A NOVEL USE OF LATENT CLASS MODELLING TO UNDERSTAND THE HETEROGENEITY OF URBAN LAND USE EFFICIENCY IN CHINA

T. Y. Wang*, L. Wan

Department of Land Economy, University of Cambridge, UK – (tw531, lw423)@cam.ac.uk

KEY WORDS: Statistical modelling, Urban modelling, Latent class analysis, Data-driven, Urban development, Land use.

ABSTRACT:

Urban and regional development disparity is a salient issue for both developed and developing countries, and increasing development control brings the efficiency of urban land use to the forefront. Understanding the heterogeneity of urban land-use efficiency is therefore important for balancing growth-oriented land supply and sustainable development. However, two research gaps can be identified from existing studies. Firstly, the disparities of urban land-use efficiency are usually multi-faceted, the measurement of which is thus a major challenge. Secondly, a systematic investigation of land-use efficiency and the associated geographical heterogeneity using multi-dimensional, longitudinal measurements seems limited. Therefore, this paper presents a novel latent class modelling method for understanding the pattern of such heterogeneity using purposely constructed, repeated cross-sectional measurements, including local government revenue density and employment density. Using data of 272 prefecture-level Chinese cities, our model first identifies the level of urban land-use efficiency can be split into three distinct cities groups based on the performance in 2017. To further investigate the change of the efficiency over the years (2012-2017), two separate models have been applied and a total of five latent groups among Chinese cities are identified with distinct development patterns. This modelling approach represents a viable method for assessing the land-use efficacity, both statically and over time, across a large number of cities of varying development stages. Policy implications are drawn.

1. INTRODUCTION

Despite the unprecedented rate of urbanisation in China for the past decades, significant economic development disparities have been witnessed across Chinese cities and current trajectory does not suggest a narrowing trend. Figure 1a shows the change of built-up area in China from 2012 to 2017 at provincial level. Figure 1b shows the change of local government revenue for the same time period, where the gap between more prosperous provinces and the rest of the country seems widening. We choose local government revenue as the measurement over Gross Domestic Product (GDP) because the former may provide a more realistic and robust measure of the economic performance of the city. Higher level and growth of local government revenue suggests a stronger economic capacity of the city in terms of financing quality public infrastructure and services.

Monitoring and understanding the urban development pattern is crucial for informing effective policy interventions. In recent years, given the increasing trend of urbanisation process but the total amount of land resources remains constant, the efficiency of urban land use is a prerequisite for sustainable development. Particularly, China faces a huge challenge that the area of urban construction land has increased by 352.1% in the past 30 years, from 12,907.9 km^2 in 1991 to 58,355.3 km^2 in 2020, resulting in an uncoordinated urban and regional development and increasing environmental degradation issues. Therefore, to achieve the goal of China that maintaining at least 120 million ha of farmland

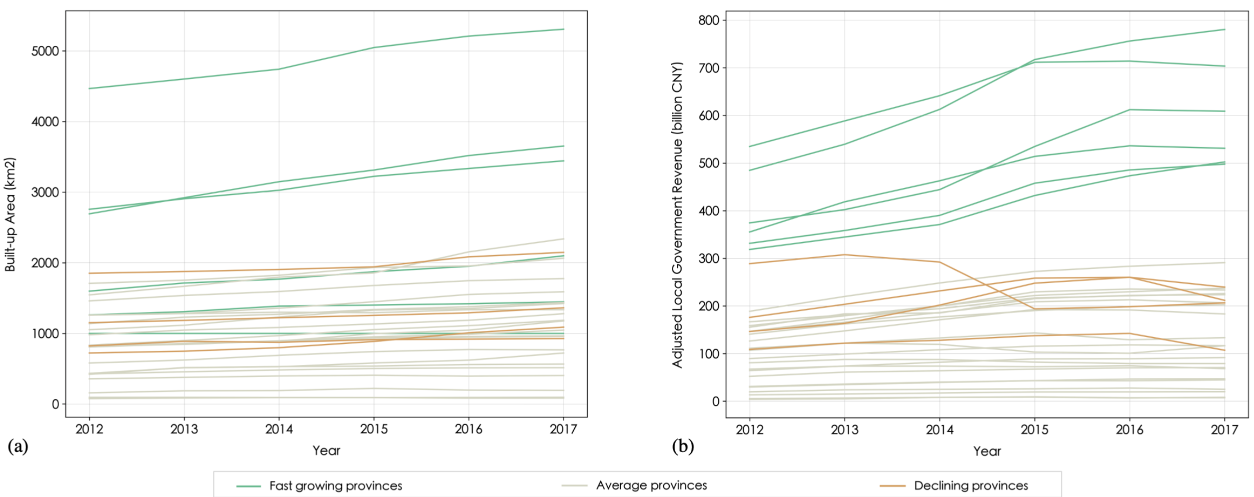


Figure 1. (a) Change of built-up area in China from 2012-2017; (b) Change of local government revenue in China from 2012-2017 (constant price based on year 2012 level; data source: *China City Statistical Yearbook* series)

* Corresponding author

nationwide, it is important for city planners and policy makers to monitor and understand their current efficiency performance of urban land use, in addition to the absolute size of urban land area, and facilitate future planning strategy making.

A multitude of existing studies have examined the urban land-use efficiency from various perspectives, while two major research gaps have been identified: 1) the disparities of land-use efficiency are usually multi-faceted, the measurement of which is thus a major challenge; and 2) a systematic investigation of urban land-use efficiency and the associated geographical heterogeneity using both multi-dimensional, longitudinal measurements seems limited.

To address the aforementioned research gaps, this research aims to provide a novel application of latent class analysis (LCA) for identifying the heterogeneity of urban land-use efficiency in Chinese cities. Specifically, LCA is used to identify latent groups of cities based on local government revenue density and employment density. The selection of the two indicators is for the following considerations. Firstly, density-based measurement can help to control for biases incurred by the different sizes of cities, which is a significant issue in the Chinese context. Secondly, both indicators are based on per km^2 economic output, which thus explicitly address the land as a key production input. In terms of the treatment of time in the model, LCA will be conducted for both static, cross-sectional data (2017) and time-series data (2012-2017). The static analysis aims to provide a ranking of cities in terms of the land-use efficiency; whilst the dynamic analysis focuses on identifying generic patterns based on the change of land-use efficiency over time. The combination of these two analyses appears novel among studies of similar topics and is expected to provide a viable analytical method for examining the heterogeneity across a large number of cities based on relevant measurements. This modelling approach could be used by local governments and institutions to identify the latent cities groups with complex development patterns, address those development needs, and facilitate future policy and planning strategy making.

The reminder of this paper proceeds as follows. Section 2 reviews existing relevant studies. Section 3 summarises data and model variables. Section 4 explains the details of modelling methods. Section 5 presents results and discussion. Section 6 provides the conclusion.

2. LITERATURE REVIEW

2.1 Multi-dimensional measurements of urban land-use efficiency

Existing studies tend to examine the efficiency of urban land use mainly through two types of methods. First, conduct density-based measurement to indicate land-use efficiency, which is the value created per unit of urban land area such as economic output per unit of urban land (Wu et al., 2017; Zhang et al., 2019) and population per unit of urban land (Chakraborty et al., 2022; Dong et al., 2019). Given the notion of land-use efficiency is usually multi-faceted, the measurement of which is a key to understand the development underlying patterns and draw effective policy implication. However, there has yet been a consensus on the measurement of urban land-use efficiency. Another group of scholars tend to describe the process of urban land use as an input-output system and assess land-use efficiency via various techniques such as Data Envelopment Analysis (Chen et al., 2016; Zhu et al., 2019), Stochastic Frontier Approach (Dong et al., 2020; Jin et al., 2018), Slack-based Measurement (Lu et al., 2018; Song et al., 2022), etc., where they evaluated the land-use efficiency from the index of material input and economic and environmental output perspective.

Although studies have applied various measurements to assess land-use efficiency, little study has examined the efficiency from the scope of local government revenue and employment yet. Particularly in China, given the land finance scheme that land-related revenue (including both land conveyance fee and its subsequent tax related income) serves as a predominant source of local government fiscal account, land has often been used as a tool to stimulate local economic output in short term. However, compared to one-off income from land conveyance fee, whether land can be realised to bring more subsequent employment growth and related tax income stream would be rather more important for long-term sustainable development. Recently, Cai (2017) and Liu (2019) have started to examine the impact of land on land-based tax revenue, which shifts the focus from overall economic total output to the pocket of local government. Nevertheless, to provide more comparable and robust evaluation of urban land-use efficiency over the years, further investigation targeting the dimension of both local government revenue and employment with the respect of unit land area would be required.

2.2 Measuring cross-city heterogeneity in land-use efficiency

In order to capture both multi-dimensional and longitudinal measurements, a proper classification with mapping is a powerful tool to help understand the non-linear heterogeneity of cities in terms of features, functionality, and problems. Numerous studies have tried to extract the underlying information and patterns by applying different techniques and mining the data. For example, the official UK Office for national Statistics (ONS) conducted K-means cluster analysis for residential-based and workplace-based area classification in the UK (ONS, 2018). Also, by using remote sensing data, Rahman et al., (2019) applied hierarchical cluster analysis and classified 331 cities of Bangladesh based on five selected spatial features.

In recent years, latent class analysis (LCA) is a well-established methodology to identify unobserved class membership within the sample objects and it has been widely applied in many research fields such as social science (Fox and Escue, 2021), psychology (Zhou et al., 2019), and biomedical area (Lee et al., 2021). It postulates that membership in the unobserved classes could explain the different patterns across cases, where the term of *pattern* is measured by several observed indicators. In other words, based on the selected indicator variables, LCA could identify which cities can be grouped together and cities in the same class often share similar characteristics. Compared with other cluster methods such as K-means and hierarchical cluster analysis, LCA, as a statistical model, provides more formal criteria of goodness-of-fit measurements such as Akaike information criterion and Bayesian information criterion, to suggest the number of classes (Rafiq and McNally, 2021). In addition, another key advantage of LCA is that it can be easily extended to incorporate covariates besides indicators themselves, which can help to accommodate model flexibility, distinguish endogenous variables, and facilitate the description of model results and underlying structures. Technical details can refer to Huang & Bandeen-Roche (2004) and Vermunt (2010). Moreover, the results of LCA can be easily interpreted and extended to other applications such as structural equation modelling (Jahanshahi and Jin, 2020).

Therefore, many urban studies have applied LCA to extract underlying patterns in recent research such as urban development (Chikaraishi et al., 2015; Wan and An, 2017), travel patterns (Haas et al., 2018; Rafiq and McNally, 2021), travel modes (Zhou et al., 2020), and location choice (Jiang et al., 2020; Liao et al., 2015), while it seems that LCA has not been applied to address the cross-city heterogeneity of land-use efficiency yet. Various aspects and extensions have then been developed under the framework of the standard LCA, particularly in terms of the selection of observed variables and latent variables, the type of the variables, and the temporal dimensions.

3. DATA

Data for this research are compiled from *China City Statistical Yearbook Series* and *China Urban Construction Statistical Yearbook Series*. The compiled data set covers 272 prefecture-level cities in China over the period from 2012 to 2017. Missing values are filled by linear interpolation process. The correction of a small number of inconsistency errors refers to our early paper, see Wan et al., (2021). In addition, all monetary variables have been adjusted by the constant price based on year 2012 through price deflators.

To assess urban land-use efficiency, local government revenue density ($revDen$) and local employment density ($empDen$) have been calculated respectively by: 1) local government revenue divided by built-up area, and 2) total number of employees in secondary and tertiary industries divided by built-up area. Notably, local government revenue particularly refers to local general public budgetary revenue in this context, which is closely related to the land development through non-transaction-based tax revenue stream such as corporation and personal tax. For employment variable, we target the number of employees working in the secondary and tertiary industries, since urban construction land is mainly for non-agricultural purposes and this approach is in line with prior relevant urban studies (Gao et al., 2020; Liu et al., 2018). To capture the growth pattern over the years and remove the scaling issue, we calculate the index value of both local government revenue density and employment density based on their value in year 2012. The prefix 'ind' in the variable names refers to the index value. In terms of the covariates, two covariates variables take the stock level in the beginning year of 2012, notably by the employment in secondary and tertiary industries ($Emp2012$) and GDP per capita ($pcGDP2012$). Table 1 lists the descriptive statistics of all variables in this study.

Variable (unit)	Mean	Std	Min	Max
empDen2017 (10,000 people/km^2)	0.920	0.505	0.082	4.227
revDen2017 (billion CNY/km^2)	0.136	0.079	0.010	0.610
indempDen2013	1.048	0.157	0.395	1.620
indempDen2014	1.062	0.207	0.390	1.748
indempDen2015	1.047	0.255	0.367	1.897
indempDen2016	1.035	0.267	0.355	2.235
indempDen2017	1.035	0.295	0.329	2.165
indrevDen2013	1.079	0.122	0.594	1.672
indrevDen2014	1.131	0.196	0.564	2.113
indrevDen2015	1.155	0.321	0.347	2.891
indrevDen2016	1.128	0.343	0.238	3.021
indrevDen2017	1.077	0.331	0.328	2.655
Emp2012 (10,000 people)	107.943	149.635	9.810	1190.948
pcGDP2012 (100,000 CNY/person)	0.541	0.339	0.035	2.159

Table 1. Descriptive statistics table

4. METHODS

In order to identify the distinct city groups for both static and dynamic manners of urban land-use efficiency across cities, two types of LCA models will be applied accordingly for 272 prefecture-level cities in China. Model 1 aims to present a static assessment to show the performance level of land-use efficiency for each city, based on the cross-sectional data in 2017. Model 2 aims to further investigate the change of land-use efficiency during the sample period and identify the generic growth patterns of all cities, based on the panel data between year 2012 and year 2017. Figure 2. summarises the structure of the overall analytical procedure in this research.

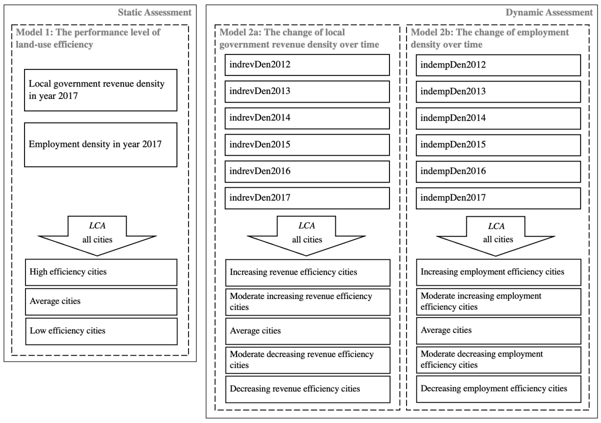


Figure 2. Overall analytical framework

The goal of these three LCA models is to identify the distinct city groups by maximising cross-group variance and minimising within-group variance. More specifically, LCA assigns each city with a probability of being in each subclass based on maximum likelihood estimation, and each city will be assigned to the class that they have the highest probability of belonging. Through the modelling, we defined the research questions as: (1) how many latent classes are underlying among Chinese cities by conducting a substantively and statistically check; (2) which city belongs to which latent class; (3) what are the patterns of each class and how these classes differ from each other. In terms of determining the number of latent classes, we follow the method by lowest Akaike information criterion (AIC) and Bayesian information criterion (BIC) (Nylund et al., 2007). All the LCA models were performed through Stata 17.0.

4.1 Model 1: The performance level of land-use efficiency

Model 1 is designed to provide a static assessment in terms of how cities performed in terms of their land-use efficiency based on multiple dimensions, which will provide a basis for the further investigation. As aforementioned, the design of LCA model includes both indicator variables and covariates. Specifically, indicator variables are used to define the latent classes, and covariates are used to predict the probability of the individual belonging to a latent class through the estimation of a structural model. Specifically, to differentiate cities by simultaneously considering both land-use efficiency and economic structure, we particularly focus on two key aspects in this study: 1) local government revenue density, and 2) local employment density. The former aims to capture the efficiency of local government financial capability, which shows the pocket money that local government can invest for building public physical and social infrastructure services; the latter aims to reflect the outcome in terms of stimulating employment opportunities. Increasing land expansion without increasing local government revenue or employment opportunity will be regarded as low efficiency of urban land use in this context. Additionally, the number of employments in secondary and tertiary industries and GDP per capita are included in the model as covariates, since these two variables could represent the size of the local economy and the level of economic welfare respectively, which tends to be relatively stable and in line with the approach in existing studies (Li and Gibson, 2015; Yan et al., 2020).

4.2 Model 2: The change of land-use efficiency over time

In order to further investigate the dynamic pattern of land-use efficiency, that how cities performed over the time, Model 2 is designed to classify cities based on index values over the years between 2012 and 2017. Similar to Model 1, we particularly address two aspects in this study: 1) the change of local government revenue density (Model 2a), and 2) the change of local employment density (Model 2b). Two covariates (the number of employees in year 2012 and GDP per capita in year 2012) are included to control the varying size of the local economy and local economic welfare, respectively. In order to better capture the growth pattern and rule out the influence of the initial level, normalised index value have been calculated for both two explanatory variables with setting the value in 2012 equals to 1. The results of latent classes between Model 2a and Model 2b will be compared, to investigate the potential correlation between two class memberships. Also, two scatter plots will be constructed to link the change of efficiency with the change of built-up area. Overall, model 2 aims to address the key question that, how cities have performed over time.

5. RESULTS AND DISCUSSION

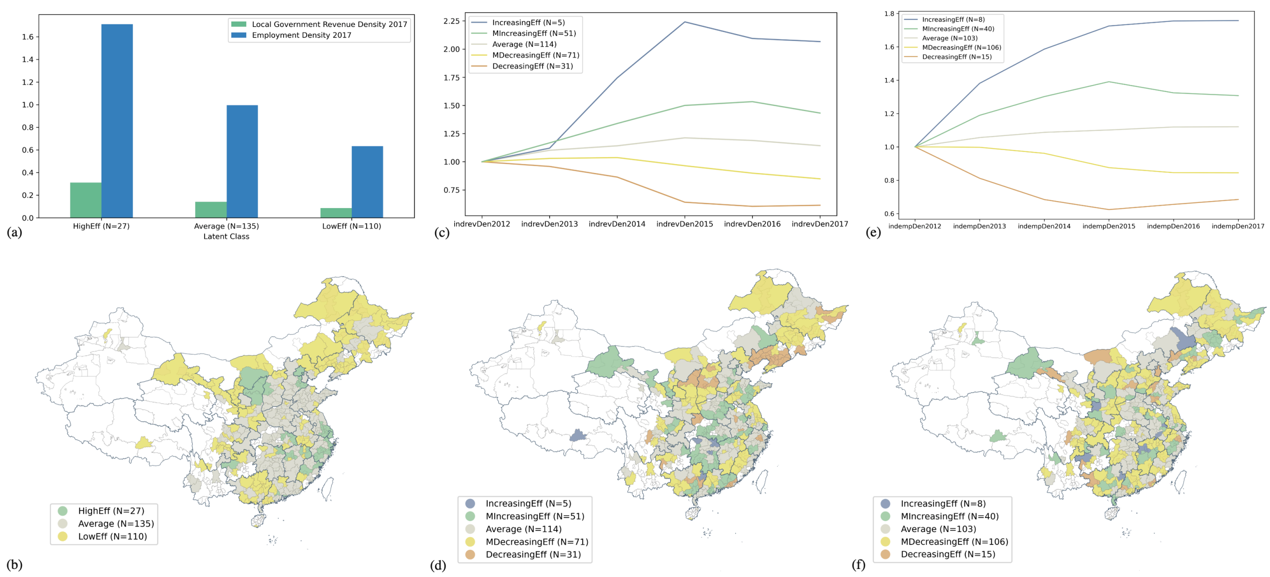
5.1 Model 1: The performance level of land-use efficiency

Table 2 lists the fit statistics and the number of cities in each class for all three models. In model 1, the lower change of both AIC and BIC imply that there is a better model performance from 2 classes to 4 classes. Given the 3-class model had relatively lower AIC and BIC values, and more importantly, the associated classes are logically interpretable, three-class model was selected in this research for model 1.

Figure 3 plots all the LCA results for three models. Specifically, Figure 3a shows the mean value of each explanatory variable (local government revenue density in 2017 and employment density in 2017) for each latent class in model 1. The result indicates that there is a significant heterogeneity of land-use efficiency exists among cities, which is in line with the findings of existing studies (Lu et al., 2018; Liu, 2018). 27 cities are classified as high-efficiency (*HighEff*) cities, featuring the highest level of both local government revenue density and employment density in 2017. By contrast, 110 cities are classified as low efficiency (*LowEff*) cities, featuring relatively lower level of local government revenue density and employment density in 2017. The other 135 cities are classified as *Average* cities, showing a relatively compatible performance level of both local government revenue and employment density. Additionally, Figure 3b provides a mapping of the city grouping. It shows that high efficiency cities are not only concentrating on east coastal side but also on few inland regions, which seems to oppose the general argument and stereotype that cities in eastern regions often have higher land-use efficiency than central regions (Lu et al., 2018). Our model based on city-level performance highlights that cities within the same region do not have same efficiency level of urban land use. Moreover, the bold blue line draws provincial boundary, which highlights cities within the same province also have distinct performance of land-use efficiency. Thus, it would be rather important to investigate city level to draw practical planning policy implication, particularly cities in the central region and lower-tier administrative hierarchy can also have the potential to achieve high land-use efficiency. The list of cities in each class is summarised in Appendix Table A.

5.2 Model 2: The change of land-use efficiency over time

5.2.1 The change of local government revenue density (Model 2a) and the change of employment density (Model 2b)



Model 1. The performance level of land-use efficiency (Yr 2017)

Model 2a. The change of local government revenue density (Yr 2012-2017)

Model 2b. The change of employment density (Yr 2012-2017)

Figure 3. The LCA results of Model 1, Model 2a, Model 2b

Model 1. Number of Classes in the LCA model				
	1-class model	2-class model	3-class model	4-class model
Class A	272	241	135	127
Class B		31	27	25
Class C			110	117
Class D				3
Total	272	272	272	272
AIC		-349.63	-384.42	-462.55
BIC		-317.18	-333.94	-394.04

Model 2a. Number of Classes in the LCA model						
	1-class model	2-class model	3-class model	4-class model	5-class model	6-class model
Class A	272	158	161	148	114	116
Class B		114	50	53	51	51
Class C			61	66	71	72
Class D				5	5	5
Class E					31	17
Class F						11
Class G						
Total	272	272	272	272	272	272
AIC			-649.42	-839.38	-911.09	-956.66
BIC			-555.67	-716.78	-759.64	-776.37

Model 2b. Number of Classes in the LCA model						
	1-class model	2-class model	3-class model	4-class model	5-class model	6-class model
Class A	272	188	125	135	106	111
Class B		84	104	97	103	91
Class C			43	21	40	25
Class D				19	15	17
Class E					8	8
Class F						20
Class G						
Total	272	272	272	272	272	272
AIC			-739.49	-837.10	-935.00	-991.28
BIC			-645.74	-714.51	-783.56	-810.99

Table 2. Fit Statistics of Model 1, Model 2a, and Model 2b

As shown in Table 2, with the increasing number of classes, both AIC and BIC values decreased for Model 2a and Model 2b. The five-class model was selected for both Model 2a and Model 2b, since five class model has relatively lower values of AIC and BIC and a relatively flattening decline pattern was observed after five-class model.

In terms of the change of local government revenue density (Model 2a), Figure 3c shows the mean value of each explanatory variable for each latent class. Specifically, 5 cities (Yueyang, Zhangjiajie, Lhasa, Xiangtan, Hezhou) are classified as increasing efficiency (*IncreasingEff*) cities, featuring a greatest increase of local government revenue density from year 2012 to year 2017. In addition, 51 cities are classified as moderate increasing efficiency (*MIncreasingEff*) cities, that their local government revenue density performance has slightly increased over the years. By contrast, 31 cities are decreasing efficiency (*DecreasingEff*) cities, featuring a decreasing of local government revenue density from 2012 to 2017. 71 cities are classified as moderate decreasing efficiency (*MdecreasingEff*) cities, as they received slightly shrinking pattern of local government revenue density, Also, 114 cities are *Average* cities, that their local government revenue density did not have much change over the years. In terms of the change of local employment density (Model 2b), Figure 3e shows that our LCA model identifies 8 increasing efficiency (*IncreasingEff*) cities, 40 moderate increasing efficiency (*MIncreasingEff*) cities, 103 *Average* cities, 106 moderate decreasing efficiency (*MDecreasingEff*) cities, and 15 decreasing efficiency (*DecreasingEff)* cities.

The mapping of the city grouping for Model 2a and Model 2b suggests that there are several lower-tier cities which also have achieved sustainable improvement in urban land-use efficiency. By comparing Figure 3d and Figure 3f, it shows that there is a distinct difference in terms of the grouping result between the change of local government revenue density and the change of employment density. In other words, cities that experienced an improvement of employment density over the years may not necessarily receive increase of local government revenue, which highlights the importance of simultaneously examine multiple dimensions to assess the urban land-use efficiency.

5.2.2 Class membership matrix between Model 2a and Model 2b

In order to identify the common patterns between the change of local government revenue density and the change of employment density from two sub-models over the years, we summary a matrix for the results of latent class membership between Model 2a and Model 2b. Table 3 shows the number of cities in each class. The list of cities is summarised in Appendix Table B.

Model 2a \ 2b	Increasing Eff (N=8)	MIncreasing Eff (N=40)	Average (N=103)	MDecreasing Eff (N=106)	Decreasing Eff (N=15)
Increasing Eff (N=5)	1	1	2	1	0
MIncreasing Eff (N=51)	3	8	22	18	0
Average (N=114)	3	17	51	40	3
MDecreasing Eff (N=71)	1	12	22	32	4
Decreasing Eff (N=31)	0	2	6	15	8

Table 3. Count matrix of latent class membership results between Model 2a and Model 2b
(Rows are class membership results based on Model 2a; Columns are class membership results based on Model 2b)

By comparing the latent class membership between Model 2a and Model 2b, 8 cities have been identified in the *MIncreasingEff-MIncreasingEff* group, as these cities have experienced moderate increasing of both local government revenue density and employment density over the years. Interestingly, refer to the list of cities in Appendix Table B, these eight cities include not only the commonly known tier-1 city (i.e., Shanghai), but also cities from relatively lower administrative ranking (i.e., Wuhu, Jinhua). A number of higher-tier cities (i.e., Beijing, Wuhan, Chongqing, Hangzhou, Chengdu) concentrate on *Average-Average* group, and *Average-MDecreasingEff* group, which may imply that more cities in higher tiers may start to adopt spatial planning control to contain the growth.

Twelve cities in *MDecreasingEff-MIncreasingEff* group deserve more particular policy attention, as these cities have experienced an increasing of employment size, but the local government revenue has decreased slightly. It implies that the stimulation of employment growth has not yet brought effective improvement of revenue stream to the local government (i.e., through personal and corporate tax revenue). Thus, the planning strategy in these cities may need further investigation, though it takes time from stimulating local employment growth to the local government revenue growth.

Cities in *DecreasingEff-DecreasingEff* group tend to be small cities, where these 8 cities have seen a decline of both local government revenue density and employment density, indicating a general shrinking pattern. The decline of those cities is likely to be associated with the rise of major cities under spatial equilibrium. However, as more higher-tier cities start to adopt development control, the planning of smaller cities within those mega city regions becomes critical for improving equality and sustainability.

5.2.3 Relationship between the change of urban land-use efficiency and the change of built-up area

https://doi.org/10.5194/isprs-annals-X-4-W3-2022-261-2022

In order to further examine the possible relationship between the change of land-use efficiency and the rate of urban spatial expansion, two scatter plots are revealing shown in Figure 4a and Figure 4b. One distinctive feature is the distribution of data points, forming a broad triangle shape on both plots. The triangular distribution of cities suggests that cities with fast rate of spatial expansion are often associated with decreasing land-use efficiency, and that for cities with high land-use efficiency, they tend to exhibit a moderate rate of spatial expansion, if not becoming even more compact. The policy implication is twofold. Firstly, the majority of Tier-1 cities in China have adopted stringent land-use control, featuring zero net addition of construction land. Such restrictions on spatial expansion would be a prerequisite for improving land-use efficiency as per our measurement. The policy experience from those cities in terms of policy design, implementation and managing unintended impacts such as potential upward pressure on local house prices/rents would be valuable for other cities.

Secondly, there seems a tendency that some cities would expect to boost local economic growth through excessive spatial expansion, notably in the form of expanding existing administrative boundaries to create more construction land for real estate development. Admittedly, such spatial expansion could increase the absolute size of the local economy, often measured by GDP or employment, but our analysis shows that it often leads to decreased land-use efficiency. The decrease may involve two different mechanisms, i) averaging effect, where the relatively lower land-use efficiency in newly expanded areas (prior to any new land development) could reduce the average efficiency for the city as a whole; and ii) perhaps more concerning, the excessive land supply enabled by the boundary expansion may induce inefficient use of land measured by per km^2 economic output. Note that our analysis is based on urban built-up area.

It could be argued that the inefficient land use may be a result of increasing competition among lower-tier cities, as the quantity, use type and leasing price of land could be used as effective levers for attracting capital investments. Arguably, promoting local economic growth at the expense of land-use efficiency may bring short-term gains in terms of winning over other competing cities. However, improving land-use efficiency at a later stage may not be readily attainable because of the durability of land and building assets. Also, the prevalence of low-efficiency land use would disincentivise compact development as well as put a strain on providing essential infrastructure. The fact that some lower-tier cities can achieve high land-use efficiency without major boundary expansions suggests that spatial expansion is not a necessary condition for sustaining local growth. Improving the efficiency of land use would be a longer-term source of growth.

6. CONCLUSION

This study aims to investigate the heterogeneity of land-use efficiency across cities in mainland China. Through a novel application of LCA, the two purpose-built land-use efficiency measurements, i.e., local government revenue density and employment density, provides a novel method for understanding and monitoring the development patterns across cities. Our model shows that significant heterogeneity exists in both temporally static and dynamic manners. For the static assessment, though higher-tier cities (e.g., national/provincial capitals) often receive more concentrated political and economic support, lower-tier cities can also achieve high land-use efficiency and even outperform some cities in higher tiers. In terms of the longitudinal change of land-use efficiency, a number of lower-tier cities have achieved sustained improvement of land-use efficiency over the study period, the experience of which would be useful reference for other cities.

In terms of policy implications, our study suggests that, although the administrative city hierarchy renders important in terms of influencing the land development pattern and trajectory of cities, but the hierarchy per se is far from a determining factor of land-use efficiency. It is possible for lower-tier cities (non-capital cities) to thrive and even surpass higher-tier cities in terms of land-use efficiency, and the underlying policy practice and insights may be transferrable to other cities. The proposed method and analysis results are expected to assist policy makers to identify the latent groups with particular development patterns and address land-use efficiency gaps through knowledge sharing and place-based interventions.

Despite our best effort, this study only provides a descriptive modelling method to identify the heterogeneity performance across cities, and the drivers leading to the land-use efficiency heterogeneity needs to be further examined in the future. In terms of future research plan, the descriptive analysis of land-use efficiency heterogeneity through latent class analysis will be expanded with regression analysis aiming to identify potential causal relationship between generic urban planning strategies and development outcomes. Generic planning strategies will be extracted from official planning documents using conventional statistical and established Natural Language Processing methods.

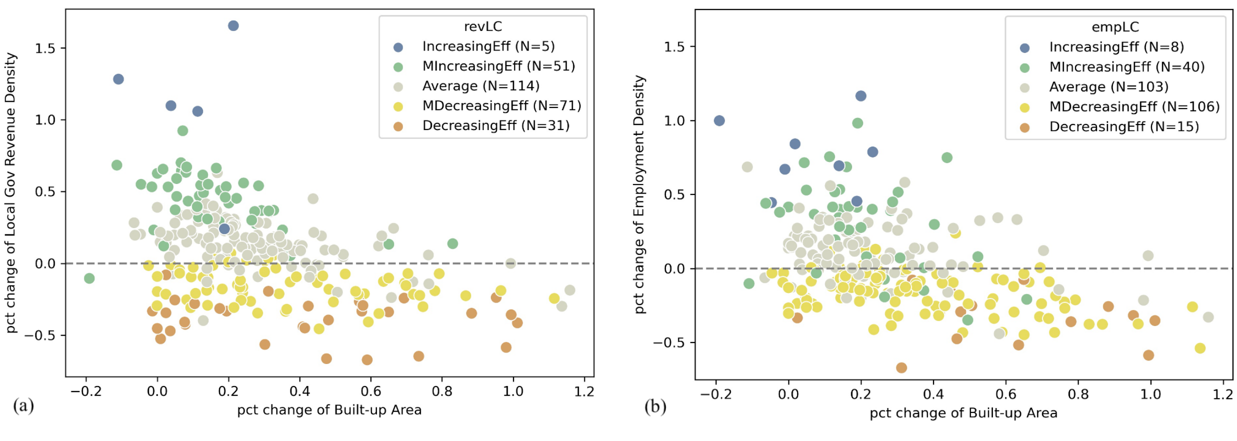


Figure 4. (a) Scatter plot of percentage change of built-up area and local government revenue density by class membership of Model 2a; (b) Scatter plot of percentage change of built-up area and employment density by class membership of Model 2b

Also, the identified latent city groups will be incorporated in the regression analysis through the use of categorical variables. The incorporation of city-group variables will help capture the non-linear variations across cities. It is expected that findings from the research would inform regional and urban planning policies that aim to tackle development disparities in China.

ACKNOWLEDGEMENTS

Tianyuan Wang would like to acknowledge Cambridge Trust and China Scholarship Council for providing the financial support for her PhD study at Cambridge. Both authors would like to thank anonymous reviewers for their insightful comments on the earlier version of this article.

REFERENCES

Cai, M., 2017. Revenue, time horizon, and land allocation in China. *Land Use Policy*, 62, 101–112.

Chakraborty, S., Maity, I., Dadashpoor, H., Novotnỳ, J., Banerji, S., 2022. Building in or out? Examining urban expansion patterns and land use efficiency across the global sample of 466 cities with million+ inhabitants. *Habitat International*, *120*, 102503.

Chen, Y., Chen, Z., Xu, G., Tian, Z., 2016. Built-up land efficiency in urban China: Insights from the General Land Use Plan (2006–2020). *Habitat International*, *51*, 31–38.

Chikaraishi, M., Fujiwara, A., Kaneko, S., Poumanyvong, P., Komatsu, S., Kalugin, A., 2015. The moderating effects of urbanization on carbon dioxide emissions: A latent class modeling approach. *Technological Forecasting and Social Change*, *90*, 302–317.

Dong, T., Jiao, L., Xu, G., Yang, L., Liu, J., 2019. Towards sustainability? Analyzing changing urban form patterns in the United States, Europe, and China. *Science of The Total Environment*, *671*,632–643.

Dong, Y., Jin, G., Deng, X., 2020. Dynamic interactive effects of urban land-use efficiency, industrial transformation, and carbon emissions. *Journal of Cleaner Production*, *270*, 122547. doi: 10.1016/j.jclepro.2020.122547.

Fox, B., Escue, M., 2021. Evaluating and comparing profiles of burglaries developed using three statistical classification techniques: cluster analysis, multidimensional scaling, and latent class analysis. *Psychology, Crime & Law*, *28*(1), 1–25. doi.org/10.1080/1068316x.2021.1880582.

Gao, X., Zhang, A., Sun, Z., 2020. How regional economic integration influence on urban land use efficiency？A case study of Wuhan metropolitan area, China. *Land Use Policy*, *90*, 104329. doi.org/10.1016/j.landusepol.2019.104329.

Haas, M. C. de, Scheepers, C. E., Harms, L. W. J., Kroesen, M., 2018. Travel pattern transitions: Applying latent transition analysis within the mobility biographies framework. *Transportation Research Part A: Policy and Practice*, *107*, 140–151. doi.org/10.1016/j.tra.2017.11.007.

Huang, G.-H., Bandeen-Roche, K., 2004. Building an identifiable latent class model with covariate effects on underlying and measured variables. *Psychometrika*, *69*(1), 5–32. doi.org/10.1007/bf02295837.

Jahanshahi, K., Jin, Y., 2020. Identification and mapping of spatial variations in travel choices through combining structural equation modelling and latent class analysis: findings for Great Britain. *Transportation*, 1–31.

Jiang, W., Feng, T., Timmermans, H. J. P., 2020. Latent class path model of intention to move house. *Socio-Economic Planning Sciences*, *70*, 100743.

Jin, G., Deng, X., Zhao, X., Guo, B., Yang, J., 2018. Spatiotemporal patterns in urbanization efficiency within the Yangtze River Economic Belt between 2005 and 2014. *Journal of Geographical Sciences*, *28*(8), 1113–1126.

Lee, W., Min, I. K., Yang, K. I., Kim, D., Yun, C.-H., Chu, M. K., 2021. Classifying migraine subtypes and their characteristics by latent class analysis using data of a nation-wide population-based study. *Scientific Reports*, *11*(1), 21595.

Li, C., Gibson, J., 2015. City scale and productivity in China. *Economics Letters*, *131*, 86–90.

Liao, F. H., Farber, S., Ewing, R., 2015. Compact development and preference heterogeneity in residential location choice behaviour: A latent class analysis. *Urban Studies*, *52*(2), 314–337. doi.org/10.1177/0042098014527138.

Liu, Y., Zhang, Z., Zhou, Y., 2018. Efficiency of construction land allocation in China: An econometric analysis of panel data. *Land Use Policy*, *74*, 261–272.

Liu, Z., 2019. Land-based finance and property tax in China. *Area Development and Policy*, 4(4), 1–15.

Lu, X., Kuang, B., Li, J., 2018. Regional difference decomposition and policy implications of China's urban land use efficiency under the environmental restriction. *Habitat International*, *77*, 32–39.

Nylund, K. L., Asparouhov, T., Muthen, B. O., 2007. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling*, *14*(4), 535–569.

Office for National Statistics (ONS)., 2018. Methodology and variables: Details of the methodology and 2011 Census statistics used for the 2011 area classification. Retrieved from: https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications/methodologyandvariables (last visit: 09 May 2022).

Rafiq, R., McNally, M. G., 2021. Heterogeneity in activity-travel patterns of public transit users. *Transportation Research Part A: Policy and Practice*, *152*, 1–18.

Rahman, Md. S., Mohiuddin, H., Kafy, A.-A., Sheel, P. K., Di, L., 2019. Classification of cities in Bangladesh based on remote sensing derived spatial characteristics. *Journal of Urban Management*, *8*(2), 206–224. doi.org/10.1016/j.jum.2018.12.001.

Song, Y., Yeung, G., Zhu, D., Xu, Y., Zhang, L., 2022. Efficiency of urban land use in China's resource-based cities, 2000–2018. *Land Use Policy*, *115*, 106009.

Vermunt, J. K., 2010. Latent Class Modeling with Covariates: Two Improved Three-Step Approaches. *Political Analysis*, *18*(4), 450–469. doi.org/10.1093/pan/mpq025.

Wan, L., An, Y., 2017. Identifying Growth Patterns of the High-Tech Manufacturing Industry across the Seoul Metropolitan Area Using Latent Class Analysis. *Journal of Urban Planning and Development*, *143*(3), 04017011.

Wan, L., Wang, T., Bao, H. X. H., 2021. Understanding the role of land finance in economic and infrastructure development in Chinese cities: new evidence from a novel structural equation modelling approach. *SSRN*. doi.org/10.2139/ssrn.3973056.

Wu, C., Wei, Y. D., Huang, X., Chen, B., 2017. Economic transition, spatial development and urban land use efficiency in the Yangtze River Delta, China. *Habitat International*, *63*, 67–78. doi.org/10.1016/j.habitatint.2017.03.012.

Yan, S., Peng, J., Wu, Q., 2020. Exploring the non-linear effects of city size on urban industrial land use efficiency: A spatial econometric analysis of cities in eastern China. *Land Use Policy*, *99*, 104944. doi.org/10.1016/j.landusepol.2020.104944.

Zhang, S., Fang, C., Kuang, W., Sun, F., 2019. Comparison of Changes in Urban Land Use/Cover and Efficiency of Megaregions in China from 1980 to 2015. *Remote Sensing*, *11*(15), 1834. doi.org/10.3390/rs11151834.

Zhou, H., Norman, R., Xia, J., Hughes, B., Kelobonye, K., Nikolova, G., Falkmer, T., 2020. Analysing travel mode and airline choice using latent class modelling: A case study in Western Australia. *Transportation Research Part A: Policy and Practice*, *137*, 187–205. doi.org/10.1016/j.tra.2020.04.020.

Zhou, J., Jin, L.-R., Tao, M.-J., Peng, H., Ding, S.-S., Yuan, H., 2019. The underlying characteristics of sleep behavior and its relationship to sleep-related cognitions: a latent class analysis of college students in Wuhu city, China. *Psychology, Health & Medicine*, *25*(7), 1–11.

Zhu, X., Li, Y., Zhang, P., Wei, Y., Zheng, X., Xie, L., 2019. Temporal–spatial characteristics of urban land use efficiency of China's 35mega cities based on DEA: Decomposing technology and scale efficiency. *Land Use Policy*, *88*, 104083.

APPENDIX

Table A. Latent class analysis results of Model 1.

	Name of Cities (in Chinese)
HighEff (N=27)	三明市, 上海, 上饶市, 丽水市, 佛山市, 北京, 南通市, 台州市, 吕梁市, 嘉兴市, 孝感市, 宁德市, 宁波市, 宜春市, 岳阳市, 廊坊市, 延安市, 无锡市, 榆林市, 毕节市, 沧州市, 泰州市, 深圳市, 苏州市, 鄂尔多斯市, 金华市, 黄冈市
Average (N=135)	三门峡市, 东莞市, 中山市, 临汾市, 临沂市, 临沧市, 乌鲁木齐市, 九江市, 云浮市, 亳州市, 保定市, 保山市, 信阳市, 六安市, 六盘水市, 兰州市, 十堰市, 南京市, 南充市, 南宁市, 南平市, 南昌市, 南阳市, 厦门市, 合肥市, 吉安市, 周口市, 哈尔滨市, 唐山市, 商丘市, 四平市, 大连市, 天津, 太原市, 威海市, 娄底市, 安阳市, 宜宾市, 宜昌市, 宣城市, 宿迁市, 巴中市, 常州市, 常德市, 平顶山市, 广州市, 庆阳市, 开封市, 张家口市, 徐州市, 德州市, 忻州市, 怀化市, 惠州市, 成都市, 扬州市, 抚州市, 揭阳市, 新乡市, 昆明市, 昭通市, 晋中市, 晋城市, 杭州市, 枣庄市, 株洲市, 梅州市, 武汉市, 永州市, 汉中市, 汕头市, 汕尾市, 江门市, 沈阳市, 河源市, 泉州市, 泰安市, 洛阳市, 济南市, 济宁市, 海口市, 淄博市, 淮安市, 清远市, 温州市, 渭南市, 湖州市, 湘潭市, 湛江市, 滁州市, 滨州市, 潍坊市, 濮阳市, 烟台市, 焦作市, 玉林市, 珠海市, 百色市, 盐城市, 石家庄市, 福州市, 绍兴市, 绥化市, 绵阳市, 聊城市, 芜湖市, 荆州市, 荆门市, 菏泽市, 萍乡市, 衡水市, 衡阳市, 衢州市, 西安市, 许昌市, 贵阳市, 赣州市, 达州市, 运城市, 邢台市, 邯郸市, 邵阳市, 郑州市, 郴州市, 重庆, 铜仁市, 镇江市, 长春, 长沙市, 长治市阜阳市, 随州市, 青岛市, 驻马店市, 黄石市
LowEff (N=110)	三亚市, 东营市, 中卫市, 丹东市, 丽江市, 乌海市, 乐山市, 伊春市, 佳木斯市, 克拉玛依市, 内江市, 包头市, 北海市, 双鸭山市, 吉林市, 吴忠市, 呼伦贝尔市, 呼和浩特市, 咸宁市, 咸阳市, 商洛市, 嘉峪关市, 固原市, 大同市, 大庆市, 天水市, 安庆市, 安康市, 安顺市, 宝鸡市, 崇左市, 巴彦淖尔市, 平凉市, 广元市, 广安市, 张家界市, 张掖市, 德阳市, 承德市, 抚顺市, 拉萨市, 攀枝花市, 新余市, 日照市, 景德镇市, 朔州市, 朝阳市, 本溪市, 来宾市, 松原市, 柳州市, 桂林市, 梧州市, 武威市, 池州市, 河池市, 泸州市, 淮北市, 淮南市, 漯河市, 潮州市, 牡丹江市, 白城市, 白山市, 白银市, 益阳市, 盘锦市, 眉山市, 石嘴山市, 秦皇岛市, 肇庆市, 自贡市, 舟山市, 茂名市, 营口市, 葫芦岛市, 蚌埠市, 西宁市, 贵港市, 贺州市, 赤峰市, 辽源市, 辽阳市, 连云港市, 通辽市, 遂宁市, 遵义市, 鄂州市, 酒泉市, 金昌市, 钦州市, 铜川市, 铜陵市, 银川市, 锦州市, 阜新市, 防城港市, 阳江市, 阳泉市, 雅安市, 鞍山市, 韶关市, 马鞍山市, 鹤壁市, 鹤岗市, 鹰潭市, 黄山市, 黑河市, 齐齐哈尔市, 龙岩市

Table B. Class membership matrix of Model 2a and Model 2b (Rows are latent classes based on Model 2a, revenue Density; Columns are latent classes based on Model 2b, employment Density; Name of cities in Chinese)

revDen LC\empDen LC	**IncreasingEff (N=8)**	**MIncreasingEff (N=40)**	**Average (N=103)**	**MDecreasingEff (N=106)**	**DecreasingEff (N=15)**
Increasing Eff (N=5)	贺州市	拉萨市	岳阳市, 湘潭市	张家界市	-
MIncreasing Eff (N=51)	乌海市, 池州市, 通辽市	上海, 丽水市, 宜昌市, 滁州市, 漯河市, 贵港市, 酒泉市, 铜仁市	中卫市, 临沂市, 佛山市, 南充市, 南阳市, 咸宁市, 商丘市, 廊坊市, 开封市, 梅州市, 武威市, 永州市, 河源市, 济南市, 玉林市, 荆州市, 荆门市, 遂宁市, 郴州市, 驻马店市, 鹤壁市, 黄石市	九江市, 周口市, 固原市, 天水市, 宣城市, 巴中市, 怀化市, 惠州市, 日照市, 株洲市, 江门市, 深圳市, 潍坊市, 珠海市, 衡水市, 邵阳市, 钦州市, 随州市	-
Average (N=114)	宜春市, 宝鸡市, 马鞍山市	东莞市, 丽江市, 乌鲁木齐市, 吉安市, 平凉市, 抚州市, 无锡市, 晋城市, 柳州市, 汉中市, 泸州市, 湛江市, 芜湖市, 许昌市, 金华市, 阜阳市, 鹰潭市	北京, 武汉市,鄂尔多斯市, 伊春市, 信阳市, 六安市, 台州市, 合肥市, 唐山市, 嘉兴市, 嘉峪关市, 太原市, 宁德市, 宁波市, 安庆市, 安康市, 安阳市, 宿迁市, 常德市, 平顶山市, 广安市, 张家口市, 徐州市, 扬州市, 新乡市, 洛阳市, 海口市, 淄博市, 温州市, 滨州市, 濮阳市, 焦作市, 白城市, 益阳市, 眉山市, 石家庄市, 福州市, 舟山市, 苏州市, 茂名市, 衡阳市, 衢州市, 赤峰市, 遵义市, 郑州市,金昌市, 镇江市, 长沙市, 阳江市, 黄冈市, 齐齐哈尔市	成都市, 杭州市, 东营市, 中山市, 亳州市, 保定市, 保山市, 兰州市, 包头市, 南京市, 南昌市, 厦门市, 威海市, 娄底市, 孝感市, 安顺市, 宜宾市, 广元市,昭通市, 景德镇市,枣庄市, 汕头市, 沧州市, 河池市, 泉州市, 湖州市, 烟台市, 聊城市, 自贡市, 萍乡市, 蚌埠市, 西宁市, 西安市, 邢台市, 鄂州市, 重庆, 银川市, 防城港市, 韶关市, 黑河市	云浮市, 天津, 张掖市
MDecreasing Eff (N=71)	毕节市	佳木斯市, 北海市, 南宁市, 吉林市, 延安市, 新余市, 沈阳市, 泰安市, 淮北市, 肇庆市, 邯郸市, 长春市	三亚市, 三明市, 临汾市, 临沧市, 吴忠市, 咸阳市, 常州市, 庆阳市, 德阳市, 承德市, 攀枝花市, 松原市, 桂林市, 淮安市, 白银市, 绵阳市, 菏泽市, 辽源市, 铜川市, 长治市, 青岛市, 黄山市	三门峡市, 上饶市, 乐山市, 克拉玛依市, 内江市, 南平市, 南通市, 呼伦贝尔市, 呼和浩特市, 哈尔滨市, 商洛市, 四平市, 大同市, 大庆市, 大连市, 崇左市, 广州市, 忻州市, 昆明市, 泰州市, 清远市, 渭南市, 牡丹江市, 盐城市, 石嘴山市, 绥化市, 贵阳市, 赣州市, 达州市, 运城市, 连云港市, 龙岩市	巴彦淖尔市, 德州市, 济宁市, 百色市
Decreasing Eff (N=31)	-	朔州市, 锦州市	本溪市, 汕尾市, 淮南市, 葫芦岛市, 铜陵市, 阳泉市	丹东市, 十堰市, 双鸭山市, 吕梁市, 抚顺市, 来宾市, 梧州市, 榆林市, 白山市, 盘锦市, 秦皇岛市, 营口市, 阜新市, 鞍山市, 鹤岗市	六盘水市, 揭阳市, 晋中市, 朝阳市, 潮州市, 绍兴市, 辽阳市, 雅安市