GWR-Based Analysis of Influencing Factors on Car Ownership in Zhejiang Province, China

Word count: 3048

Jan 16, 2021

Programme: MSc Smart Cities and Urban Analytics

Department: Bartlett School of Planning

GitHub Repository (Datasets): https://github.com/lizhiyuan913/GIS-assessment

Rpub (Full code and procedural walkthrough): https://rpubs.com/Zhiyuan_Li/GIS-assessment

Leaflet (Interactive map) : https://rpubs.com/Zhiyuan_Li/GIS-assessment-onlinemap

1. INTRODUCTION

With the rapid development of the automobile industry, China has become one of the largest countries in the automobile industry. By 2016, China came first in the ranking of the automobile consumed in the world, with 269.37 million civil cars, which is the second-largest automobile owner in the world (Yan, and Li Zheng). The Chinese automobile market is expected to reach 30 million in 2025. Motorized traffic is a major cause of air pollution, traffic congestion, rising temperatures, and energy shortage. While enjoying the convenience of cars, China is also suffering most from the negative effects of the rapid growth of car ownership.

In recent decades, both native and foreign scholars have studied car ownership. However, most of the studies use a traditional linear regression model and linear forecasting method to explore the influencing factors and predict the trend of car demand. But these studies ignore the spatial autocorrelation of the variables. Both the value and impact coefficients of influential factors like disposable income and urbanization may vary regionally and affect the fit of models. Therefore, this study applies both linear and geographically weighted regressions and chooses Zhejiang Province lying on the southeast coast of China as a sample. We firstly explore the relationship between car ownership and related factors with OLS, and then introduce the spatial factors, more reasonably and effectively simulate the relationship of the factors' impact on car ownership with the GWR.

The objective of this study is to help policymakers better understand the factors influencing car ownership and put forward sustainable policies to prepare the cities for the fast-growing number of cars, reduce car dependency, and best allocate resources to achieve healthy and stable development of the automobile industry.

2. LITERATURE REVIEW

A growing amount of literature has focused on the factors affecting car ownership in different regions and countries. Ingram and Liu (1999) observed a strong correlation between the growth of private car ownership and **road infrastructure**. Riley found out a larger **population** drives car ownership due to agglomeration effects(2002). Population density is another social-economic factor that will affect people's will to own a car. Holtzclaw and Clark (2016) illustrated that higher **population density** has undermined car ownership per capita. Large population density and low mileage per capita will easily cause traffic jams and public transportation will be a better choice in this situation. According to Li's research, **the distance from household to the city centre** also has an important influence on private car ownership (Jieping Li, 2010).

Economic factors play a significant role in car ownership and many related research models, **income** is seen as the only explanatory variable (e.g. Button et al. 1982; Dargay and Gately, 1999), which is the same with Lam and Tam's conclusion. They found that disposable income plays a major role in car ownership. Khan (1986) chose **GNP** per capita as an independent variable in his study and found the higher the GNP per capita is, the larger the car ownership is. According to Cao(2013), Vehicle ownership increase is at about twice the rate of **per capita income** growths. **Housing price** is another factor that cannot be ignored to a Chinese family. It will greatly affect residents' disposable income and in turn, affect their budget of purchasing a private car.

In the past, scholars usually used the traditional multiple linear regression to reflect the correlation between independent variables and dependent variable, assuming that there is homogeneity between variables. As a result, it can only reflect the overall average situation in the research area. However, spatial characteristics of the data have not been taken into consideration (Brunsdon C, Fotheringham A S, and Charlton M. 1996). Although lots of researchers did plenty of work to find out how the social-economic factors are related to car ownership, very few choose different counties of the same province as a sample. Therefore, this study considers **the population, per capita GDP, household price, population density, urbanization, per capita roadway mileage, and whether being the downtown or municipal district of prefecture-level cities**, and studies different county-level units in Zhejiang Province by using a combination of linear regression and geographically weighted regression (GWR).

3. METHODOLOGY

3.1 Data Collection and Study Area

i. Study Sample

Zhejiang is one of the smallest provinces of China, and one of the most affluent and densely populated. It has 13 prefecture-level municipalities containing 90 county-city units. We studied the study at the county level.

We choose Zhejiang for 2 reasons. Firstly, with the 4th strongest economy in China, it also has a 5th largest car ownership which is still growing rapidly. Its economy has been boosted by convenient road traffic and large car ownership, but it's also faced with problems such as traffic congestions and shortage of parking space.

Secondly, Zhejiang is a province composed of municipalities with different stages of urbanization and varying economy. It has Hangzhou, one of China's top megacities in densely populated coastal plains, and also municipalities ranked outside the top 200 in isolated mountain areas. Differences in economy and geography also lead to large differences in car ownership.

Figure 1. Zhejiang Province's location in China.

ii. Data and Variables

This study uses 2 kinds of data: one is data on the car ownership, as the dependent variable, and the other one is data on factors possibly influential on car ownership, as the independent variables. Based on the findings of the previous studies and the availability of relevant data, 9 variables that might impact car ownership are selected and classified into 3 types: variables representing **peoples purchasing power** , **the development of the built environment**, and **the convenience of road traffic**.

Car ownership is defined as the ratio of the total number of cars to 1000 people. Since our data sources do not include the data of income and the per capita GDP has an obvious positive correlation with economic status, we used per capita GDP to represent purchasing power. Built environment variables include household price(per square meter), population density, urbanization, population, and 2 dummy variables of whether being downtown or municipal districts of prefecture-level cities. Variables on road traffic convenience are mileage of roadway network per capita, and the ratio of annual bus passengers to population.

Most data used were collected and calculated from Zhejiang Statistical Yearbook 2019 and statistical yearbooks of prefecture-level cities in Zhejiang in 2019, which reflect the statistics of 2018. The only exception is the household price which is from Chinas housing price market platform run by Xitai corporation and China Real Estate Association. Missing data were replaced by interpolation of adjacent years.

Based on the availability of and accessibility to the statistical data, we merged some of the districts into the downtown of a prefecture-level city, after which the map we used has only 73 geographical units.

Table 1. Descriptive statistics of variables' characteristics.

Variable	Classification	Explanation	Min	Max	Mean
Car Ownership	——	Unit per 1000 people	7.05	45.06	23.93
GDP	**peoples purchasing power**	Per capita GDP, yuan	7.08	433.70	91.05
Population	**built environment**	Every 10,000 people	38767	161052	85964

Household price	**built environment**	Yuan per square meter	7868	32009	15492
Population Density	**built environment**	People per square kilometre	56.42	2260.59	609.33
Urbanization	**built environment**	Per cent	47.99	90.90	64.49
Downtown	**built environment**	Dummy variable, whether downtown of a prefecture-level city	0	1	0.17
District	**built environment**	Dummy variable, whether municipal district of a prefecture-level city	0	1	0.11
Roadway Mileage	**the convenience of road traffic**	Kilometre per capita	0.41	17.51	3.37
Highway Passenger	**the convenience of road traffic**	Ratio of highway passengers to population, every 10,000 person-times per 1000 people	1.07	52.17	14.68

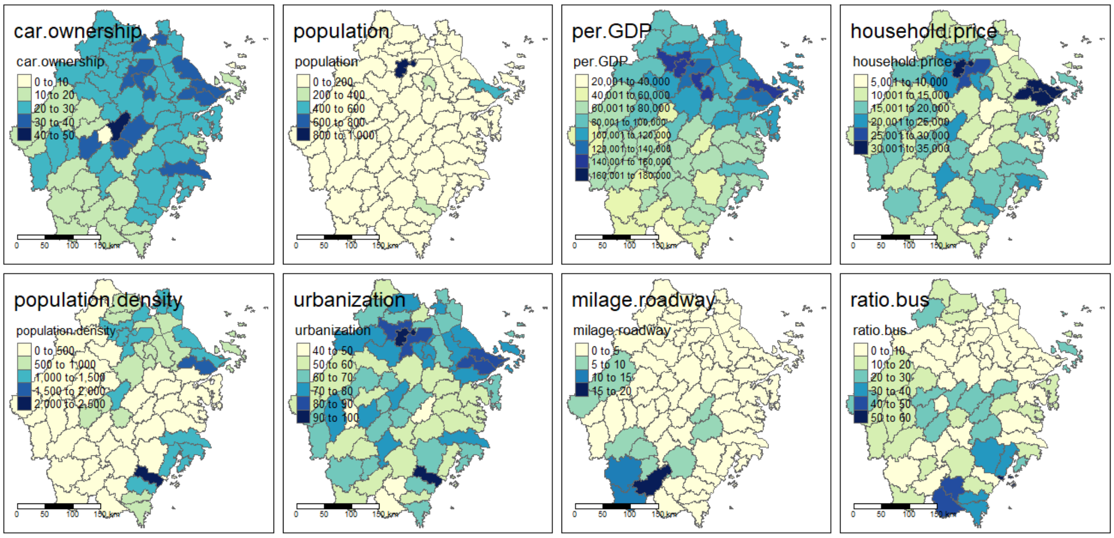


Figure 2. Spatial distribution of variables (dummy variables excluded).

3.2 Chosen Measure for Analysis

The analysis was twofold. Firstly, we employed the **Linear Regression** to check whether there is a basic linear relationship between car ownership and the independent variables, and ranked the importance of the independent variables. secondly, assuming the distribution of the regression model's residuals is geo-related, we performed the **Geographically Weighted Regression** and tried to improve the R^2 value. With the spatial distribution of the coefficients of the independent variables, we interpreted the pattern and tried to unveil the geographical mechanism of the variables' impact on car ownership.

i. Linear Regression

We have multiple model choices on hand, like classic OLS (Ordinary Least Squares), SLM (Spatial Lag Model), SEM (Spatial Error Model) and SDM (Spatial Durbin Model). Considering the spatial autocorrelation will be examined and treated in the GWR model, we ignored the effects of spatial autocorrelation in the first place by choosing the original OLS model. We calculated the standardized coefficients of the variables and recorded how much the adjusted R^2 drops when they were removed one at a time, to find out the importance of each variable.

ii. Geographically Weighted Regression

Firstly we plotted a map of the residuals from the OLS model to identify whether there is spatial autocorrelation of residuals and calculated the Moran's I for all independent and variables. With strong autocorrelation detected, we applied the basic GWR model and plotted the maps of the coefficients, which made it possible to interpret the spatial variation of the variables' influence on car ownership.

iii. Leaflet Interactive Mapping

Finally, we drew the local R^2, the few most impactable independent variables and their coefficients in GWR model into one leaflet online map as different layers and uploaded it online, for comparison, demonstration and further interpretation.

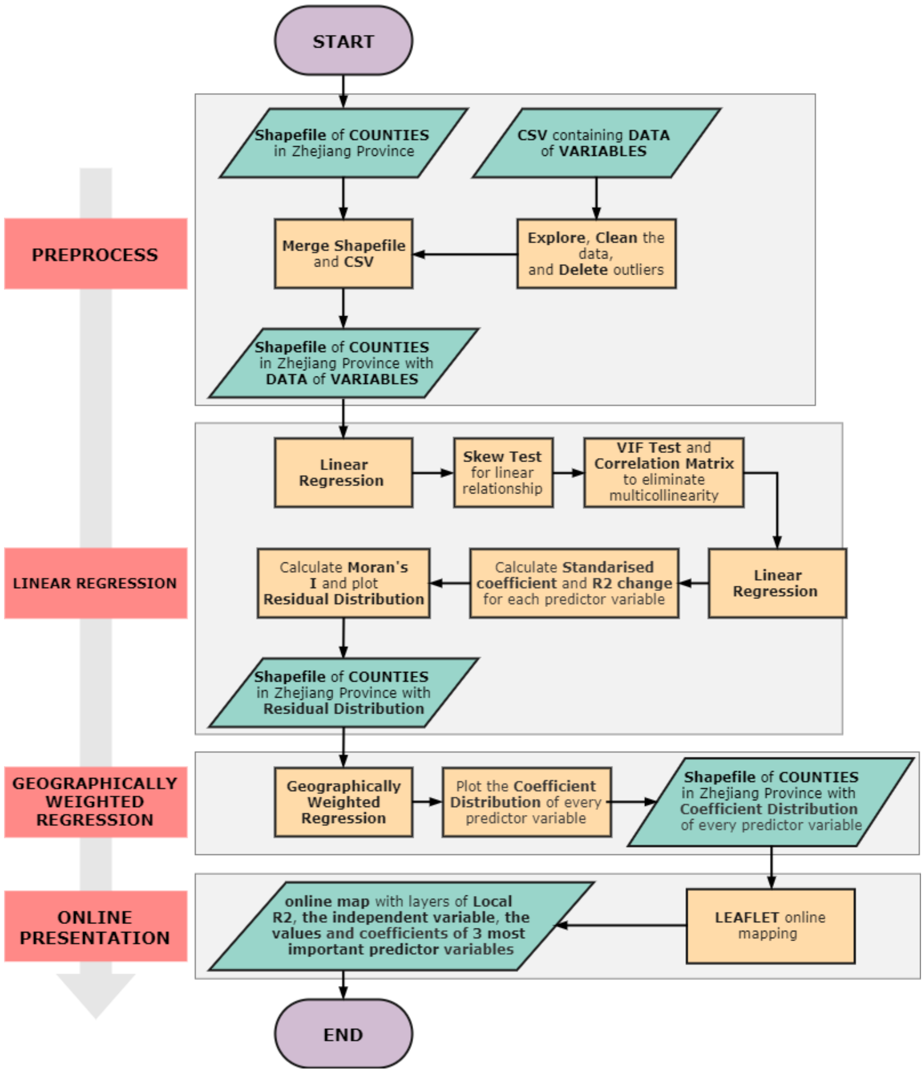


Figure 3. Workflow of data processing conducted in R

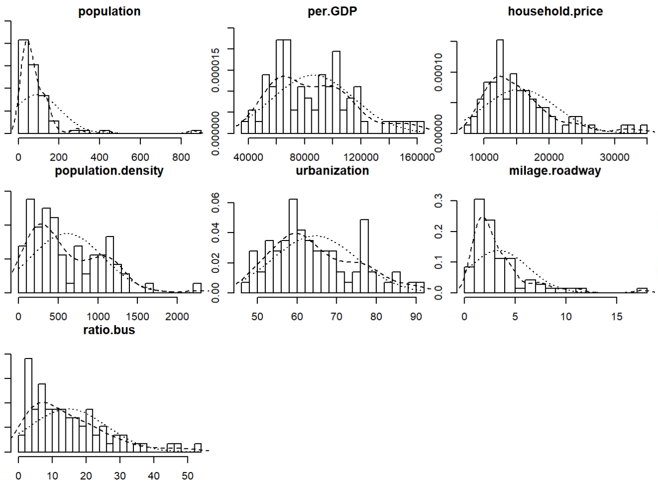
4. RESULTS

4.1 Summarizing the Data

We identified and removed one outlier in Jindong district which has multiple values of 0 (figure 2). In the skew test for a linear relationship, the population, household price, population density, roadway mileage, and highway passenger were found to be positively-skewed (table 2, figure 4), and treated with log transformation. With the VIFs of every independent variable under 5, the level of multicollinearity is at a low level (figure 5). The distribution of residuals is roughly normal, and homoscedasticity can be found (figure 6).

Table 2. VIFs and skewnesses of independent variables(except for the dummy variables)

Independent variables	VIF	Skewness before transformation	Skewness after transformation
GDP	2.428	0.6771	0.67714
Population	2.513	**4.5786**	0.09817
Household price	2.892	**1.4238**	0.55627
Population Density	2.790	**0.9688**	-0.42585
Urbanization	3.835	0.5689	0.56888
Roadway Mileage	2.132	**2.3596**	0.22024
Highway Passenger	1.517	**1.2082**	-0.40566
Downtown	2.195	——	——
District	1.691	——	——



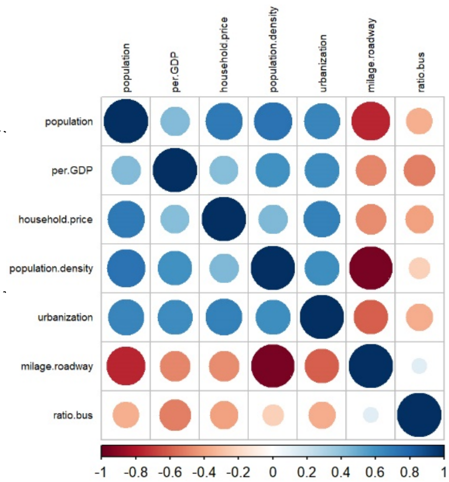
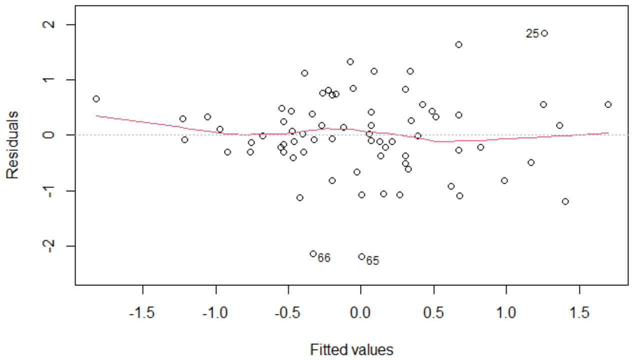


Figure 4. Histograms of independent variables(except for the dummy variables),

Figure 5. Correlation matrix of independent variables.

4.2 Regressions' Outcome

With an R^2 value of 0.4402, and P-value below 0.00002, the adjusted OLS model showed a solid relationship, but the independent values might only explain the car ownership to a limited extent in the global model. The distribution of residuals also suggests a strong spatial autocorrelation. After the GWR, the R^2 witnessed a huge improvement to 0.82, which means that the relationships between variables vary locally and can be explained better with the GWR model.



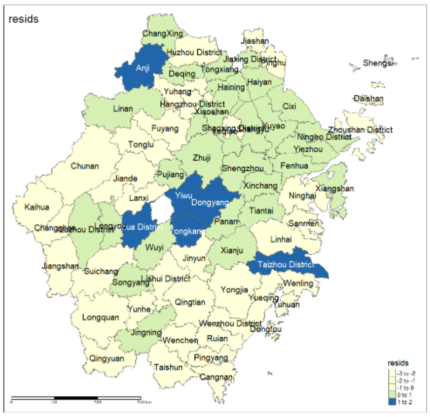


Figure 6. Residual-versus-fits plot of the OLS model and Figure 7. Distribution of residuals of the OLS model.

The distribution of local R^2 in the GWR model in figure 8. shows that the GWR model explains the relationship between dependent and independent variables better in the centre of Zhejiang, but a little bit loosely in northeast and southwest parts of Zhejiang.

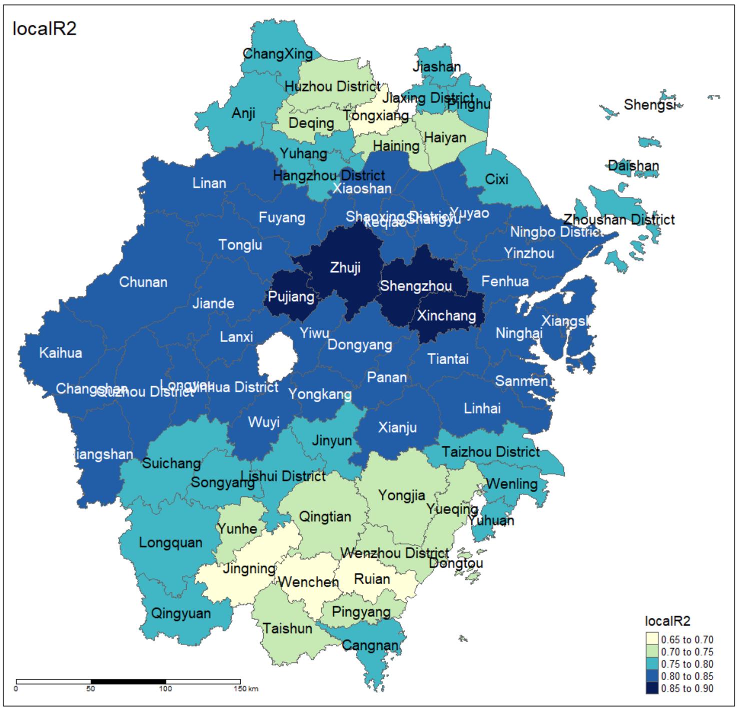


Figure 8. Distribution of local R^2 in the GWR model.

4.3 Relationships between Dependent and Independent Variables

The standardized coefficients from the linear regression show positive correlations between **GDP, household price, population density, urbanization** and car ownership, which means these factors possibly promote the growth of car ownership. P**opulation,**

roadway mileage, highway passenger, downtown, district, on the other hand, are found negatively impacting the car ownership.

Standardized coefficients suggest that **household price, roadway mileage**, and **urbanization** are the top 3 most influential factors on car ownership, followed by **population, GDP, downtown, district, population density**, and **highway passenger** according to priority. Change in the adjusted R^2 coincides with the standardized coefficients.

Table 3. Standardized coefficients, drop in adjusted R^2, and Moran's I of the variables.

Independent variables	**Standardized coefficient**	**Drop in adjusted R^2**	**Moran's I**
Car ownership	——	——	0.1964
GDP	-0.080991729	-0.007131	**0.6084**
Population	-0.122611225	-0.003434	0.1181
Household price	**0.446202613**	**0.06742**	0.2067
Population Density	0.048418098	-0.009229	**0.4483**
Urbanization	**0.273112952**	**0.01174**	0.1938
Roadway Mileage	**-0.310027311**	**0.04063**	**0.5811**
Highway Passenger	-0.069334484	-0.006605	**0.3554**
Downtown	-0.185991931	0.007585	-0.09234
District	-0.113198521	-0.001637	0.1211

As is shown in figure 9, of all the 9 independent variables, 8 showed a clear spatial autocorrelation pattern. The distribution of the coefficients for population, GDP, population density, urbanization, roadway mileage, and downtown has a single kernel pattern, and the distribution of the coefficients for household price and roadway mileage shows a gradual axial change. The coefficient for district has little change and no autocorrelation.

5. DISCUSSION

5.1 Linear Regression Result

Among factors positively correlated with car ownership, GDP, urbanization and population density are in line with common sense and our assumption. Compared with GDP per capita, representing purchasing power, and urbanization and population density as indicators of development of the built environment, in which an increase would reasonably lead to higher automobile consumption, household price's impact is rather vague. Higher household price may represent higher purchasing power and a more traffic-friendly built environment, but on the other hand, may also exhaust people's wallet and leave them with little money to buy cars as well. In China, people have a special preference for owning houses, and the positive relationship between household price and car ownership may be exclusive to more developed regions in China.

It's also reasonable that population, highway passenger, downtown, district have a negative correlation with car ownership. Higher population (not population density) may lead to more competition for license plates and more traffic-related problems. Citizens in Hangzhou may spend years in auctioning for a license plate which symbolizes the right to own a car, before really buying a car. Highway passenger represents public transportation which undermines the need for a private vehicle. Being downtown or a municipal district of a large city may indicate more traffic jams and low commuting efficiency. Similarly, a larger population may have a similar effect. Roadway mileage's negative relationship with car ownership may be intriguing, but also reasonable, as in mountainous southern Zhejiang there are many economically underdeveloped counties. Subject to the terrain, these counties have higher road mileage per capita but low car ownership.

5.2 GWR Coefficients' Distribution

The strong single-kernel and axial patterns detected in most coefficients in the GWR model show geography's influence on dependent and independent variables. Zhejiang is a mountainous province. Apart from the few large coastal plains lying on the northern and eastern edge, the vast majority of Zhejiang is full of hills, mountains and basins. Under the same economic conditions, the road mileage per capita in these regions is longer, and a higher degree of population concentration can be witnessed. So we can see a kernel pattern in coefficients for population, GDP, population density, highway mileage.

On the other hand, some large city clusters emerge on the northeast coastal plains of Zhejiang. The kernels in the distribution of urbanization and downtown coefficients are Hangzhou and Ningbo, the 2 largest and most developed cities, merging into a metropolitan area together with Shanghai. The positive influence of urbanization can be more intense. Yet, the downtown of Hangzhou, Ningbo and even Shaoxin (on the south of Hangzhou) is usually too crowded and expensive for driving or parking a car.

The most mountainous counties and the most developed coastal city centres are all faced with a shortage of land for construction, which results in the axial pattern of household price with high values on both northeast and southwest ends. The axial pattern of road mileage also coincides with the convenience of the road system of northeast plains.

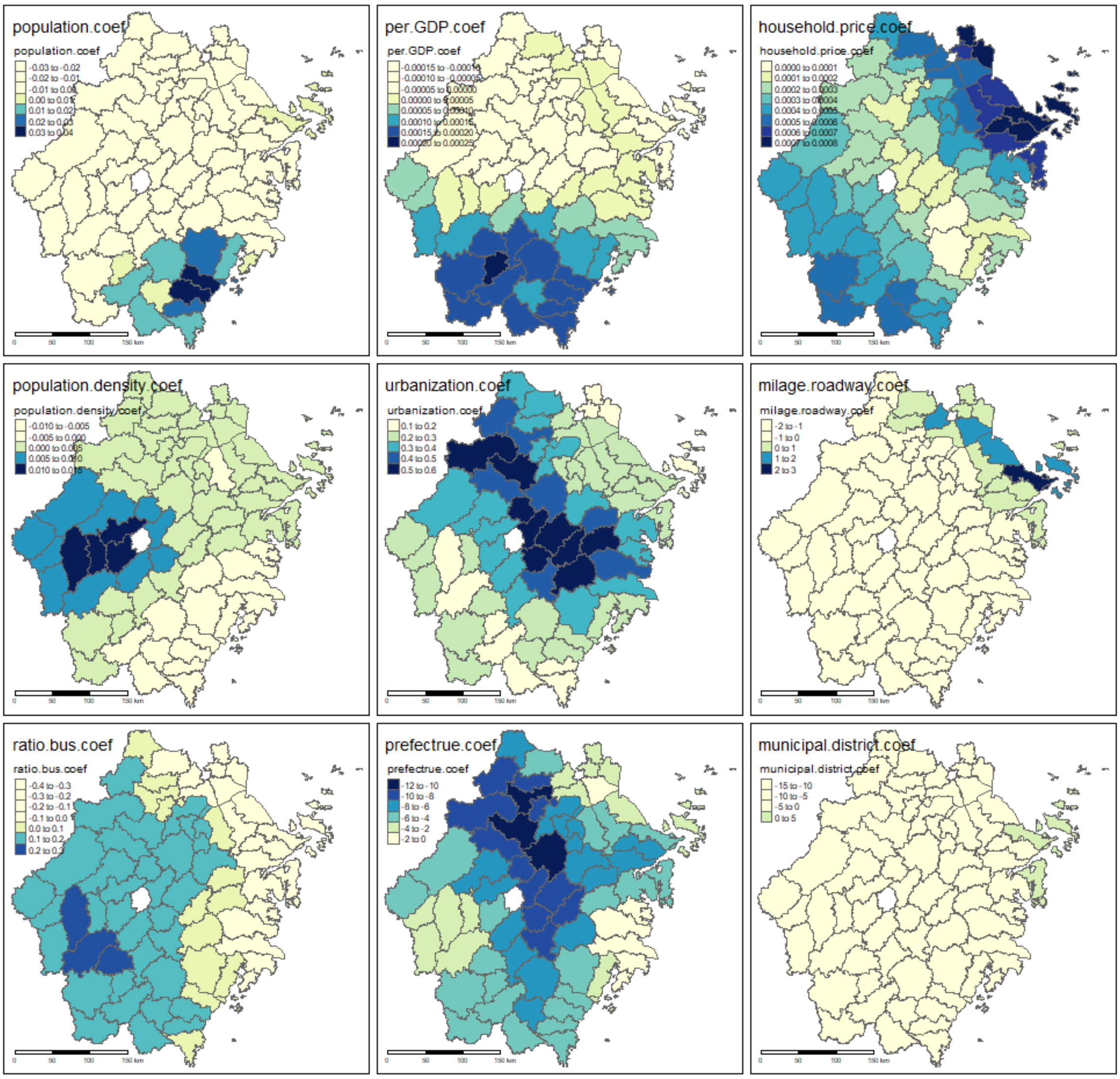


Figure 9. Distribution of coefficients of independent variables in the GWR model.

5.3 Recommendations

Firstly, apart from urbanization, the strong relationship between household price and car ownership is meaningful in that we can predict faster growth in car ownership if the household price soars.

Secondly, downtown suffering most from traffic-led problems may possess fewer cars than the surrounding districts. So we should indeed provide the surrounding districts with more parking space than suspected.

Our findings also suggest that in regions with poor traffic conditions, the provision of public transportation is more effective and should be promoted more to restrain explosive automobile consumption.

5.4 Limitations

This study has several limitations. With limited statistical data sources, some important variables were ignored, compromised or replaced. For instance, the average disposable income was replaced by GDP per capita, and the population density was not population divided by the area of the built environment, but by the area of the county as a whole. The influence of other traffic facilities such as gas stations, train stations, or undergrounds was ignored.

With an R^2 of 0.82, the GWR model can be further optimized and more relevant factors taken into consideration, especially policy factors like the auction mechanism for license plates. As the GWR model explains the relationship between dependent and independent variables better in the centre of Zhejiang, but a little bit loosely in northeast and southwest parts of Zhejiang, the distribution of local R^2 in the GWR model also suggests that there might be some other factors influencing the mountainous and coastal parts of Zhejiang exclusively.

Besides, this study only performed a regression within Zhejiang, but no comparison with other provinces or snapshots of another time slot. Thus we displayed no spatialtemporal pattern of car ownership.

6. CONCLUSION

This study set out to examine the influence of different factors on car ownership in Zhejiang with OLS and GWR models. The OLS model shows that GDP, household price, population density, urbanization positively impact car ownership. Population, roadway mileage, highway passenger, downtown, district, negatively impact car ownership. Household price, roadway mileage, and urbanization are the top 3 most influential factors on car ownership. The GWR model shows single kernel and axial patterns in the distribution of the coefficients which can be explained by the geo-economic characteristics of Zhejiang.

The results show a strong positive relationship between household price and car ownership, which we can not explain thoroughly. Further research could be done to examine this relationship, testifying whether it is exclusive to Zhejiang, or generic in China, and interpreting the deep mechanism of this relationship.

7. REFERENCE

Brunsdon, C., Fotheringham, A.G., and Charlton, M. (1996) “Geographically weighted regression: A method for exploring spatial nonstationary”, *Geographical Analysis*, 28, 4, 281-298.

Button, K. J., Pearman, A. D., and Fowkes, A. S. (1982) *Car Ownership Modelling and Forecasting*, Aldershot: Gower Publishing.

Cao, X. and Huang, X. (2013) “City-level determinants of private car ownership in China”, *Asian Geographer*, 30, 1, 37-53.

Clark, S. (2007). “Estimating Local Car Ownership Models”, *Journal of Transport Geography*, 15, 3, 184-197.

Dargay, J., and Gately, D. (1999) “Incomes Effect on Car and Vehicle Ownership, Worldwide: 1960-2015”, *Transportation Research Part A: Policy and Practice*, 33, 2, 101-138.

Holtzclaw, J., Clear, R., Dittmar, H., Goldstein, D., and P. Haas. (2002) “Location Efficiency: Neighborhood and Socioeconomic Characteristics Determine Auto Ownership and Use—Studies in Chicago, Los Angeles, and San Francisco”, *Transportation Planning and Technology*, 25, 1, 1-27.

Li, J., Walker, J.L., Srinivasan, S., and Anderson, W.P. (2010) “Modeling Private Car Ownership in China: Investigation of Urban Form Impact Across Megacities”, *Transportation Research Record*, 2193, 1, 76-84.

Khan, M. A., and Willumsen L. G.(1986) “Modelling Car Ownership and Use in Developing Countries”, *Traffic Engineering and Control*, 27, 554-560.

Lam, W., and Tam., M. (2002) “Reliability of Territory-wide Car Ownership Estimates in Hong Kong”, *Journal of Transport Geography*, 10, 1, 51-61.

Liu, Z. and Ingram, *G. K.* (1999) *Determinants of Motorization and Road Provision*, Washington: The World Bank.

Riley, K. (2002) “Motor Vehicles in China: The Impact of Demographic and Economic Changes”, *Population and Environment*, 23, 5, 479-493.