribution of employment opportunities leads to important differences in commuting distances between urban and rural areas [21]. Zhong et al. [22] used travel diary data collected in the Tampa Bay Region of the United States to develop an activity-based model, which was in turn used to simulate travel patterns of all residents in a rural county. They found that the mode choices of urban and rural residents were different. A study by Korzhenevych and Jain [23] revealed that urban residents in India were more likely to use individual motorized transport for both short and long trips, while rural areas had high levels of non-motorized travel modes and a large share of long trips. In contrast, in the U.S., residences and activity sites are widely dispersed in rural areas, resulting in longer trip distances that force rural residents to rely heavily on the car, so mobility levels in rural areas are generally higher than in urban areas [24]. Additionally, a result from a large representative sample in the UK shows that urban residents are more likely to engage in active travel than rural residents [25].

Past research has also shown that there is a relationship between travel behavior and income levels. Several studies report that personal and household income levels are positively associated with vehicle ownership [26,27] and some studies also report negative relationships between public transit use and personal and household income levels at individual and regional scales [28]. Comparisons between vehicle ownership rates of urban and rural households show that rural households are more likely to own a private vehicle than urban households with the same economic standards [29]. Ulfarsson et al. [30] surveyed the 1160 residents in the Reykjavik Capital Region of Iceland and found that about 30% of the participants made fewer trips after a major economic crisis due in part to reduced income and/or unemployment. They also reported that about 20% of the participants perceive bus transit as more important than before, which is, to some extent, at the expense of personal automobiles [30]. These studies clearly suggest that income does have a direct impact on travel behavior.

Current researches have mainly focused on statistical analyses of the differences in income levels and travel behaviors between urban and rural households and have ignored other important factors that are commonly considered in the urban planning field, such as family size, space, vehicle ownership

and trip frequency. Furthermore, they have not discussed what kind of implications or impacts such considerations could have on various planning models, such as travel demand models (TDMs) and integrated land-use transport models (ILUTMs), especially in terms of modeling accuracy and data requirements aspects. These effects are particularly important due to increasing policy requirements to consider such differences. In light of this gap, this paper intends to study if such differences exist between rural and urban households, including among their subgroups. This paper's research uses a set of statistical tests and a travel survey dataset from the city of Wuhan, China to demonstrate that such differences exist, and evaluates and quantifies the impact of these differences on the modeling process, data requirements, and resulting accuracy.

In this study, a literature review of the differences in socioeconomics and travel behavior of rural and urban households is first provided in the "Introduction" section. A brief introduction outlining the data and research methods are then presented in the Section 2. Then, the results of the study are presented in the Sections 3 and 4, including detailed descriptions and discussions of each test completed. Finally, the major findings, conclusions and future areas of research of this study are given.

2. Methodology

2.1. Study Area

This study uses data from a travel survey conducted by the City of Wuhan in 2008. Wuhan is the largest city in central China and is the capital of Hubei province. It is the largest transportation hub in China and serves as a gateway connecting all corners of the country. The data used in this study is aggregated on the basis of traffic analysis zones (TAZs), which divide the study area into smaller zones with similar characteristics. Points of interest (POI) are used together with zonal characteristics to identify the home zone of a given set of households as either rural or urban. The survey's raw data assigns each household a unique household ID and identifies the ID of the TAZ wherein it is located. Geographic Information System (GIS) is used to analyze and visualize the locations of each TAZ and categorize them into urban (UZ) or rural zones (RZ). TAZs located within the Third Ring Road and any other TAZs with commercial and office buildings (identified from the POI data) were categorized as urban zones (UZ) and all others were categorized as rural zones (RZ). The entire study area includes 690 TAZs, of which 399 were categorized as urban zones and 291 as rural zones, as shown in Figure 1.

Figure 1. Traffic analysis zone (TAZ) map of the City of Wuhan (urban TAZs are highlighted in red and rural TAZs are highlighted in blue).

2.2. Data

The travel survey data contains data from a total of 36,300 households and includes two types of data, socioeconomic data and travel behavior data. The socioeconomic data includes the household

size, household income, housing type, and highest household mobility. This data also provides the age, education level, gender, and marital status of each household member. The travel behavior data provides the trip-making history of all the household members. Within the data set, all the households are also categorized based on their location as rural or urban, in order to facilitate the analysis of the differences between the two categories. The rural and urban households are further divided into more homogeneous groups in terms of their income level and size (number of people in the household). This data set is used to study the “scissors differences” between rural and urban households in terms of their income, size, housing consumption, uses of transportation vehicles and trip frequency/distance.

Before any analysis work was undertaken, the data was first checked for consistency and 1950 data observations with either missing values or outliers were removed, leaving 34,350 valid observations to be used for further processing and analysis. Both rural and urban households were divided into the following three classes: small-size households (SH, households with less than 2 residents), medium-size households (MH, households with 2 or more residents but less than 7 residents), and large-size households (LH, households with 7 or more residents). Additionally, all households were classified into the following three classes based on their income level: low-income households (LH, the lowest 20% of household income in the whole survey data), medium-income households (MH, the middle 60% of household income in the whole survey data), and high-income households (HH, the highest 20% of household income in the whole survey data).

Table 1 shows the summary statistics of the household attributes considered in this paper, including the minimum, maximum, range, mean, standard deviation (SDV) and coefficient of variation (COV) of household size (HS), income (HI) (unit: 10,000 Yuan), building area (BA) (unit: m²), the highest household mobility mode (HHM) and travel distance (TD) (unit: km) of the urban (U) and rural (R) households. These basic summary statistics can be compared so as to highlight the major differences between the two groups of houses. The COV value provides a measure of the variability of a particular attribute within the major group (urban or rural) with higher values, indicating that there is a larger variation of the attribute within the group. In our results, the mean values of HS, BA and HHM of rural households were higher than those of the urban households. In contrast, the mean values of HI and TD of rural households were lower than those of urban households. In general, the larger BA values for rural households was expected, due to the higher HS values. Rural households also has higher COV values for HI, HS, and TD, but lower COV values for BA and HHM, which indicates that the rural households show a larger variation in the first three aspects, but a smaller variation in the last two aspects.

Table 1. Descriptive statistics for urban and rural household size, income, space consumption, the highest household mobility mode and travel distance.

	U or R	HS	HI	BA	HHM ¹	TD
Min.	U	0	0	6	0	0.01
	R	1	0	10	0	4.6
Max.	U	11	18	255	6	90.2
	R	10	18	250	6	87
Range	U	11	18	249	6	90.2
	R	9	18	240	6	82.1
Mean	U	2.9	5.1	71.8	1.2	27.7
	R	3.2	4.2	108.8	1.8	25.9
SDV	U	0.84	3.8	36.8	1.4	18.7
	R	1.1	3.5	44.7	1.5	18.1
COV (%)	U	28.9	74.5	51.3	116.7	67.5
	R	34.4	83.3	41.1	83.3	69.9

¹ The highest household mobility mode (HHM) refer to the “fastest” transport mode available to a given household, which was divided into the following six grades: bicycle (1), electric bicycle (2), motorcycle (3), other vehicles (4), private car (5) and company car (6). Also, “the other vehicles” refer to truck, tractor, etc.

2.3. Estimation Techniques

(1) Kolmogorov–Smirnov (K-S) test

The one-sample K-S test is a non-parametric test of the equality of continuous, one-dimensional probability distributions, which is typically used to compare a sample with a reference probability distribution [31]. The normal distribution is applied as a reference distribution and the test is applied to determine if the survey data from the city of Wuhan is normally distributed. If the data is shown to be normally distributed, a parametric test such as z-test could then be applied to test whether there are significant differences between urban and rural areas considering household income (HI), household size (HS), building area (BA), highest household mobility (HHM) and travel distance (TD). If the data is shown to not be normally distributed, then non-parametric tests will be used to check the existing problem of significant differences.

(2) Mann–Whitney U (MWW) test

In this work, a non-parametric test is used for the comparison of the distribution of two independent samples in the case where there is no clear understanding of the overall distribution. In statistics, the MWW test is a non-parametric test of the null hypothesis, that two samples come from the same population against an alternative hypothesis. Particularly, the test is often used to show that a population may have larger values than the other [31]. The MWW test does not need the data to obey normal distribution.

(3) Kruskal–Wallis H test

The Kruskal–Wallis H test is the nonparametric equivalent of the analysis of variance F test. It tests the null hypothesis that all k populations possess the same probability distribution against the alternative hypothesis that the distribution differs in its location. Compared with the F test, it does not need the assumptions of the nature of the sampled populations [31]. It is also the extension of the MWW test if the number of groups is over two, and its parametric equivalent is the one-way analysis of variance. An important Kruskal–Wallis H test shows that at least one sample randomly dominates another sample. One of the advantages of using the Kruskal–Wallis H test on this sample of data is that there are 9 groups in each of the urban and rural households that are defined in terms of HS and HI (e.g., small-size and low-income household) and this test can be used to identify whether any statistically significant differences exist among each group.

3. Accuracy and Data Requirement Assessment

After quantifying the “scissors differences” in the socioeconomics, and travel and space consumption behavior between the rural and urban households, the impact of such differences on an integrated model’s accuracy was examined, based on the number of households considered and data required. This study analyzes data using the Wuhan PECAS model, which represents the production, exchange and consumption allocation system and is one of the latest ILUTM models, developed on the basis of previous experiences with MEPLAN (Marcial Echenique & Partners’ software package) and TRANUS (Transporte y Uso del Suelo). Although the PECAS framework was used to demonstrate these impacts, these results are universally applicable to any travel demand or integrated land use transport models. This section outlines the design diagram of this model and provides detail on the way that the modeling accuracy and data demand was calculated. Within this process, the following model scenarios were considered:

3.1. Scenario 1: ILUTM-1

Under this scenario, only the following two household groups were considered in the Wuhan PECAS model: urban and rural.

3.2. Scenario 2: ILUTM-2

Under this scenario, households in the city of Wuhan were classified into the following four groups according to urban/rural status and household size: (1) small-size urban, (2) large-size urban, (3) small-size rural, and (4) large-size rural. “Small-size” households were defined as having less than 5 residents while “large-size” households had 5 or more residents.

3.3. Scenario 3: ILUTM-3

Under this scenario, households were classified into six groups according to their urban/rural designation and household size (small, medium and large). “Small-size” in this scenario was defined as any household with less than 3 residents, “medium-size” was defined as any household with 3 or more residents but less than 7 residents, and the “large-size” was any household with 7 or more residents.

3.4. Scenario 4: ILUTM-4

Under this scenario, households were classified into eight groups according to their urban/rural designation and household size (small and large) and household income. Household income was defined with two levels: “below average” and “above average”. Average income was estimated from the survey data as 50k/year.

3.5. Scenario 5: ILUTM-5

Under this scenario, households were divided into 12 groups based on their urban/rural designation, household size and household income as listed above. Household size had three levels, whereas the household income had two levels according to the categories described in scenarios 3 and 4.

3.6. Scenario 6: ILUTM-6

Under this scenario, 18 household groups were defined, and this scenario served as the most “accurate” household model. Households were grouped by their urban/rural designation, household size and household income. Household size used three levels: small-size, medium-size, large-size and household income also used three levels: low, medium and high.

Figure 2 shows that the design diagram of two PECAS models with the first one, shown in Figure 2a, grouping households into only two categories (urban and rural), while the other, shown in Figure 2b, grouping households into 18 categories (by urban/rural, household size and household income). The only difference between these two models is in the number of household groups considered, and all of the other aspects are identical. The household module interacts with labor production, goods and services consumption, space consumption labor consumption, space consumption and transportation modules. The modeling accuracy is quantified by calculating the within-group mean square error using Equation (1) and the number of data points required is calculated using Equation (2).

In order to assess the modeling accuracy of different household classification methods, the following within-group mean square error (WG-MSE) was used:

$$s^2 = \frac{\sum_{i=1}^{n_1} (y_{i1} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_{i2} - \bar{y}_2)^2 + \cdots + \sum_{i=1}^{n_n} (y_{in} - \bar{y}_n)^2}{n_1 + n_2 + \cdots + n_n - n} = \frac{SSE}{n_1 + n_2 + \cdots + n_n - n'} \quad (1)$$

where

n_n = number of group n from 1 to n

SSE = the sum of squares error

The input data required for developing a full land-use transport model, under the above classification methods, was also calculated based on the PECAS framework, as follows:

$$DD = n_g * (n_s + n_t), \quad (2)$$

where:

n_g = total number of groups used to classify households

n_s = total number of groups used to categorize household socioeconomics data

n_t = total number of groups used to categorize household travel data

Wuhan PECAS Demo Model																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
Design Diagram Structure																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
		PUTS (inputs and outputs)		COMMODITIES AND SERVICES																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
				First Industry-Product					Secondary Industry-Product					Tertiary Industry-Product					Internal management					Government Services					Output																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
				LABOUR					Management & Technical Labour					Retail Service Labour					Outdoors Labour					Operational Labour					Other Labour																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
				SPACE					Residential Space					Space-for-First Industry					Space-for-Second Industry					Space-for-Tertiary Industry					Office Space																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
				C01					C02					C03					C04					C05					L01					L02					L03					L04					L05																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
				S01					S02					S03					S04					S05					S06																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
PRODUCING ACTIVITIES																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
First Industry-Operation	A01																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			

(a)

Figure 2. Cont.

(b)

Figure 2. The ILUTM diagram structure of PECAS: (a) grouping households into only two categories; (b) grouping households into 18 categories.

4. Study Results

4.1. Kolmogorov–Smirnov and Mann–Whitney U Tests

The Kolmogorov–Smirnov test was applied to test whether the distribution of the above household attributes (e.g., HS and TD) was normally distributed. The results of the analysis showed that none of the data were normally distributed, as the test statistics for all data were all under the level of significance of 0.05.

Since the household attributes considered were not normally distributed, the non-parametric Mann–Whitney U test was used to test if there was a statistically significant difference in all attributes between urban and rural households. Table 2 shows the test statistic results for five of the most important household attributes: household size (HS), building area (BA), household income (HI), highest household mobility (HHM), and travel distance (TD). For the first four attributes, the test statistic results were all below 0.05 ($p < 0.05$), which implied that the difference in the values between urban and rural households was statistically significant for these attributes. However, the test statistic for the travel distance (TD) was well above the level of significance of 0.05, indicating that TD values did not differ significantly between urban and rural households.

Table 2. The Mann–Whitney U Test of socioeconomic and behavioral variables between urban and rural households.

	HS	HI	BA	HHM	TD
Mann–Whitney U	84319684.0	87886455.0	50819744.0	77175930.5	775402.0
Z	−27.5	−19.5	−67.8	−35.0	−1.4
Sig.	0.000	0.000	0.000	0.000	0.157

The results for the TD data were both interesting and unexpected, and further analysis was undertaken to determine if the result was correct. A trip length distribution diagram of the sampled urban and rural households was developed and is shown in Figure 3 as a frequency histogram for urban and rural households. The red histograms show the travel distance distribution for urban households and the green ones show the travel distance distribution for the rural. In Figure 3, the average travel distance for urban households is 4.26 km, which is much lower than that of the rural household, 23 km. For urban households, the frequency of trips decreases as the travel distance is increased, implying that many of the trips taken by urban residents are short in nature and less than 20 km. Conversely, it can be found from Figure 3b that more than half of the trips made by rural households are beyond 20 km, with a maximum distance of more than 60 km. From Figure 3, it is clear that there is significant difference between the travel distance of urban and rural households, although the Mann–Whitney U test failed to recognize this difference as statistically significant. On further analysis, it is found that this failure is largely due to the aggregation effect of samples, and the travel distances are found to be statistically different when such a test was carried out at the subgroup level.

Figure 3. Frequency diagram of travel distance distribution of the members of urban and rural households: (a) frequency histogram for urban households; (b) frequency histogram for rural households.

4.2. The Results for Kruskal–Wallis H Test

Next, the HI and HS were used as grouping variables to test how the household size and household income impact the household's behavior in terms of household building area, vehicle ownership rates and travel behavior. Here, both rural and urban households were further classified into three subgroups based on either HS or HI. The Kruskal–Wallis H test instead of the Mann–Whitney U Test was used, as the latter is limited to just two-sample tests. The Kruskal–Wallis H test is a non-parametric method which can be used to compare more than two populations of equal or different sample size.

The results for these comparisons are highlighted in Table 3. When grouping the analysis by household income, the test statistics for all three attributes of the urban households and two out of three of the attributes for rural households (except for TD) met the level of significance requirement of $p < 0.05$. This means that HI has a significant impact on the BA, HHM and TD of urban households and the BA and HHM of the rural households. Similarly, the same analysis was repeated for groupings by household size for an almost identical result with the same significant parameters. It is interesting to note that neither HI nor HS influenced the travel distance of rural households.

Table 3. Kruskal–Wallis H test of the impact of household income (HI) and household size (HS) on urban and rural households in household building area, vehicle ownership, and travel distance.

	URBAN			RURAL		
	BA	HHM	TD	BA	HHM	TD
HI	2210.5	506.3	17.4	25.9	13.5	0.9
(Sig.)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.616)
HS	127.0	380.5	28.1	152.1	215.4	2.1
(Sig.)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.142)

After considering the impact of household size and household income separately, both household size and household income were then considered as an overall grouping variable. Each of the urban and rural households were further classified into 9 subgroups based on the combinations of the two categorical variables: small-size and low-income (SL), small-size and medium-income (SM), small-size and high-income (SH), medium-size and low-income (ML), medium-size and medium-income (MM), medium-size and high-income (MH), large-size and low-income (LL), large-size and medium-income (LM), large-size and high-income (LH). The Kruskal–Wallis H test again was applied to the 9 urban or rural subgroups to check if the interaction effect of HI and HS makes a difference in their travel distance, and the results are shown in Table 4.

Table 4 shows the *p*-values from the Kruskal–Wallis H tests for comparing the average travel distance of household members among the 9 types of urban and 9 types of rural households classified according to their size and income. The red part shows the test results for the travel distance among urban households and the green part shows the results of the same test for rural households. It should be noted that there was not enough data for two of the large household groups (LL and LH groups, in both urban and rural areas) and therefore the corresponding test statistics are unavailable for these groups.

For households in urban areas (red part of Table 4), it can be observed that the *p*-values for the small and medium households generally meet the level of significance value of $p < 0.05$. For example, the *p*-values between the SM and SH group is 0.000, which is less than 0.05 and implies that there is a statistically significant difference between the travel distances of SM and the SH households. In contrast, for the large-size and medium-income (LM) households, none of the comparisons meet these criteria, except for the case when LM is compared to the SL group, as shown by the circle on the bottom of the Table 4. Similarly, the *p*-value between the SM and the LM group is 0.099, which means there is no significant difference between the average trip distances made by members of the SM and the LM households. This suggests that the income gap between the medium-income households and the high-income households is not significant and therefore, the travel distance of their household members is similar to each other. It also appears that the income thresholds used for classifying households into different income groups may also play a role in determining whether the members from different household groups would show a similar or different travel distance. This will be further examined in future study.

It is interesting to note that the household members from both urban and rural SL group show a significantly lower average travel distance when compared to all other groupings of urban and rural household groups. It appears that urban and rural SL households are among the most disadvantaged groups and therefore their average travel distances are consistently lower than those of the other groups. There are several situations that commonly give rise to such families in China. One example is the case of elderly people living alone who do not need to work and live on a retirement pension. Additionally, they may have health issues and often do not have a driver's license (this is common in China), thus making long trips more challenging for them. Similarly, a second example is low-income young and middle-aged families without children. These families typically have low-wage jobs, such as waiters or farmers, and their work locations are often close to their residences, and thus, they do not need to travel far distances when commuting.

For households in rural areas (green part of Table 4), the results of the Kruskal–Wallis H test show that the rural LM household group have trip distances that are significantly different statistically ($p < 0.05$). LM households were found to make trips that are longer on average than the other 6 rural household groups. Again, it should be noted that there are not enough data for the other two large household groups and therefore only the test statistic for the rural large household group—LM—is available. On the other hand, the analysis shows that the five rural household groups (SM, SH, ML, MM, MH) do not have statistically significantly different travel distances. For example, the p -values between the SM and SH group is 0.147, which is higher than the threshold of 0.05, implying that, statistically, there is no difference between the small-medium and the small-high households in their travel distance. This result is interesting as the results may imply that the household size and income of these groups still allow them to live similar lifestyles and make trips of similar distances.

Table 4. Kruskal–Wallis H Test of the impact of HI and HS impact on average travel distance of the members of urban and rural households.

	SL	SM	SH	ML	MM	MH	LL	LM	LH
SL	---	0.024	0.000	0.000	0.000	0.000	---	0.028	---
SM	0.059	---	0.147	0.801	0.911	0.300	---	0.000	---
SH	0.000	0.000	---	0.202	0.103	0.361	---	0.000	---
ML	0.000	0.008	0.001	---	0.824	0.447	---	0.000	---
MM	0.000	0.002	0.000	0.973	---	0.097	---	0.000	---
MH	0.000	0.000	0.167	0.001	0.000	---	---	0.000	---
LL	---	---	---	---	---	---	---	---	---
LM	0.019	0.099	0.675	0.367	0.357	0.912	---	---	---
LH	---	---	---	---	---	---	---	---	---

It is very interesting to notice from Table 4 that, except for the SL group, the urban household groups show an almost exactly opposite result to that of the rural household groups. For example, the oval and triangle drawn for the urban households on Table 4 show that the test statistics for the five urban household groups are generally significantly different to each other, whereas for those of the LM group, most are not significant. In contrast, the test results for the rural household groups show the opposite result, with the five household groups not being significantly different from each other, and the LM group is significantly different from the others. It appears that, in general, the households in the rural areas tend to have very similar lifestyles (including their travel distance), regardless of their size and income. However, the urban households tend to have a much wider range of lifestyles depending on the socioeconomic status of their household members, including such factors as education attainment, occupation and income. The above opposite results also emphasize that the travel behavior of the rural and urban households is clearly different and should be explicitly considered in the relevant modeling context (developing countries with similar urban/rural “scissors differences”).

Figure 4 shows the box-plot of space consumption patterns for the urban and rural households classified by both household size and income, where the green boxes represent those of the rural households and the red represent the urban households. The household groups number represent SL, SM, SH, ML, MM, MH, LL, LM, LH households numbered from 1 to 9. In Figure 4, it can be seen that the average space consumed by a particular class of households (SL) in the rural area is consistently higher than their counterpart in the urban area. Obviously, the cheaper land and lower rent/price of residential space contributes to such differences. In addition, it is clear from Figure 4 that

the size of the household affects space consumption consistently for rural households, regardless of their income, with larger households generally using more space. In contrast, in urban households, both household size and income level affect how much residential space is consumed. The average space quantities consumed by urban households follow a “wave” pattern, where the consumed space quantity increases within a same household-size group as income rises, then it falls when moving to a lower-income group with a larger size. As can be seen in Figure 4, this “wave” pattern repeats from the small-size group to medium-size groups and then to the large-size group.

Similar Kruskal–Wallis H tests were carried out to compare the residential space consumed by the 9 types of urban and 9 types of rural household subgroups classified according to their size and income (detailed results are not presented here due to limited space). This analysis revealed that, in general, the space consumption of large-size households (LL, LM and LH) and the small and medium-size households with a medium or high income in the urban area are not different to each other in 11 out of 16 cases. It appears that, due to high acquisition cost or rent of urban space, large households, especially those with low or medium income, tend to compromise on their needs for space and have similar space consumption patterns as the small- and medium-size households with a medium and high income. Conversely, it is found that the household size does make a difference in the space consumption in the rural areas, and that household income plays less of a role. The reason appears to be that the space in rural areas is generally inexpensive, and as the size of households increase, they tend to consume more space in order to meet their needs.

Figure 4. Box-plot of building area consumed by urban and rural households (the green boxes for rural households and the red for urban households).

4.3. The Results for Modeling Accuracy and Data Requirements

Figure 5 shows how the mean squared error (MSE) and data requirements (DR) change as the household module of the Wuhan PECAS model is set up under different modeling scenarios (from Scenario 1 to Scenario 6), which offer different number of household categories. Figure 5a shows how MSE and DR values associated with household size and household income change as the number of household categories considered vary across the modeling scenarios. It is clear that, as the number of the household groups considered increases, the MSEs associated with household size and household income decrease, albeit with some minor bumps between similar models. On the other hand, the DR increases consistently and rapidly over this process. For example, Scenario 1’s DR is 40 points of data and the MSE of HS is 0.93; Scenario 2’s DR is 80 points of data and the MSE of HS is 0.53; Scenario 3’s DR is 378 points of data and the MSE of HS is 0.36; Scenario 4’s DR is 504 points of data and the MSE of HS is 0.53; Scenario 5’s DR is 756 points of data and the MSE of HS is 0.36; and Scenario 6’s DR is 1134 points of data and the MSE of HS is 0.35. Figure 5b shows a similar result to Figure 5a,

where a similar trend for MSE and DR related to the consumption behavior of built space is found. The trend lines and corresponding equations show how the MSE values and the data points required change as the number of household categories increases from 2 to 18 across the six model scenarios. The results show that is reasonable to expect that increasing the number of groups used in the model results in higher model complexity and requires more data. The decrease of the MSE implies that the model accuracy increases as the model considers more household categories, but this result comes at the cost of a higher demand for data and an increased modeling workload.

(a)

(b)

Figure 5. Mean squared error (MSE) of the household sub-model and the corresponding data requirements across model scenarios: (a) household size and household income; (b) built space.

5. Discussion

For urban households, household income has become a decisive factor in determining the building area (BA) consumed. However, because of the low cost of living space in rural areas, family size has

been found to be the primary influencer of BA consumed in rural areas. Based on our data, there is little difference between the minimum and maximum BA consumed in urban and rural areas, respectively. However, the average BA consumed in the urban areas is 71.8 square meters, which is much smaller than that of the rural areas (108.8 square meters). This is related to the fact that urban land and housing prices are higher than those in rural areas, and this phenomenon will eventually lead to urban sprawl and suburbanization. In order to obtain better living conditions and take advantage of better transportation systems, urban residents will choose relatively low-priced rural and suburban areas to purchase houses or commercial space, forming new economic growth points and resulting in the continuous expansion of the city.

Wuhan is a new modern metropolitan area and is following the fast urbanization processes previously observed in cities such as Beijing and Shanghai over the past 10 to 20 years. In Figure 3b, we see that long travel distances are common for rural residents, which indicates that some residents still live in the countryside but now work in the city. Zhao et al. [32] found that, in Beijing, living in low-density areas increased the length of commuting trips made to the city center by residents, but that living in compact suburban neighborhoods reduced such needs. Similar observations have also been made by other researchers in other countries [33,34]. Another study by the same authors showed that employment subcenters could reduce the need for suburban workers to commute to the city center [35]. However, Hu et al. [36] found that from 2000 to 2008 in inner Beijing, existing and emerging center areas induced longer commutes than non-center areas. In the inner-ring suburbs, commutes to emerging centers were shorter than those to existing centers. In the outer-ring suburbs, emerging center areas contributed to longer commutes, while existing center areas facilitated shorter commutes. The findings of these researchers are similar to ours, particularly so for the travel distances of rural residents. The results indicate that promoting polycentric urban development could potentially increase commutes in Chinese cities, and the city of Wuhan is not immune to these negative effects.

In the urban areas of the city of Wuhan, residents usually travel by car or public transport. However, in the rural areas, families own many vehicles, but most of them are bicycles, electric vehicles or motorcycles. The difference of mobility choice between urban and rural families appears to be as a result of income differences. However, in developed countries (such as the United States), because the income gap between urban and rural areas is not significant, over 97% of rural households own at least one car vs. 92% of urban households, and 91% of trips are made by car in rural areas vs. 86% in urban areas. Regardless of age, income and race, almost everyone in rural areas relies on the private car for most of their travel needs. Mobility levels in rural areas are generally higher than in urban areas [24]. Therefore, urban–rural differences show different characteristics in different stages of urban development in different countries, which is largely determined by household income.

Additionally, many researchers have shown that urban bias is rising very rapidly in China [37] and other developing countries [38]. Studies suggest that urban sprawl can cause many secondary negative effects and equity issues [39], such as traffic congestion (including its implications for pollution), the loss of open space at the urban fringe, and unrecovered infrastructure costs associated with new low-density development [40,41]. Because of the above issues, researchers are exploring the usefulness of classical urban development theories to mitigate these effects, such as the theory of growth poles [42]. These issues and theories are frequently discussed in the urbanization process of developing countries, and in future studies, these theories can be applied to analyze the root of “scissors difference” and identify the causes of urban–rural disparities in developing countries.

6. Conclusions and Recommendations

In many developing countries, there is significant “scissors difference” between the income of rural and urban households, and therefore, it is required to explicitly consider such a difference in any modeling practice. This can be done by classifying households into urban and rural subcategories. However, there has not been any study to explicitly examine this issue and no guidance exists on how to classify households under the case that significant “scissors difference” presents.

In this study, travel survey data from the City of Wuhan collected in 2008 was used to study the differences between the rural and urban households in terms of their household size, income, space consumption, vehicle ownership rate and travel distance. The study results clearly show that “scissors differences” do exist between urban and rural households and such differences are also present among their subgroups. In particular, the following conclusions can be drawn from the study results:

1. Urban household subgroups classified based on either household income or size show statistically significant differences in the building area they consume, the highest household mobility and their travel distance.
2. Rural household subgroups classified based on either household income or size show statistically significant differences in the building area they consume and in the highest household mobility. However, no significant differences in travel distances between the rural household subgroups were observed.
3. In urban areas, household income appears to be the most important factor that determines the quantity of space consumed, when excluding the effect of “rigid demand” related to the household size. Conversely, in the rural areas, due to the low price of residential space/land, household size dominates the decision related to space consumption.
4. The rural household subgroups (classified by both household income and household size) show almost exactly the opposite results to their urban counterparts. The results of this study seem to suggest that the households in the rural areas have a very similar lifestyle (including their travel distance), regardless of their size and income, when compared to urban households in general.
5. The study results clearly show that, as the number of household categories considered within ILUTMs increase, the model is able to more accurately represent household behavior (with a smaller MSE), but at a cost of a higher demand for data. According to the results of the PECAS models tested, there are significant differences in family size, family income, space consumption and mobility choice when family types are divided into two categories (urban and rural), but there is no difference in travel behavior between them. However, when family income and family size are used as classification indicators and households are classified into 18 categories (9 types of urban households and 9 types of rural households), the travel distance as well as other attributes were found to be significantly different between urban household and rural subgroups.

The study results can be used to guide model design in order to explicitly consider the “scissors differences” that exist in household income, size, space consumption, vehicle ownership rates, and travel distance. For example, if a travel model is desired, then household groups with similar travel behavior can be merged together to simplify the modeling tasks. Conversely, when an integrated land use transport model is considered ideal, it would make more sense to separately model those household groups by considering their differences in several important aspects, such as space consumption, mobility choice and travel distance, all together.

While this study has looked at the “scissors difference” problem in detail, it still has the following limitations: only rural/urban, family size and family income were used for the household classification based on the traditional “Four-Step” travel demand modeling practice. Currently, this is constrained by the data used, which was extracted from the latest travel survey of the city of Wuhan. Furthermore, this data has not been updated since 2008. It is necessary that better classification schemes be developed in order to improve the overall planning practice and use integrated land-use transport models (ILUTMs). These models have the best performance when household behavior, in terms of space consumption, mobility choice, location choices of home and work and consumption of various goods and services, is explicitly considered. In addition, to improve the accuracy of the results of this study, other sources of data, such as mobile phone signal and login location data of Internet communication software/Apps, can be used for small-scale verification.

Author Contributions: Conceptualization, M.Z. and J.D.H.; Data curation, Q.T.; Formal analysis, X.M.; Investigation, Q.T.; Methodology, X.M.; Software, J.D.H.; Supervision, M.Z. and X.M.; Validation, M.Z. and X.M.; Visualization, M.Z. and X.M.; Writing—original draft, Q.T.; Writing—review and editing, M.Z. and X.M.

Funding: National Key Research and Development Program of China (2018YFB1600900).

Conflicts of Interest: The authors declare no conflicts of Interest.

References

1. Seton, F. Scissor crises, value-prices, and the movement of value-prices under technical change. *Struct Chang. Econ. Dyn.* **2000**, *11*, 13–24. [CrossRef]
2. Zhao, L.M.; Li, L.; Wu, Y.J. Research on the coupling coordination of a sea-Land system based on an integrated approach and new evaluation index system: A case study in Hainan Province, China. *Sustainability* **2017**, *9*, 859. [CrossRef]
3. Yue, D.; Xu, X.; Li, Z.; Hui, C.; Li, W.; Yang, H.; Ge, J. Spatiotemporal analysis of ecological footprint and biological capacity of Gansu, China 1991–2015: Down from the environmental cliff. *Ecol. Econ.* **2006**, *58*, 393–406. [CrossRef]
4. Jiang, H.B.; Wu, Y.F. Scissors difference problem under the market economy. *Shanghai Stat.* **2003**, *7*, 39–41. (In Chinese)
5. Wang, S.; Tan, S.; Yang, S.; Lin, Q.; Zhang, L. Urban-biased land development policy and the urban-rural income gap: Evidence from Hubei Province, China. *Land Use Policy* **2019**, *87*, 104066. [CrossRef]
6. Jones, G.A.; Corbridge, S. The continuing debate about urban bias: The thesis, its critics, its influence, and implications for poverty reduction strategies. *Prog. Dev. Stud.* **2010**, *1*, 1–18. [CrossRef]
7. Lipton, M. Urban bias: Of consequences, classes and causality. *J. Dev. Stud.* **1993**, *29*, 29. [CrossRef]
8. Kuznets, S. Economic growth and income inequality. *Am. Econ. Rev.* **1955**, *5*, 1–28.
9. Khor, N.; Pencavel, J. Income inequality, income mobility, and social welfare for urban and rural households of China and the United States. *Res. Labor Econ.* **2010**, *30*, 61–106.
10. Cao, Y.; Chen, X.; Ma, Y. Urbanization, urban-rural income gap and economic growth-based on Chinese provincial panel data empirical analysis. *Stat. Res.* **2010**, *3*, 29–36.
11. Fang, C. Rural urban income gap and critical point of institutional change. *Soc. Sci. China* **2003**, *5*, 16–25.
12. Eastwood, R.; Lipton, M. Rural and urban income inequality and poverty: Does convergence between sectors offset divergence within them? *Inequal. Growth Poverty Era Lib. Glob.* **2004**, *1*, 112–143.
13. Oyekale, A.S.; Adeoti, A.I.; Oyekale, T.O. Measurement and Sources of Income Inequality among Rural and Urban Households in Nigeria. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=985929 (accessed on 16 July 2019).
14. Chen, B.K.; Lin, J.Y. Development strategy, urbanization and the urban-rural income gap in China. *Soc. Sci. China* **2014**, *35*, 5–20.
15. Jin, H.; He, Y.X.; Zhao, C.H. Social security expenditure, demographic structural changes, and urban-rural income gap in China: Evidence from the provincial data. *Rev. Cercet. Interv. Soc.* **2018**, *62*, 254–277.
16. Su, C.W.; Liu, T.Y.; Chang, H.L. Is urbanization narrowing the urban-rural income gap? A cross-regional study of China. *Habitat Int.* **2015**, *48*, 79–86. [CrossRef]
17. Harun, M.; Mat, S.H.C.; Jalil, A.Z.A. Public expenditure expansion and inter-ethnic and rural-urban income disparity. *Procedia Econ. Financ.* **2012**, *1*, 296–303. [CrossRef]
18. Chotia, V.; Rao, N.V.M. Investigating the interlinkages between infrastructure development, poverty and rural-urban income inequality: Evidence from Brics Nations. *Stud. Econ. Financ.* **2017**, *34*, 466–484. [CrossRef]
19. Tiwari, A.K.; Shahbaz, M.; Islam, F. Does financial development increase rural-urban income inequality? *Int. J. Soc. Econ.* **2013**, *40*, 151–168. [CrossRef]
20. Sehrawat, M.; Giri, A.K. Panel data analysis of financial development, economic growth and rural-urban income inequality: Evidence from SAARC Countries. *Int. J. Soc. Econ.* **2016**, *43*, 998–1015. [CrossRef]
21. Jain, M.; Korzhenevych, A.; Hecht, R. Determinants of commuting patterns in a rural-urban megaregion of India. *Transp. Policy* **2018**, *68*, 98–106. [CrossRef]

22. Zhong, M.; Shan, R.; Du, D.; Lu, C.Y. A comparative analysis between traditional four-step and activity-based travel demand modeling: A case study of Tampa, Florida. *J. Transp. Plan. Technol.* **2015**, *38*, 517–533. [\[CrossRef\]](#)
23. Korzhenevych, A.; Jain, M. Area- and gender-based commuting differentials in India's largest urban-rural region. *Transp. Res. Part D Transp. Environ.* **2018**, *63*, 733–746. [\[CrossRef\]](#)
24. Pucher, J.; Renne, J.L. Rural mobility and mode choice: Evidence from the 2001 National Household Travel Survey. *Transportation* **2005**, *32*, 165–186. [\[CrossRef\]](#)
25. Hutchinson, J.; White, P.C.L.; Graham, H. Differences in the social patterning of active travel between urban and rural populations: Findings from a large UK household survey. *Int. J. Public Health* **2014**, *59*, 993–998. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Liddle, B. The systemic, long-run relation among gasoline demand, gasoline price, income, and vehicle ownership in OECD countries: Evidence from panel cointegration and causality modeling. *Transp. Res. Part D Transp. Environ.* **2012**, *17*, 327–331. [\[CrossRef\]](#)
27. Dargay, J.; Gately, D. Income's effect on car and vehicle ownership, worldwide: 1960–2015. *Transp. Res. A Policy Pract.* **1999**, *33*, 101–138. [\[CrossRef\]](#)
28. Paulley, N.; Balcomebe, R.; Mackett, R. The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transp. Policy* **2006**, *13*, 295–306. [\[CrossRef\]](#)
29. Choudhary, R.; Student, M.T.; Vasudevan, V. Study of vehicle ownership for urban and rural households in India. *J. Transp. Geogr.* **2017**, *58*, 52–58. [\[CrossRef\]](#)
30. Ulfarsson, G.F.; Steinbrenner, A.; Valsson, T. Urban household travel behavior in a time of economic crisis: Changes in trip making and transit importance. *J. Transp. Geogr.* **2015**, *49*, 68–75. [\[CrossRef\]](#)
31. Kottogoda, N.T.; Rosso, R. *Applied Statistics for Civil and Environmental Engineers*; Blackwell Publishing: Hoboken, NJ, USA, 2008; pp. 260–312.
32. Zhao, P.J.; Lu, B.; de Roo, G. Urban expansion and transportation: The impact of urban form on commuting patterns on the city fringe of Beijing. *Environ. Plan. A* **2010**, *42*, 2467–2486. [\[CrossRef\]](#)
33. Olsson, J. Changing Rural–urban Accessibility Linkages: Long Distance Mobility Behavior over Two Decades in the Philippines. In Proceedings of the 12th World Conference on Transport Research (WCTR), Lisbon, Portugal, 11–15 July 2010.
34. Andersson, M.; Lavesson, N.; Niedomysl, T. Rural to urban long-distance commuting in Sweden: Trends, characteristics and pathways. *J. Rural. Stud.* **2018**, *59*, 67–77. [\[CrossRef\]](#)
35. Zhao, P.J.; Lu, B.; de Roo, G. The impact of urban growth on commuting patterns in a restructuring city: Evidence from Beijing. *Pap. Reg. Sci.* **2011**, *90*, 735–754. [\[CrossRef\]](#)
36. Hu, L.Q.; Sun, T.S.; Wang, L.L. Evolving urban spatial structure and commuting patterns: A case study of Beijing, China. *Transp. Res. Part D Transp. Environ.* **2018**, *59*, 11–22. [\[CrossRef\]](#)
37. Eastwood, R.; Lipton, M. Pro-poor growth and pro-growth poverty reduction: Meaning, evidence, and policy implications. *Asian Dev. Rev.* **2000**, *18*, 22–58.
38. Connell, J. A new inequality? Privatisation, urban bias, migration and medical tourism. *Asia Pac. Viewp.* **2011**, *52*, 260–271. [\[CrossRef\]](#)
39. Nechyba, T.J.; Walsh, R.P. Urban sprawl. *J. Econ. Perspect.* **2004**, *18*, 177–200. [\[CrossRef\]](#)
40. Brueckner, J.K. Urban sprawl: Diagnosis and remedies. *Int. Reg. Sci. Rev.* **2000**, *23*, 160–171. [\[CrossRef\]](#)
41. Brueckner, J.K. Urban sprawl: Lessons from urban economics. In *Brookings-Wharton Papers on Urban Affairs*; Brookings Institution Press: Washington, DC, USA, 2001; pp. 65–89.
42. He, W.; Tang, B.L. Development mode of mid-small cities in northern Jiangsu Based on the growth pole theory. *Asian Agric. Res* **2012**, *4*, 63–76.