


Article

An Urban Scaling Estimation Method in a Heterogeneity Variance Perspective

Wenjia Wu ^{1,2,3}, Hongrui Zhao ^{1,2,3,*}, Qifan Tan ^{1,2,3} and Peichao Gao ⁴ 

¹ Institute of Geomatics, Department of Civil Engineering, Tsinghua University, Beijing 100084, China; wu-wj14@mails.tsinghua.edu.cn (W.W.); tqf17@mails.tsinghua.edu.cn (Q.T.)

² 3S Center, Tsinghua University, Beijing 100084, China

³ National Engineering Laboratory for Green and Safe Construction Technology in Urban Rail Transit, Tsinghua University, Beijing 100084, China

⁴ Faculty of Geographical Science, Beijing Normal University, Beijing 100084, China; gaopc@bnu.edu.cn

* Correspondence: zhr@tsinghua.edu.cn; Tel.: +86-010-62782682

Received: 29 January 2019; Accepted: 25 March 2019; Published: 28 March 2019



Abstract: Urban scaling laws describe powerful universalities of the scaling relationships between urban attributes and the city size across different countries and times. There are still challenges in precise statistical estimation of the scaling exponent; the properties of variance require further study. In this paper, a statistical regression method based on the maximum likelihood estimation considering the lower bound constraints and the heterogeneous variance of error structure, termed as CHVR, is proposed for urban scaling estimation. In the CHVR method, the heterogeneous properties of variance are explored and modeled in the form of a power-of-the-mean variance model. The maximum likelihood fitting method is supplemented to satisfy the lower bound constraints in empirical data. The CHVR method has been applied to estimating the scaling exponents of six urban attributes covering three scaling regimes in China and compared with two traditional methods. Method evaluations based on three different criteria validate that compared with both classical methods, the CHVR method is more effective and robust. Moreover, a statistical test and long-term variations of the parameter in the variance function demonstrate that the proposed heterogeneous variance function can not only describe the heterogeneity in empirical data adequately but also provide more meaningful urban information. Therefore, the CHVR method shows great potential to provide a valuable tool for effective urban scaling studies across the world and be applied to scaling law estimation in other complex systems in the future.

Keywords: scaling laws; heterogeneity; China; urban systems; complex systems

1. Introduction

Urban scaling law, as one of the most significant regulations in urban science [1], has drawn much attention for providing powerful universalities of the variations of urban attributes with city size [2–6]. The urban scaling law state that the relationship between urban attributes and population is well-described by the power-law scaling of the form $y = ax^b$ [2], which was inspired by the allometric scaling in biology [7,8]. This law provides a systematic and quantitative framework for urban sustainable development [2,9]. Urban scaling laws reveal empirical regularity for cities across different urban attributes [3,10,11], nations [12–15], and times [11], validated in both empirical [2] and theoretical [3] ways. Urban scaling studies also provide empirical evidence for the economies of scale and the agglomeration effect beneath the sustainable development of cities [10,16]. Urban scaling laws have been widely applied to analyzing urban hierarchy [17], infrastructure strategies [18], urban designs [19], and environmental strategies [20].

The universality of urban scaling has been questioned in recent studies [21]. It has been observed that the scaling exponent is sensitive to the definition of the city [13], the utilized database [22], and the regression method [23]. For example, the scaling of transportation-related CO₂ emissions changes with the population results in different conclusions at different aggregation levels [7]. Opposite conclusions about the foreign investment in the French urban system can be reached using the traditional ordinary least squares (OLS) approach and the computing framework avoiding zero values proposed by Finance [23]. These situations call for more solid statistical support for the study of urban scaling laws.

Challenges still exist for precise statistical estimation methods of the scaling exponent [24,25]. The linear regression method on log-transformed data is utilized widely due to its simplicity, convenience without any iterative process [2,26]. However, the analysis using logarithmic scales involves rotational distortion [27], undetected outliers [28], a multiplicative error without a justification and interpretation [27,29], and a heterogeneous error without justification [30], resulting in a biased estimation result. Researchers suggested that the analysis of urban scaling laws should be carried out on the original scale using a nonlinear regression method [31]. However, heterogeneity of variance has been observed, and the underlying assumption in the regression method is invalid [32]. Furthermore, Leitão [33] argued that challenges should be attributed to statistical properties of cities; namely, the heterogeneous fluctuations of urban attribute Y as a function of population P . Moreover, the ordinary least squares regression method may result in systematic errors in the power laws [25] and makes an underlying assumption of homogeneous variance without a statistical test. Therefore, the variance of the error model should be considered in urban estimation methods.

The properties of the variance of the error model remain largely unexplored and require further study [26]. Previous studies were interested in obtaining estimates of the urban scaling exponents but did not take enough consideration of the error structure [29]. Both traditional methods utilized the OLS regression with an underlying assumption of homogeneity. However, it has been suggested that heterogeneous variance appeared ubiquitously in complex systems [34,35], and nontrivial fluctuations have been reported in previous studies of urban data [36–38]. The heterogeneous variance of the error structure should be considered in the scaling regression, which has been documented in the fields of biology [29] and ecology [32]. It is believed that there exists important information given the heterogeneity of the variance, which should be thoroughly investigated in the models of urban scaling.

In this paper, a statistical framework based on the maximum likelihood estimation considering the lower bound constraints and the heterogeneous variance regression of the error structure, termed as CHVR, is proposed. Traditional approaches for estimating urban scaling parameters are firstly explored, and their limitations are summarized. The properties of heterogeneous variance are explored and integrated into the maximum likelihood-based method. The CHVR method is applied to estimating six urban attributes in China and compared with two traditional methods. Method assessments and a statistical test are performed to validate the CHVR method. Furthermore, the interpretation of the parameter in the variance model is explored to provide extra urban information.

2. Study Area and Data

2.1. Study Area

China is considered as the study area for analyzing urban scaling estimation methods, given the rapid urban development [39] and a complex urbanization pattern [40]. Much attention has been given to empirical urban scaling studies in China from different perspectives [41,42]. There are five levels of administrative divisions in mainland China, namely, province, prefecture, county, township, and village. Urban scaling exploration in China needs an appropriate definition of a city, since the urban scaling exponent is sensitive to it, as has been examined theoretically and empirically [13]. There are three statistical units related to the definition of a city in the statistical data, including prefecture (or total city), municipal district (or districts under the city), and built-up area (or built district), as shown in Figure 1a,b. Considering the comparison to Metropolitan Statistical Areas (MSA) in the

USA [22], a consistent definition of functional cities, and previous studies in China [15,22,31], urban scaling estimation methods and results are explored using data aggregated at the municipal district (“shixiaqu”) scale for each prefecture. Due to the lack of data and the removing of the inaccurate records, about 275 out of 333 prefectures in mainland China were chosen for the analysis at the municipal district scale, as shown in Figure 1a.

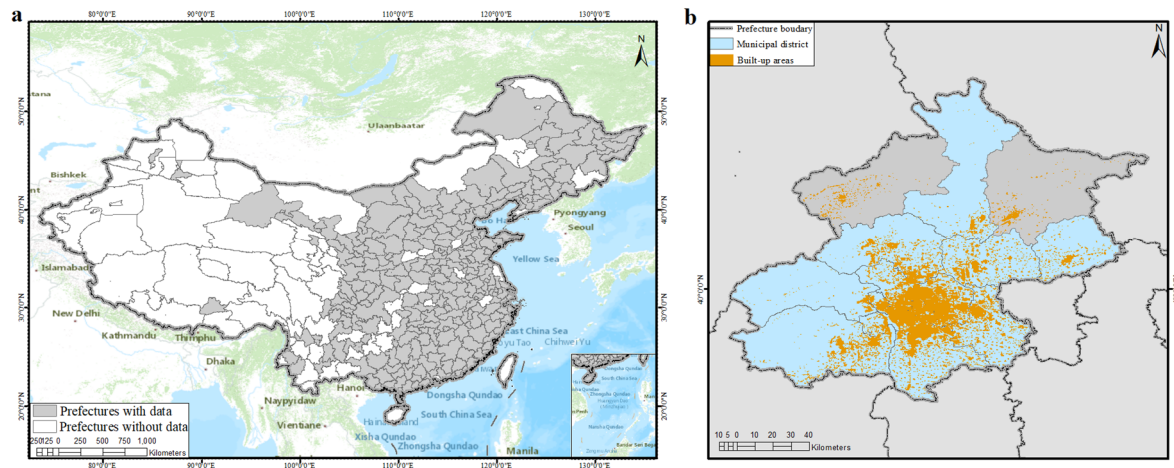


Figure 1. Study areas and statistical units in China in 2010. (a) Prefectures in China in 2010. Prefectures in grey represent selected prefectures and prefectures in white, mainly located in the western part, represent prefecture without appropriate data excluded in this study. (b) Example of three statistical units related to the definition of the city in the statistical data. Taking Beijing in 2010 as an example, the prefecture in grey refers to the whole Beijing, the municipal district in blue refers to the sum of 16 districts except two counties, and built-up area in orange refers to centralized contiguous parts in the central city and urban construction land with complete infrastructure in suburban areas.

2.2. Demographic Data

Demographic data are important for urban scaling as an independent variable. There are two kinds of demographic data in mainland China: household-registered population according to the ‘Hukou’ system, and resident population considering migration. The resident population is more closely related to actual citizens [43] who engage in social interactions and shape city functions; moreover, household-registered population deviates significantly from the resident population for a given city. For example, the size of the household-registered population was only one quarter of the resident population in Shenzhen, China, in 2010. Thus, resident population data at the municipal district level are calculated from the census data in 2010 and are then used to analyze and compare three urban scaling estimation methods in Section 6. Resident population data at the municipal district level in 2000, 2005, 2010, and 2015 are gathered to explore the urban interpretation of the proposed parameter in Section 7.3. Resident population in 2000 and 2010 are provided by censuses, while resident population in 2005 and 2015 are gathered from the China micro-census with the supplement of City Statistical Yearbook and yearly statistical bulletin on national economic and social development for each city at the prefecture level.

2.3. Urban Attributes Data

Statistical information for various urban attributes, namely the Gross Domestic Product (GDP), wage data, water consumption data, electricity consumption data, road area data, and built-up area data at the municipal district level in China for years 2000, 2005, 2010, and 2015 were gathered as the dependent variable in the urban scaling law analysis. These variables were selected with the intention to encompass a range of urban domains in a descriptive way and cover the scaling regimes (super-linear, linear and sub-linear), based on the hypothesis [2,3,10] that super-linear scaling are

manifested by quantities related to socioeconomic productivity (GDP, total income), linear scaling are usually associated with basic individual services (household water consumption, household electricity consumption), and sub-linear scaling are manifested by infrastructure demand (built-up area, road surface area). Metadata are collected from the China City Statistical Yearbook. Specifically, GDP at constant price are utilized for comparative studies by year, and paved roads area data are utilized as the measure of road surface area. Then, the necessary pre-processing, including data structuring and deleting outliers, was performed on the metadata to derive the structured urban attributes data.

3. Urban Scaling Laws: Principles and Problems with Estimation

In this section, background knowledge of urban scaling laws is introduced firstly, and then classical approaches to estimating urban scaling parameters and the limitations of such approaches are presented. In addition, heterogeneity of variance of the traditional methods is explored further.

3.1. Principles of Urban Scaling Laws

Urban scaling laws [2], also termed as allometric scaling laws for cities, stated that urban attributes Y has a power-law relationship with urban population P , which can be written as:

$$Y = Y_0 * P^\beta, \quad (1)$$

where Y_0 is a standardized constant and β is termed as the scaling exponent. Unlike the allometric scaling in biology, whereby all metabolic rates scale sub-linearly with body mass with the exponent of $\frac{3}{4}$ [8,44], the scaling exponents β a wide range of urban properties falls into three universal categories [2,3]:

- $\beta > 1$ shows a super-linear scaling regime associated with social currencies (e.g., GDP, and the number of patents [45]), describing the agglomeration effect. For example, although the population of Beijing is double that of Wuhan, the GDP of Beijing is more than double that of Wuhan.
- $\beta = 1$ denotes a linear scaling regime associated with individual human needs (e.g., jobs and household water consumption).
- $\beta < 1$ means a sub-linear scaling regime associated with infrastructural variables (e.g., urban areas and road length), describing economies of scale.

These urban scaling regimes (super-linear, linear, and sub-linear) for urban attributes can reflect non-linear urban development mechanisms, and scaling exponent can provide a measurement of agglomeration effect and scale effect for urban attributes [2,3].

The urban system is regarded as an evolving set of geographically and economically coherent [46] cities at the macroscopic scale [41], characterized by specific behaviors [2]. Empirical studies verified that urban scaling law reflects universal scaling relationships that apply to all urban systems with different levels of development (including the USA [12], Britain [13], Brazil [14], and China [15]).

In urban scaling estimations, the main question is: given a set of cities, how can the exponent be estimated? Is this scaling super-linear, linear, or sub-linear? Moreover, is the scaling exponent well-evaluated? The scaling exponent has important meanings in urban dynamics studies for the economies of scale and the agglomeration effect beneath the sustainable development of cities.

3.2. Traditional Urban Scaling Statistic Methods

Firstly, the estimation of urban scaling parameters is described in detail from the mathematical perspective. Given a set of cities $\{n_i, i = 1, 2, \dots, n\}$, the relationship between the set of urban attributes observations $\{y_i, i = 1, 2, \dots, n\}$ and the set of urban population observations $\{p_i, i = 1, 2, \dots, n\}$ can be formed as:

$$y_i = Y_0 * p_i^\beta. \quad (2)$$

However, based on the urban scaling laws, Equation (2) describes the ideal situation where the urban attributes set $\{y_i, i = 1, 2, \dots, n\}$ and population set $\{p_i, i = 1, 2, \dots, n\}$ are observed without any error. Considering the reality, we assume that the scaling form in Equation (2) achieves on average. This notion can be formulized by utilizing the conventional mean urban attributes $E(y_i|p_i)$, and Equation (2) can be rewritten as:

$$E(y_i|p_i) = Y_0 * p_i^\beta. \quad (3)$$

There are two common kinds of methods for parameter evaluation in Equation (2). Linear regression based on OLS on logarithmic scales, represented by the LOG_LR method in this paper, is widely used in scaling studies in urban [11] and biology field considering its simplicity [27]. In detail, Equation (2) is log-transformed as linear form as follows:

$$\log(y_i) = \beta * \log(p_i) + \log(Y_0). \quad (4)$$

Then, the linear regression is conducted based on ordinary least squares criterion. That is to say, the error ε_i is assumed to be normally distributed on log-transformed data as:

$$\log(y_i) = \beta * \log(p_i) + \log(Y_0) + \varepsilon_i, \text{ and } \varepsilon_i \sim N(0, \sigma^2), \quad (5)$$

where σ^2 represents the variance of the error model ε . The error ε_i can be referred on the arithmetic scale as follows:

$$y_i = Y_0 * p_i^\beta * e^{\varepsilon_i}, \text{ and } \varepsilon_i \sim N(0, \sigma^2), \quad (6)$$

corresponding to lognormally distributed multiplicative errors.

The other method applying widely in estimation scaling parameters for urban scaling laws [31] and allometric scaling laws [27] utilized nonlinear regression on the original arithmetic scale, represented by the NLR method. It is assumed that error ε_i is normally distributed as:

$$y_i = Y_0 * p_i^\beta + \varepsilon_i, \text{ and } \varepsilon_i \sim N(0, \sigma^2). \quad (7)$$

Then, the optimization algorithm, such as the Gauss-Newton method, is conducted to solve the non-linear least squares problems.

3.3. Limitations

As to the LOG_LR method, although linear regression on double-log-transformed data is widely utilized due to its convenience and the lack of recursive iteration, there are four limitations resulting from the transformation to logarithmic scales and a lack of model diagnosis as follows:

1. The LOG_LR method transforms the power function relating explanatory and response variables into a linear function. Thus, the optimal function has changed in the OLS method, potentially resulting in bias in the estimation. It has been demonstrated that the LOG_LR method predicts geometric means for urban attributes instead of arithmetic means [30]. Log-linear regression based on geometric mean is valuable to the hydrological community [6]; however, scenarios of urban scaling estimations are different from Horton laws originated by entropy maximization [7]. Here, we try to give a statistical justification. From Equation (5), we obtain:

$$E(\log(y_i)|p_i) = \beta * \log(p_i) + \log(Y_0). \quad (8)$$

The conditional probability of $\log(y_i)$ can be written as:

$$P(\log(y_i)|p_i) = \frac{1}{\sqrt{2\pi} * \sigma} * e^{-\frac{(\log(y_i) - \mu_i)^2}{2\sigma^2}}, \text{ where } \mu_i = \log(Y_0) + \beta * \log(p_i). \quad (9)$$

Utilizing the integration formula for a compound function, we can derive:

$$P(y_i|p_i) = \frac{1}{\sqrt{2\pi} * \sigma} * \frac{1}{y} * e^{-\frac{(\log(y)-\mu_i)^2}{2\sigma^2}} \quad (10)$$

From Equation (10), it can be inferred that:

$$E(y_i|p_i) = \int_0^{+\infty} P(y|p_i) * y * dy = \int_0^{+\infty} \frac{1}{\sqrt{2\pi} * \sigma} * \frac{1}{y} * e^{-\frac{(\log(y)-\mu_i)^2}{2\sigma^2}} * y * dy. \quad (11)$$

Omitting the specifics of the integration process, Equation (11) can be computed as:

$$E(y_i|p_i) = Y_0 * p_i^\beta * e^{\frac{\sigma^2}{2}}. \quad (12)$$

Thus, it can be verified that:

$$E(y_i|p_i) \neq e^{E(\log(y_i)|p_i)}, \quad (13)$$

that is $E(y_i|p_i) \neq Y_0 * p_i^\beta$. Therefore, it can be verified that there are transformation biases.

2. Log-transformation narrows the gap in data; therefore, several outliers may remain undetected [28].

3. Multiplicative errors on the original scale [27] lack statistical justification and meaningful interpretations [29].

4. The LOG_LR method makes an underlying assumption that errors are heterogeneous on the arithmetic scale [30], which needs further verification. If the errors are heterogeneous on the arithmetic scale, the weighting may be inaccurate due to a lack of analysis of errors model. Thus, it demonstrated that the estimation of urban scaling parameters should be performed on the original arithmetic scale.

Regarding the NLR method, however, previous non-linear estimation of scaling laws needs more consideration of the model diagnostics [47] and a common heterogeneous error structure for complex systems.

For both methods, it is only if certain assumptions hold [47] that parameters estimated by the least squares method are consistent with estimators obtained by the maximum likelihood method that has been verified to be more accurate. Therefore, utilizing the least squares method in both traditional methods may introduce some systematic errors [25].

In summary, limitations still exist for both methods, which result from the log transformation, a lack of model structure and statistical justification, and the least squares method. In the following sections, the properties of variance are explored; afterwards, an alternative statistical method based on the maximum likelihood fitting is proposed to estimate the parameters and test urban scaling laws.

3.4. Variance Homogeneity of the Fitting Method

Variance homogeneity is one of the important parts of a model structure. Heterogeneity refers to the spread of the data not being the same at each value of the independent variable [32]. The assumption of homogeneity means that conditional on the independent variables, the variance of errors is constant. The alternative means that the variance of errors varies at different values of independent variables, as shown in Figure 2.

The homogeneity of variance can be evaluated through a visual examination of residuals. Figure 3 shows the plot of the log-transformed (to make the effect clearer) fitted value versus the residuals for the NLR method. The plot seems to have the shape of a horn with diameter varying from the left to the right. More specifically, as the fitted city attributes increase, it seems that although there is some unevenness in general, it is growing. It can be inferred that the variance is an increasing function of the mean. This finding demonstrates that the assumption of homogeneity of variance is not satisfied. As shown in Figure 2, the variances of adaptive ability differ for cities at different levels of the hierarchical structure of urban systems, including small, medium and large cities. As we know, the

population may follow the power law distribution, so that extreme values will occur. There are model violations related to errors being normally distributed with heterogeneous variance [47]. Since a larger population results in larger residuals, the residuals from the larger population will be overestimated compared with those for a relatively small population. Figures 2 and 3 show the tendency of the variation to increase with the mean value [47].

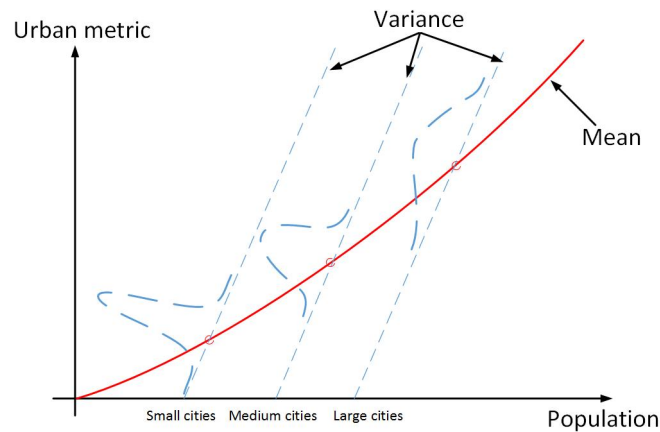


Figure 2. Schematic diagram of heterogeneous variance in the fitting method of urban scaling laws. The red line refers to the fitting method, and the blue dashed lines represent error models with various values of the variance at different values of the independent variable, e.g., for small, medium, and large cities.

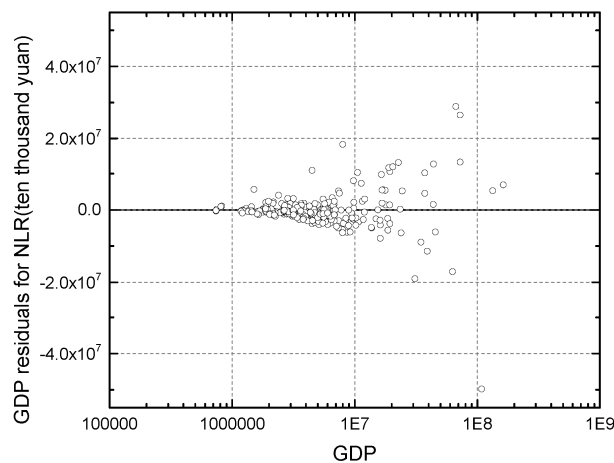


Figure 3. Plot of the log-transformed (to make the effect clearer) fitted values versus the residuals of the Gross Domestic Product (GDP) for the nonlinear regression method.

More statistical methods are needed for testing heterogeneity further. If there are available replicated datasets, a statistical test can be performed quantitatively, such as Levene's test [48]. For urban scaling estimations without replicates, Breusch–Pagan test [49], which tests whether the variance of the residuals is dependent with the independent variables, will be utilized and discussed in Section 6.2.

Therefore, heterogeneous variance is discovered in fitting methods and should be considered in the proposed method for accuracy and to obtain extra urban information [32].

4. The CHVR Method

In this section, the maximum probability estimation is utilized to estimate the scaling exponent. The heterogeneous variance will be explored and considered on the error model. The maximum

likelihood estimation integrated with the lower bound constraints and the heterogeneity variance regression of error structure, termed as CHVR, is presented in the following sections.

4.1. Maximum-Likelihood Fitting Method Based on Variance Function

For CHVR, the first central question is the likelihood \mathcal{L} function of the data. As mentioned in Section 3, in the traditional non-linear fitting method, we assume that the error model is subject to normally distributed, that is:

$$y_i = Y_0 * p_i^\beta + \varepsilon_i, \text{ and } \varepsilon_i \sim N(0, \sigma^2). \tag{14}$$

Then, the conditional probability $P(y|p)$ has the form:

$$P(y|p) = \frac{1}{\sqrt{2\pi} * \sigma} * e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \text{ where } \mu = Y_0 * p^\beta. \tag{15}$$

In the CHVR method, heterogeneous variance refers that the variance is dependent on the mean which implies cities with larger population obtain larger errors. Therefore, the heterogeneity will be integrated into the error model, and the variance of the error model is assumed as:

$$\text{var}(\varepsilon_i) = \sigma^2(Y_0 * p_i^\beta)^{2\theta}, \tag{16}$$

where θ denotes the spread of the variance [50] and typically $\theta \in [0.5, 1]$ in complex systems [34]. This assumption is motivated by Taylor’s law [51] in the complex science field, and power-of-the-mean variance model [47] in the statistic field. The variance in Equation (16) describes the heterogeneity property, which may be due to mixed effect from the variation of human activity and the fluctuation of data.

Therefore, based on the variance of the error model, we propose the urban scaling law model considering heterogeneity as:

$$y_i = Y_0 * p_i^\beta + \varepsilon_i, \text{ and } \varepsilon_i \sim N(0, \sigma^2(Y_0 * p_i^\beta)^{2\theta}). \tag{17}$$

The conditional probability $P(y|p)$ in this method can be computed as:

$$P(y|p) = \frac{1}{\sqrt{2\pi} * \sigma(Y_0 * p^\beta)^\theta} * e^{-\frac{(y-\mu)^2}{2\sigma^2(Y_0 * p_i^\beta)^{2\theta}}}, \tag{18}$$

where $\mu = Y_0 * p^\beta$. Based on the definition of the maximum-likelihood fitting method, the likelihood function \mathcal{L} can be written as:

$$\begin{aligned} \mathcal{L}(\beta, Y_0, \theta, \sigma|p, y) &= P(p, y|\beta, Y_0, \theta, \sigma) \\ &= P(p_1, y_1|\beta, Y_0, \theta, \sigma)P(p_2, y_2|\beta, Y_0, \theta, \sigma) \dots P(p_n, y_n|\beta, Y_0, \theta, \sigma), \end{aligned} \tag{19}$$

that is:

$$\mathcal{L}(\beta, Y_0, \theta, \sigma|p, y) = \prod_{i=1}^{i=n} P(p_i, y_i|\beta, Y_0, \theta, \sigma). \tag{20}$$

In order to maximize the likelihood function, we log-transformed the Equation (20) as:

$$\log(\mathcal{L}(\beta, Y_0, \theta, \sigma|p, y)) = \sum_{i=1}^{i=n} \log(P(p_i, y_i|\beta, Y_0, \theta, \sigma)). \tag{21}$$

The joint probability density function can be expanded as:

$$P(p_i, y_i|\beta, Y_0, \theta, \sigma) = P(y|p_i, \beta, Y_0, \theta, \sigma) * P(p_i|\beta, Y_0, \theta, \sigma), \tag{22}$$

where $P(p_i|\beta, Y_0, \theta, \sigma)$ denotes the prior probability of the urban population. The prior probability can be obtained by the maximum entropy statistics of the urban system [52,53]. However, it can be suggested that $P(p_i|\beta, Y_0, \theta, \sigma)$ is independent of any of estimators of urban scaling law. Thus, in our method, $P(p_i|\beta, Y_0, \theta, \sigma)$ can be regarded as a constant in the process of fitting, that is:

$$\log(\mathcal{L}(\beta, Y_0, \theta, \sigma|p, y)) = \sum_{i=1}^{i=n} \log(P(y_i|p_i, \beta, Y_0, \theta, \sigma)) + C, \tag{23}$$

where C represents the sum of $\log(P(p_i|\beta, Y_0, \theta, \sigma))$, as a constant. Then, Equation (18) can be substituted into Equation (23) and we have:

$$\log(\mathcal{L}(\beta, Y_0, \theta, \sigma|p, y)) = \sum_{i=1}^{i=n} \log\left(\frac{1}{\sqrt{2\pi} * \sigma (Y_0 * p_i^\beta)^\theta} * e^{-\frac{(y_i - Y_0 * p_i^\beta)^2}{2\sigma^2 (Y_0 * p_i^\beta)^{2\theta}}}\right) + C. \tag{24}$$

Omitting detailed computing process, Equation (24) can be rewritten as:

$$\log(\mathcal{L}(\beta, Y_0, \theta, \sigma|p, y)) = -n * \log(\sqrt{2\pi} * \sigma) - \sum_{i=1}^{i=n} \log\left(\left(Y_0 * p_i^\beta\right)^\theta\right) - \frac{1}{2 * \sigma^2} * \sum_{i=1}^{i=n} \frac{(y_i - Y_0 * p_i^\beta)^2}{(Y_0 * p_i^\beta)^{2\theta}} + C. \tag{25}$$

Until here, the estimating parameters about urban scaling can be computed by maximizing Equation (25) or minimizing the opposite, which is different from the optimization function in the ordinary least square method. Furthermore, $y_i \leq 0$ is included in this method so that observations with $y_i \leq 0$ can be considered. However, it has been suggested that observations with $y_i < 0$ are invalid. Therefore, we are looking forward to a revised method with the lower bound.

4.2. Meeting the Lower Bound Constraints

As mentioned previously, for empirical urban data, urban scaling laws are obtained generally only for values of y_i above some lower bound y_{min} . Therefore, in this section, we build the maximum-likelihood fitting method with the lower bound, where the central question is how to integrate the lower bound into the method.

Firstly, the integral of the conditional probability $P(y|p)$, Equation (18), within the interval $y \in [y_{min}, +\infty]$ can be calculated as:

$$\int_{y_{min}}^{+\infty} P(y|p) dy = Q\left(\frac{y_{min} - \mu}{\text{var}(\varepsilon)}\right), \tag{26}$$

where Q denotes the Q function in statistic, which is the tail distribution function of the standard normal distribution, μ and $\text{var}(\varepsilon_i)$ are generated from Equations (15) and (16), respectively. Since we assume that the probability $P(y|p)$ holds zero for all $p < 0$, the probability $P(y|p)$ for $p \geq 0$ in Equation (18) is remained but transformed by multiplying by a standardized function, the reciprocal of Equation (26), to guarantee the entire function is integrated into 1 in the valid domain for urban data. Therefore, we get the conditional probability $P'(y_i|p_i)$ with lower bound as:

$$P'(y_i|p_i) = P(y_i|p_i) * \frac{1}{Q\left(\frac{y_{min} - \mu_i}{\text{var}(\varepsilon_i)}\right)}, y_i > 0. \tag{27}$$

Then, the likelihood function \mathcal{L}' can be described as:

$$\mathcal{L}'(\beta, Y_0, \theta, \sigma|p, y) = \prod_{i=1}^{i=n} P(y_i|p_i)' = \prod_{i=1}^{i=n} P(p_i, y_i|\beta, Y_0, \theta, \sigma) * \frac{1}{Q\left(\frac{y_{min} - \mu_i}{\text{var}(\varepsilon_i)}\right)}. \tag{28}$$

In order to maximize the likelihood function, Equation (28) can be log-transformed as:

$$\log(\mathcal{L}'(\beta, Y_0, \theta, \sigma|p, y)) = \sum_{i=1}^{i=n} \log(P'(p_i, y_i|\beta, Y_0, \theta, \sigma)). \tag{29}$$

Similar to the procedures in Section 4.1, the optimal function in this method can be computed as:

$$\begin{aligned} \log(\mathcal{L}'(\beta, Y_0, \theta, \sigma|p, y)) &= -n * \log(\sqrt{2\pi} * \sigma) - \sum_{i=1}^{i=n} \log\left(\left(Y_0 * p_i^\beta\right)^\theta\right) - \frac{1}{2 * \sigma^2} * \\ &\sum_{i=1}^{i=n} \frac{\left(y_i - Y_0 * p_i^\beta\right)^2}{\left(Y_0 * p_i^\beta\right)^{2\theta}} - \frac{1}{\sigma^2} * \sum_{i=1}^{i=n} Q\left(\frac{y_{min} - Y_0 * p_i^\beta}{\left(Y_0 * p_i^\beta\right)^{\frac{\theta}{2}}}\right) + C. \end{aligned} \tag{30}$$

However, it can be verified that there is a biased estimation in the method. In this method, the expectation of y_i with p_i can be computed as:

$$E'(y_i|p_i) = \int_{y_{min}}^{+\infty} P'(y|p) * y * dy. \tag{31}$$

Omitting detailed computing process, Equation (31) can be rewritten as:

$$E'(y_i|p_i) = \frac{var(\sigma_i)}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2 * var(\sigma_i)^2}} + var(\sigma_i) * Q\left(-\frac{\mu}{var(\sigma_i)}\right) \neq Y_0 * p^\beta, \tag{32}$$

where $\mu = Y_0 * p^\beta$. Aiming at the unbiased estimation, Equation (32) holds by choosing appropriate parameters μ' , which meet the following conditions:

$$E'(y_i|p_i) = \frac{var(\sigma_i)}{\sqrt{2\pi}} e^{-\frac{\mu'^2}{2 * var(\sigma_i)^2}} + var(\sigma_i) * Q\left(-\frac{\mu'}{var(\sigma_i)}\right) = Y_0 * p^\beta. \tag{33}$$

Apparently, there is no analytical solution for both parameters; however, numerical approaches can be achieved cursively in the optimization process.

Then, Equation (33) can be maximized using certain optimization method. In the optimization process, some constraints should be taken into consideration. We can see that the expectation of y with p must be located within the valid interval $[y_{min}, +\infty]$. That is, there is a constraint for the method:

$$E'(y_i|p_i) = Y_0 * p_i^\beta \geq y_{min}. \tag{34}$$

Thus far, the estimating parameters about urban scaling can be computed by maximizing $E'(y_i|p_i)$ in Equation (33), represented by CHVR. Furthermore, how to choose the lower bound y_{min} will be discussed. Empirically, the urban population has a certain lower bound according to the definition of the city from governments in different countries, for example, the lower limit for the population is 80,000 for China [54], 60,000 for Japan [54], and 50,000 for the USA [55]. To define metropolitan areas, the U.S. Census Bureau also utilizes the population size, density and commuting flow data [56]; however, there is no clear lower limits for urban attributes, such as GDP and wages. Some suggest that y_{min} should be estimated statistically to trade off data discarding against scaling performance, which lacks practical justification and meaning. In our method, y_{min} is set to zero. Moreover, estimators on the condition that y_{min} with other values can also be estimated utilizing our method.

4.3. Defining a Self-Starter Function

The initial value is important to the estimating process; thus, a self-starter function is defined to determine the initial value of the parameters in urban scaling laws. Since the log-transformed linear regression method computes the geometric mean of the estimators simply without cursive iterations, the results from the LOG_LR method are utilized as the self-starter function in our method.

5. Method Evaluation Criteria

The goodness-of-fit of the fitting method for urban scaling law estimators is usually quantified by the coefficient of determination R^2 , which can be computed as:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}, \quad (35)$$

where SS_{res} and SS_{tot} denote the sum of squares of residuals and the total sums of squares, respectively. However, the coefficient of determination describes the percentage of the response variable explained by the urban scaling method, which evaluates the statistical significance of the method from one perspective. In addition to R^2 , cross-validation, distance based on Monte Carlo simulation, and the Breusch–Pagan test are also adopted to assess the performance of the proposed CHVR method and provide criteria for compared with two traditional methods, LOG_LR and the NLR method.

For the second evaluation method, cross-validation is performed to evaluate whether the methods can be extended to more generalized datasets [57] avoiding overfitting and selection bias [58]. The specific method of cross-validation is that 80% of the data set was chosen randomly as the training set and the other 20% as the test set; then, the mean squared error (MSE) MSE_i is computed to assess the fitting, and this process has been iterated cursively for m times so that the index of the cross-validation for the method, termed as Cross-Validation Index (CVI) can be quantified as:

$$CVI = \frac{1}{m} * \sum_{i=1}^{i=m} MSE_i. \quad (36)$$

The cross-validation evaluation method implements the basic assessment out of the limit of independent variables.

The third evaluation method, the Monte Carlo simulation method, is utilized to evaluate the performance of the regression method achieving comparing the simulated and the original data directly [29]. In detail, the target method, such as the CHVR method, is first computed to evaluate the scaling parameters of the empirical dataset. Then, the scaling parameter results can be written as parameter list A. Then $n = 10,000$ synthetic datasets of independent value can be derived from the conditional probability distribution $P'(y_i|p_i)$ in the method with the fixed parameter sets A via Monte Carlo simulation. Lastly, the average Euclidean distance between the synthetic data sets and empirical data is computed as the goodness-of-fit for the method, termed as the Monte Carlo Index (MCI). The Monte Carlo simulation-based evaluation method implements the basic assessment out of the limit of dependent variables.

The last evaluation method, the Breusch–Pagan test (BP test), is performed to measure heterogeneity of the method. BP test was first put forward by Breusch and Pagan in [49]. In addition to the graphical test for heterogeneity, it can also evaluate if the variance of the residuals of the data is dependent on the variables and provide a valid statistical method. Suppose that we estimate the regression method:

$$y = f(x, \beta), \quad (37)$$

where y represents the dependent variable related to the independent variable through a function f (urban scaling law method), and β denotes the parameter sets. Traditional methods assume that the variance is independent of the variables, termed as the homogeneity assumption. The violation of homogeneity may be referred to as that the variance has a linear relationship with the independent variables, that is:

$$\sigma^2 = \gamma_0 + \gamma_1 x, \quad (38)$$

where σ^2 represents the variance of the disturbance while γ_0 and γ_1 denote the coefficients of this auxiliary regression. Then the null hypothesis of homogeneity can be written as:

$$H_0 : \gamma_0 = \gamma_1 = 0. \quad (39)$$

Two different statistical tests are usually conducted in the BP test [11]. Direct F-test assesses the hypothesis that the error variance does not depend on the independent variables. Indirect Chi-square test is conducted on the auxiliary Lagrange Multiplier (LM) statistic. As the LM test overstates the impact of small or relatively large samples [59], the F-statistic is preferred in our evaluations.

6. Method Evaluations and Results

Method evaluations were performed to compare the performance of the LOG_LR, NLR, and CHVR methods. Three comparing methods are applied to estimating the urban scaling exponents β of six urban attributes in China, including GDP, total income, household water consumption, household electricity consumption, road surface area, and built-up area. The LOG_LR, NLR, and CHVR methods have been compared based on evaluation criteria in Section 5. Furthermore, the standard errors of the estimated β were estimated using bootstrapping with replacement [1]. The summary of the urban scaling exponents, the corresponding standard errors, and method evaluations for these six urban attributes, using the demographic data with high accuracy provided by the census, are demonstrated in Table 1, which were derived from the CHVR method and the traditional methods in China in 2010.

Table 1. Summary of the urban scaling exponents and method evaluation to three methods.

Urban Attributes	Methods	β	SE	θ	R ²	CVI	MCI
GDP	LOG_LR	1.2162	0.0345	/	0.7814	3.4616×10^{13}	2.1376×10^8
	NLR	1.1580	0.0591	/	0.9099	3.5702×10^{13}	2.0037×10^8
	CHVR	1.2022	0.0300	0.6813	0.9342	3.3173 $\times 10^{13}$	1.2806 $\times 10^8$
Total income	LOG_LR	1.1809	0.0397	/	0.7367	3.1409×10^{12}	3.3998×10^7
	NLR	1.4063	0.2401	/	0.8145	3.6023×10^{12}	4.7743×10^7
	CHVR	1.2032	0.0395	0.8737	0.9625	2.4320 $\times 10^{12}$	3.0365 $\times 10^7$
Household water consumption	LOG_LR	1.1278	0.0377	/	0.7283	2.2582×10^7	1.2999×10^5
	NLR	1.1121	0.0727	/	0.8389	2.2155×10^7	1.5787×10^5
	CHVR	1.1131	0.0332	0.8561	0.9346	2.0520 $\times 10^7$	1.0287 $\times 10^5$
Household electricity consumption	LOG_LR	1.1493	0.0331	/	0.8278	2.4318×10^9	1.6676×10^6
	NLR	1.1130	0.0518	/	0.9371	2.6383×10^9	1.6592×10^6
	CHVR	1.1484	0.0310	0.7539	0.9574	2.2693 $\times 10^9$	1.0758 $\times 10^6$
Road surface area	LOG_LR	0.9862	0.0287	/	0.7034	8.9181×10^5	3.0466×10^4
	NLR	0.7984	0.0430	/	0.8155	6.7414 $\times 10^5$	2.8398×10^4
	CHVR	0.9417	0.0236	0.7727	0.8573	8.8028×10^5	2.0146 $\times 10^4$
Built-up area	LOG_LR	0.8790	0.0287	/	0.7544	3.2347×10^3	1.5083×10^3
	NLR	0.8758	0.0544	/	0.8790	3.4481×10^3	1.9226×10^3
	CHVR	0.8900	0.0273	0.7840	0.8882	2.9697 $\times 10^3$	1.2085 $\times 10^3$

Note: SE denotes the standard errors of the estimated β . Each evaluation index marked with bold indicates the best performance among three methods for the urban attribute according to the evaluation criterion.

6.1. Method Evaluation Results of the LOG_LR, NLR, and CHVR Methods

The CHVR method presents significant improvements over both LOG_LR and NLR methods in terms of the coefficient of determination R², the Cross-Validation Index (CVI), and the Monte Carlo Index (MCI).

(1) The coefficient of determination R²: The value of R² in the CHVR method is the highest, and that in the NLR method is the second highest among these three methods for six urban attributes. Taking the total income as an example, almost 96.25% of the response variable is explained by the CHVR method, compared with 73.67% explained by the LOG_LR method, and 81.45% explained by the NLR method. It can be concluded that the CHVR method shows a significant improvement over traditional methods in terms of R².

(2) The Cross-Validation Index (CVI): The value of CVI in the CHVR method is much lower than the respective values for the LOG_LR and NLR for urban attributes, clearly indicating that the CHVR method performs better than the other two methods. For urban attributes (namely, GDP, total income, household electricity consumption, and built-up area), it can be observed that CVI in the

CHVR is lower than that in the LOG_LR method, and CVI in the NLR method is the highest. For household water consumption, it can be observed that that CVI in the CHVR is lower than that in the NLR method, and CVI in the LOG_LR method is the highest. These comparisons demonstrate that the CHVR method outperforms both LOG_LR and NLR methods in the cross-validation test. Road surface area is a special case, where CVI in the CHVR method is higher than CVI in the NLR method; however, the proposed method still performs better than the LOG_LR method. The selection bias in the dimension of the dependent variable and imprecisions in data gathering may be possible reasons for this exception, which will be further examined.

(3) The Monte Carlo Index (MCI): The value of MCI of the CHVR method is much lower than the respective values of the LOG_LR and NLR methods for urban attributes, clearly indicating that the CHVR method performs better than the other two methods. For urban attributes, namely, GDP, household electricity consumption, and road surface area, $MCI_{CHVR} < MCI_{NLR} < MCI_{LOG_LR}$, demonstrating that the CHVR method performs the best, and the NLR method performs second best among the three methods. For urban attributes, namely, total income, household water consumption, and built-up area, it can be observed that $MCI_{CHVR} < MCI_{LOG_LR} < MCI_{NLR}$, showing that the CHVR method performs the best, and the LOG_LR method performs second best among the three methods. It can be concluded that the CHVR method can derive urban scaling exponents with higher accuracy in terms of MCI out of limits of dependent variables.

In summary, the results obtained using three kinds of comparison criteria verify the outperformance of the CHVR method in terms of a basic statistical assessment, an assessment based on random independent variables, and dependent variables. The effectiveness of the CHVR method is mainly attributed to the variance function with heterogeneity describing mixed effects accurately and the consideration of the lower bound constraints.

6.2. Heterogeneity Results of the LOG_LR, NLR, and CHVR Methods

To further explore the effectiveness of the variance function describing heterogeneity in the CHVR method, a visual inspection of residuals and the BP test were utilized to examine how heterogeneity varies between the traditional methods and the CHVR method.

The plot of the log-transformed fitted values versus residuals of urban attributes was utilized to compare heterogeneity in three methods qualitatively. Taking GDP as an example, as shown in Figure 4, the plots of the log-transformed fitted values versus residuals of the GDP for both LOG_LR and CHVR methods show that the residuals are independent of the fitted value, indicating homogeneity. However, the shape of a horn for the NLR method indicates heterogeneity, as described in Section 3.4.

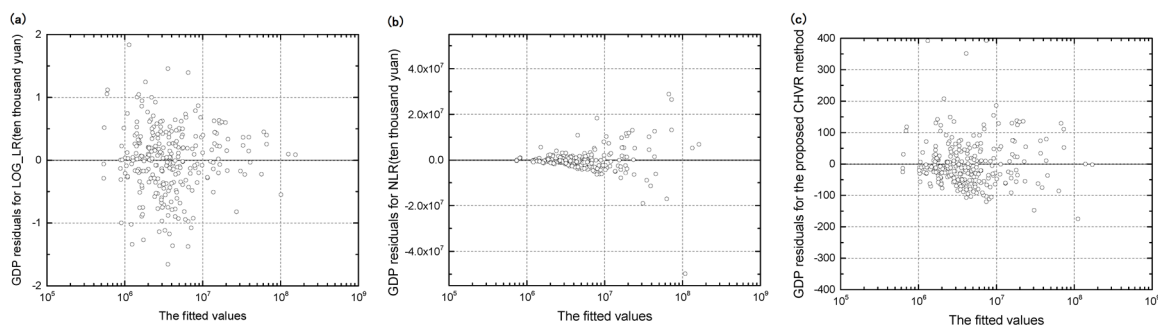


Figure 4. Plot of the log-transformed fitted values versus residuals of the GDP for (a) the LOG_LR method, (b) the NLR method, and (c) the CHVR method.

Furthermore, the BP test was implemented to test the heterogeneity between the LOG_LR, NLR, and CHVR methods quantitatively. The F-statistics values and their corresponding p -values are listed in Table 2. Sequential Bonferroni [60] was utilized to control the experiment-wise Type I error rate. The Sequential Bonferroni method sorts the p -values in ascending order and compares them to

nominal alpha levels of $\frac{\alpha}{m}$ to α , where α denotes the desired overall alpha level and m is the number of hypotheses. In detail, as for GDP, if $\alpha = 0.05$ is assumed as the overall alpha level, the p -values for the F-statistics in three methods are firstly sorted from small to large, and then, homogeneity hypothesis is tested at a significance level of $\frac{0.05}{3}$ for the method with the smallest p -value, $\frac{0.05}{2}$ for the method with medium p -value, 0.05 for the biggest. If the p -value for the F-statistics is greater than the testing significance level, the null hypothesis can be accepted so that homogeneity appeared in this method according to the BP test. Otherwise, heterogeneity can be concluded and marked with an asterisk ‘*’.

Table 2. Summary of F-statistics results in the BP test for the LOG_LR, NLR, and CHVR method for six urban attributes.

Parameter	LOG_LR		NLR		CHVR	
	F-Value	p -Value	F-Value	p -Value	F-Value	p -Value
GDP	3.3241	0.0694	48.9939	1.99×10^{-11} *	0.6134	0.4341
Total income	0.0237	0.8778	35.2855	8.64×10^{-9} *	3.3044	0.0702
Water consumption	6.5394	0.0111*	40.9598	6.74×10^{-10} *	0.7598	0.3842
Electricity consumption	5.6067	0.0186*	46.8604	5.15×10^{-11} *	0.0579	0.8101
Road surface area	2.2672	0.1333	91.3284	7.44×10^{-19} *	0.2567	0.6128
Built-up area	0.3846	0.5357	81.3017	3.51×10^{-17} *	0.1677	0.6825

Note: Water consumption and electricity consumption denote household water consumption, and household electricity consumption, respectively. Those p -values less than the testing significance level are marked with an asterisk ‘*’ indicating heterogeneity.

It is noteworthy that the homogeneity assumption was accepted for the CHVR method for all six urban attributes. In contrast, the homogeneity assumption was rejected for the NLR method for all six urban attributes; in addition, as for the LOG_LR method, the homogeneity assumption was accepted for four urban attributes (GDP, total income, road surface area, and built-up area), but not for household water and electricity consumption. Specifically, significant heterogeneity is observed in the NLR method, which reduces the accuracy of exponent estimation. In the LOG_LR method, an underlying heterogeneity assumption in the linear regression method on the logarithmic scale can account for homogeneous results for four urban attributes. Considering that heterogeneity is not described by a specific variance model, but rather by the underlying result of log-transformation, heterogeneity is not characterized effectively for household consumption attributes; thus, the homogeneity assumption is rejected. Furthermore, it is assumed that the larger the p -value is for the F-statistics that corresponds to the null hypothesis of homogeneity, the more homogeneous the residuals of the model. The highest p -value in the CHVR method for five urban attributes other than the total income shows that significant heterogeneity has been eliminated through the variance function in the CHVR method. As to the total income, a lower p -value in the CHVR method compared with that of the LOG_LR method may be ascribed to the inappropriate description of heterogeneity, while the NLR method is more appropriate. In essence, the heterogeneity can be appropriately described in our framework.

The above suggests that the variance function in the CHVR method can describe heterogeneity adequately and, compared with traditional estimation methods, results in a more efficient estimation of urban scaling exponents.

7. Discussions

In this section, scaling functions for six urban attributes based on the CHVR method are analyzed to explore the urban development mechanisms in China; then, exponents among three methods are discussed regarding how the estimated exponent depends on the method. In addition, interpretation of the parameter in the variance model is explored from a dynamic perspective to provide extra urban information.

7.1. Urban Scaling Law in CHINA Based on the CHVR Method

The similarities and differences of urban scaling regimes for six attributes in China compared with developed countries quantitatively revealed the urban development trends in China.

The exponents of urban scaling for some of the urban attributes in China fall into the same scaling regimes as those of other countries in the previous studies [3,61], including super-linear scaling for socioeconomic attributes, linear scaling for household-related attributes and sub-linear scaling for infrastructure-related attributes. However, there are also some unique phenomena of exponents of the urban scaling laws in China as shown in Table 1. As to socioeconomic attributes, the GDP scaling exponent of China is 1.2022, representing super-linear scaling and indicating the urban agglomeration effect in the economy in China. Furthermore, the exponent is greater than the expected value of $\frac{7}{6}$ proposed by Bettencourt [3] and 1.11 in the United States [3]. This finding indicates that urban population agglomeration contributed more to China's economic growth in 2010. A similar phenomenon has also been explored for the total income with $\beta_{CHVR} = 1.2032$. Considering household-related attributes, both the household water consumption and the electricity consumption fall into the super-linear regime instead of the expected linear regime, with exponents of 1.1131 and 1.1484, respectively. This finding may be attributed to lower energy consumption efficiency characterizing the stage of extensive economic growth in China in 2010. As to infrastructure-related attributes, the built-up area's scaling exponent in China is 0.8790 and greater than the theoretically expected value of $\frac{5}{6}$ and 0.6~0.8 in developed countries [3]. The comparative exponents results indicate that the built-up areas in China display weaker scale effects than developed countries and the optimized situation proposed by Bettencourt [3]. Thus, this finding indicates lower built-up land use efficiency compared with developed countries and an unbalanced relationship between land urbanization and population urbanization in China in 2010. The policies for intensive land use and optimization of land use structure could be a better solution in the process of urbanization in China, consistent with "economical and intensive land use" policies in China proposed in 2008 and developed in 2012 [15,16]. The scaling exponent of the road surface area is 0.9471, far greater than the theoretical value of $\frac{5}{6}$ [3] and even closer to 1. This exponent indicates a nearly linear scaling relationship between road surface area and population size. Some may argue that this relationship is ambiguous between linear and sub-linear scaling. However, it can be observed that this ambiguous scaling relationship does not reflect the economies of scale for the road surface area in previous studies in developed countries [19]. The observed exponent suggests that the total scale of road infrastructure construction has become large enough, but the efficiency is not optimal, and there is a lack of cost advantages of the economies of scale. This may result from the unbalanced relationship between road infrastructure supply (e.g., road network and transport infrastructure structure) and demand (e.g., urban morphology and land use pattern).

7.2. Dependence of the Estimated Exponent on the Method

How the scaling exponent depends on the estimation method is discussed from both the urban attributes dimension and the estimation method dimension.

From the urban attributes dimension, the dependence of the exponent on the estimation method exists in different urban attributes. There are different magnitudes of variations observed for different urban attributes under study. Exponents of household water consumption and the built-up area show less dependence on the estimation methods, while exponents of other urban attributes (GDP, total income, household electricity consumption, road surface area) fluctuate widely for different estimation methods, as shown in Figure 5. For instance, the scaling exponent of the total income in the NLR method is much larger than the exponent in the CHVR and LOG_LR method. Therefore, different estimation methods will result in different scaling regimes for these urban attributes. Considering the road surface area, the scaling exponents were less than 1 in the CHVR and NLR method, indicating a sub-linear scaling regime for this urban attribute and corroborating the expected scaling regime [3]. However, the scaling exponents obtained from the LOG_LR were approximately 1, indicating a nearly

linear instead of sub-linear scaling relationship. The different urban scaling laws of the road surface area in the three methods are illustrated in Figure 6. Furthermore, different estimation methods will result in different numerical relationships with theoretically expected values of these urban attributes. In particular, for the total income, the estimated exponents in the LOG_LR method and the CHVR method were close to the theoretical value $\beta = \frac{7}{6}$; however, the exponent in the NLR method was far above the expected value. The dependences on the estimation method for different attributes have important theoretical and empirical implication for urban science. The universality of urban scaling law has been doubted due to inconsistent results obtained for different countries [13,21] compared with theoretical and empirical regimes reported by Bettencourt [2,3]. However, this discrepancy may occur affected by the urban scaling estimation method. For example, the scaling exponent of the road surface area obtained using the CHVR method complies with urban scaling law, even though there is a paradox for the LOG_LR method. In addition, the urban scaling exponent β is affected by not only the estimation method but also the urban definition, i.e., the data aggregation scale. Therefore, urban scaling regimes cannot be obtained independently from the analysis of the variance of the error model in the estimation method.

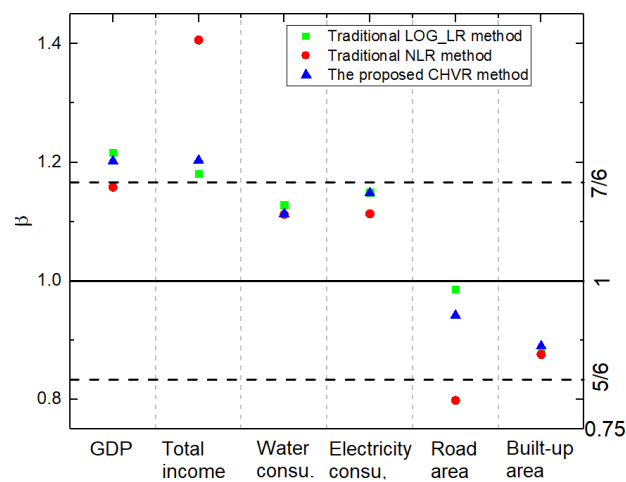


Figure 5. Comparison of results for urban scaling exponents for the LOG_LR, NLR, and CHVR method in China in 2010. The green squares, red circles, and blue triangles denote the scaling exponents calculated from the LOG_LR, NLR, and CHVR methods, respectively. The dotted lines refer to the theoretically expected value [3] $\beta = \frac{7}{6}$ for social currency-related attributes (the super-linear regime), $\beta = 1$ for human needs-related attributes (the linear regime), and $\beta = \frac{5}{6}$ for infrastructure-related attributes (the sub-linear regime), respectively.

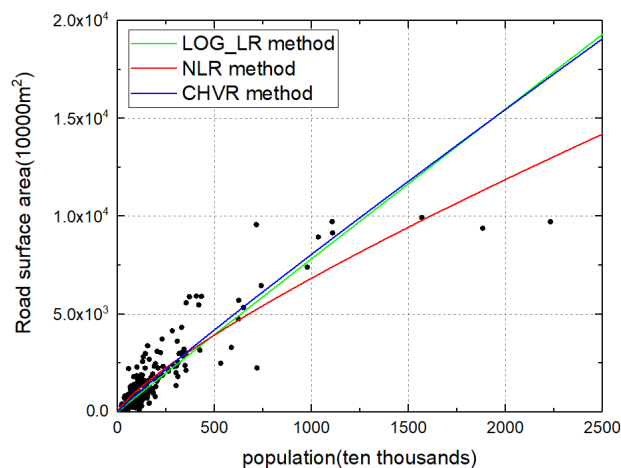


Figure 6. Different urban scaling laws of road surface area from the LOG_LR, NLR, and CHVR methods.

From the estimation method dimension, the exponent from the CHVR method locates in the middle of the LOG_LR method and NLR method, indicating that the variance of error model in the proposed method is analogous to a weighted model of the variance in LOG_LR and NLR methods. It can be observed that there is certain comparative relationship among three methods, i.e., the exponent obtained using the CHVR method is larger than that of one traditional method and smaller than that of the other. As shown in Figure 5, blue dots are located in the middle of red and green dots for all urban attributes apart from the built-up area, where the exponents from three methods were nearly consistent. The LOG_LR method assumes the lognormally distributed and multiplicative error by default, and the NLR method assumes the normally distributed and additive error by default. In our method, properties of variance have been explored and derived as the power-of-the-mean variance model. This result indicates that our method is analogous to a weighted model of LOG_LR and NLR methods. In other words, the CHVR method is not an ordinary weighted model of two traditional models. This model can be viewed as an intermediate state between the LOG_LR and NLR models. It is analogous to that of a small-world network being an intermediate state between the rule network and the random network [62] in the complex network field. Similarly to the small-world network, the CHVR method is based on a heterogeneous model of variance in real-world urban systems, where the error is not purely multiplicative or additive, and variances for different residuals are not purely dependent or independent. Thus, our method can effectively describe not only the combination of different effects, such as the variation of human activity and the fluctuation of data, but also the intermediate state originating from the real-world urban complex system, such as the error model and the relationship between variances for different residuals.

7.3. Urban Interpretations of θ from a Dynamic Perspective

In this section, the interpretation of the parameter θ in the variance function of the CHVR method is explored to discover extra urban information beneath the heterogeneity in urban scaling law. The parameter θ was calculated in the long term for all six urban attributes. Specifically, as shown in Figure 7, the values of θ were derived by the CHVR method for all six urban attributes for the year 2000, 2005, 2010, and 2015 considering the urban development process in China and data limits described in Section 3. The variation of the parameter θ is examined to understand the meaning of this parameter in urban systems.

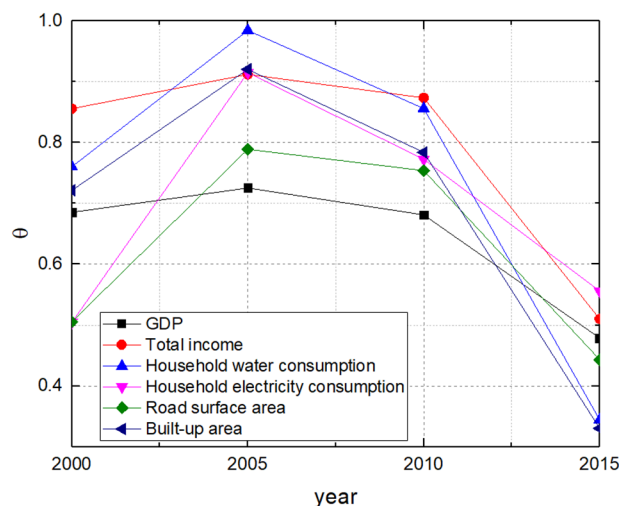


Figure 7. The values of θ are derived by the CHVR method for all six urban attributes for the year 2000, 2005, 2010, and 2015.

Regarding all six attributes belonging to three scaling regimes, the value of the parameter θ in 2005 was higher than that in 2000, and then it decreased continuously from 2005 to 2015, as shown in Figure 7.

For instance, for GDP, the value of parameter θ increased from $\theta = 0.6855$ in 2000 to $\theta = 0.7255$ in 2005, when the value reached the maximum; then, the value of parameter θ declined to $\theta_{2010} = 0.6813$ in 2010 and $\theta_{2015} = 0.4788$ in 2015. The peak in 2005 shows that there was a transition point around 2005 in the process of urban development in China so that the urbanization in China can be categorized into two stages according to the value of θ . This result is consistent with the findings by Chen [63], who discovered that urbanization velocity attained a peak in the year of 2004 or 2005 bringing two stages in the process of China's urbanization, as well as the statistical results from China Statistical Yearbook, where the peak arrives at around 2003. Statistically, the high value of the parameter θ means overdispersion within the data. It was suggested by West and Bettencourt [64] that the greater degree of dispersion in urban systems, which is in the self-organized evolution process, represents the shorter time for the urban systems to move toward the ideal optimal configuration represented by the scaling law. Therefore, the continuous reduction of θ since 2005 shows less dispersion over time after 2005, and further implies that urban system in China moves toward the ideal optimal configuration represented by the scaling law unceasingly from 2005 to 2015. In addition, the unexpected growth of θ from 2000 to 2005 may be attributed to that self-organized behaviors appear, so that an urban scaling effect is not apparent.

The variation of the parameter θ indicates that the exponent θ in urban systems can be interpreted as not only the degree of overdispersion statistically but also the index of matureness of the self-organized urban systems toward the ideal configuration represented by the scaling law, which is related with urbanization velocity, inequality between large cities and small cities, urban efficiency, and spatial synchronization. The interpretation of the variance function in the form of Taylor's Law has remained of interest over time, with particular effort directed to understanding the exponent θ in different complex systems [37]. The meaning of θ has been interpreted to different meanings depending on the system studied, synchronization in transportation system [65], randomness and aggregation in ecology systems [51], and population growth in human geography systems [36].

8. Conclusions

Urban scaling law is the key to establishing a predictive and quantitative theory for urban organization and sustainable development. Although urban scaling law has been explored theoretically and empirically in previous work, accurate scaling exponent estimation is still challenging. Limitations exist in the traditional methods, and the variance structure need be explored further.

In this paper, a maximum likelihood-based statistical framework named CHVR is proposed to estimate the urban scaling exponents considering both the lower bound constraints and the error structure of heterogeneity. The CHVR method has been applied to estimating the scaling exponents for six urban attributes in China. Method evaluations have been performed utilizing the coefficient of determination, the cross-validation index, and Monte Carlo-based index. Evaluation results show that all three evaluation indexes for the CHVR method yield improvement over the LOG_LR and NLR methods for all six urban attributes (e.g., GDP, total income, household water consumption, household electricity consumption, road surface area, and built-up area). The properties of variance heterogeneity were explored through the Breusch–Pagan test, demonstrating that the variance function in the CHVR method describes heterogeneity adequately and results in a more effective estimation of urban scaling exponents than do traditional estimation methods. The variation of parameter θ in the long term shows that exponent θ implies not only the overdispersion degree in statistics but also an index of maturity for self-organized urban systems to the optimal configuration represented by the scaling law. It can be concluded that the CHVR method is more effective and robust than traditional scaling estimation methods and has great potential in analyzing urban development mechanisms.

Future research is recommended in the following areas. Firstly, other possible forms of methods considering heterogeneous variance can be explored. Secondly, the CHVR method can be applied to more countries, urban attributes, and periods of time. Thirdly, the variation of θ in finer granularity

time can be explored to further unearth the interpretation of θ . Finally, this study has been conducted at the municipal district scale in China, and empirical studies can be applied to more spatial scales.

Author Contributions: W.W. and H.Z. conceived and designed the experiments; W.W. and Q.T. performed the experiments and analyzed the data, W.W., H.Z., and P.G. wrote and revised the paper.

Acknowledgments: This research was supported by the Fund of National Natural Science Foundation of China (grant number 41571414) and the Tsinghua University Initiative Scientific Research Program (grant number 2015THZ01). We thank anonymous reviewers for their constructive comments, which greatly improved this article. We also thank Xu Ma at Tsinghua University for helpful discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Batty, M. A theory of city size. *Science* **2013**, *340*, 1418–1419. [[CrossRef](#)]
- Bettencourt, L.M.; Lobo, J.; Helbing, D.; Kühnert, C.; West, G.B. Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7301–7306. [[CrossRef](#)]
- Bettencourt, L.M. The origins of scaling in cities. *Science* **2013**, *340*, 1438–1441. [[CrossRef](#)]
- Pumain, D.; Paulus, F.; Vacchiani-Marcuzzo, C.; Lobo, J. An evolutionary theory for interpreting urban scaling laws. *Cybergeo Eur. J. Geogr.* **2006**. [[CrossRef](#)]
- Chen, Y.G. and Liu, M.H. Study on fractal dimension of size distribution of cities. *Bull. Sci. Technol.* **1998**, *14*, 395–400.
- Nordbeck, S. Urban allometric growth. *Geogr. Ann. Ser. B Hum. Geogr.* **1971**, *53*, 54–67. [[CrossRef](#)]
- Kleiber, M. Body size and metabolism. *Hilgardia* **1932**, *6*, 315–353. [[CrossRef](#)]
- West, G.B.; Woodruff, W.H.; Brown, J.H. Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2473–2478. [[CrossRef](#)] [[PubMed](#)]
- Batty, M. *The New Science of Cities*; The MIT Press: Cambridge, MA, USA, 2013; pp. 123–126.
- Bettencourt, L.; West, G. A unified theory of urban living. *Nature* **2010**, *467*, 912–913. [[CrossRef](#)]
- Bettencourt, L.M.; Lobo, J.; Strumsky, D.; West, G.B. Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLoS ONE* **2010**, *5*, e13541. [[CrossRef](#)] [[PubMed](#)]
- Louf, R.; Barthélemy, M. How congestion shapes cities: From mobility patterns to scaling. *Sci. Rep.* **2014**, *4*, 5561. [[CrossRef](#)] [[PubMed](#)]
- Arcaute, E.; Hatna, E.; Ferguson, P.; Youn, H.; Johansson, A.; Batty, M. Constructing cities, deconstructing scaling laws. *J. R. Soc. Interface* **2015**, *12*. [[CrossRef](#)] [[PubMed](#)]
- Alves, L.G.; Ribeiro, H.V.; Lenzi, E.K.; Mendes, R.S. Distance to the scaling law: A useful approach for unveiling relationships between crime and urban metrics. *PLoS ONE* **2013**, *8*, e69580. [[CrossRef](#)]
- Chen, Y. Characterizing growth and form of fractal cities with allometric scaling exponents. *Discret. Dyn. Nat. Soc.* **2010**, *2010*, 194715. [[CrossRef](#)]
- Li, R.; Dong, L.; Zhang, J.; Wang, X.; Wang, W.-X.; Di, Z.; Stanley, H.E. Simple spatial scaling rules behind complex cities. *Nat. Commun.* **2017**, *8*, 1841. [[CrossRef](#)]
- Pumain, D.; Rozenblat, C. Two metropolisation gradients in the European system of cities revealed by scaling laws. *Environ. Plan. B Urban Anal. City Sci.* **2018**. [[CrossRef](#)]
- Arbabi, H.; Mayfield, M.; Dabinett, G. Urban performance at different boundaries in England and Wales through the settlement scaling theory. *Reg. Stud.* **2018**. [[CrossRef](#)]
- Louf, R. Wandering in cities: A statistical physics approach to urban theory. *arXiv* **2015**, arXiv:1511.08236.
- Oliveira, E.A.; Andrade, J.S., Jr.; Makse, H.A. Large cities are less green. *Sci. Rep.* **2014**, *4*, 4235. [[CrossRef](#)]
- Cottineau, C.; Hatna, E.; Arcaute, E.; Batty, M. Diverse cities or the systematic paradox of urban scaling laws. *Comput. Environ. Urban Syst.* **2017**, *63*, 80–94. [[CrossRef](#)]
- Dong, L.; Wang, H.; Zhao, H. The definition of city boundary and scaling law. *Acta Geogr. Sin.* **2017**, *72*, 213–223.
- Finance, O.; Cottineau, C. Are the absent always wrong? Dealing with zero values in urban scaling. *Environ. Plan. B Urban Anal. City Sci.* **2018**. [[CrossRef](#)]
- Stumpf, M.P.; Porter, M.A. Critical truths about power laws. *Science* **2012**, *335*, 665–666. [[CrossRef](#)] [[PubMed](#)]
- Clauset, A.; Shalizi, C.R.; Newman, M.E.J. Power-law distributions in empirical data. *Siam Rev.* **2009**, *51*, 661–703. [[CrossRef](#)]

26. Bettencourt, L.M.; Lobo, J. Urban scaling in Europe. *J. R. Soc. Interface* **2016**, *13*. [[CrossRef](#)] [[PubMed](#)]
27. Packard, G.C. On the use of logarithmic transformations in allometric analyses. *J. Theor. Biol.* **2009**, *257*, 515–518. [[CrossRef](#)]
28. Packard, G.C.; Birchard, G.F.; Boardman, T.J. Fitting statistical models in bivariate allometry. *Biol. Rev.* **2011**, *86*, 549–563. [[CrossRef](#)]
29. Xiao, X.; White, E.P.; Hooten, M.B.; Durham, S.L. On the use of log-transformation vs. Nonlinear regression for analyzing biological power laws. *Ecology* **2011**, *92*, 1887–1894. [[CrossRef](#)]
30. Packard, G.C.; Birchard, G.F. Traditional allometric analysis fails to provide a valid predictive model for mammalian metabolic rates. *J. Exp. Biol.* **2008**, *211*, 3581–3587. [[CrossRef](#)] [[PubMed](#)]
31. Li, X.; Chen, G.; Xu, X. Urban allometric growth in China: Theory and facts. *Acta Geogr. Sin.* **2009**, *64*, 399–407.
32. Alain, F.; Zuur, E.N.L. *Mixed Effects Models and Extensions in Ecology with R*; Springer: Berlin/Heidelberg, Germany, 2009.
33. Leitão, J.C.; Miotto, J.M.; Gerlach, M.; Altmann, E.G. Is this scaling nonlinear? *R. Soc. Open Sci.* **2016**, *3*, 150649. [[CrossRef](#)]
34. Eisler, Z.; Bartos, I.; Kertész, J. Fluctuation scaling in complex systems: Taylor’s law and beyond. *Adv. Phys.* **2008**, *57*, 89–142. [[CrossRef](#)]
35. Gao, P.; Liu, Z.; Tian, K.; Liu, G. Characterizing traffic conditions from the perspective of spatial-temporal heterogeneity. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 34. [[CrossRef](#)]
36. Greig, A.; Dewhurst, J.; Horner, M. An application of Taylor’s Power Law to measure overdispersion of the unemployed in English labor markets. *Geogr. Anal.* **2015**, *47*, 121–133. [[CrossRef](#)]
37. Hanley, Q.S.; Khatun, S.; Yosef, A.; Dyer, R.-M. Fluctuation scaling, Taylor’s law, and crime. *PLoS ONE* **2014**, *9*, e109004. [[CrossRef](#)] [[PubMed](#)]
38. Nomaler, Ö.; Frenken, K.; Heimeriks, G. On scaling of scientific knowledge production in US metropolitan areas. *PLoS ONE* **2014**, *9*, e110805. [[CrossRef](#)] [[PubMed](#)]
39. UN. *World Urbanization Prospects: The 2014 Revision*; UN: New York, NY, USA, 2014.
40. Deng, X.; Huang, J.; Rozelle, S.; Uchida, E. Growth, population and industrialization, and urban land expansion of China. *J. Urban Econ.* **2008**, *63*, 96–115. [[CrossRef](#)]
41. Ye, X.; Xie, Y. Re-examination of Zipf’s law and urban dynamic in China: A regional approach. *Ann. Reg. Sci.* **2012**, *49*, 135–156. [[CrossRef](#)]
42. Chen, Y.; Zhou, Y. Scaling laws and indications of self-organized criticality in urban systems. *Chaos Solitons Fractals* **2008**, *35*, 85–98. [[CrossRef](#)]
43. Ramaswami, A.; Jiang, D.; Tong, K.; Zhao, J. Impact of the economic structure of cities on urban scaling factors: Implications for urban material and energy flows in China. *J. Ind. Ecol.* **2018**, *22*, 392–405. [[CrossRef](#)]
44. West, G.B.; Brown, J.H.; Enquist, B.J. A general model for the origin of allometric scaling laws in biology. *Science* **1997**, *276*, 122–126. [[CrossRef](#)]
45. Pumain, D.; Paulus, F.; Vacchiani-Marcuzzo, C. Innovation cycles and urban dynamics. In *Complexity Perspectives in Innovation and Social Change*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 237–260.
46. Cristelli, M.; Batty, M.; Pietronero, L. There is more than a power law in Zipf. *Sci. Rep.* **2012**, *2*, 812. [[CrossRef](#)]
47. Ritz, C.; Streibig, J.C. *Nonlinear Regression with R*; Springer: Berlin/Heidelberg, Germany, 2008; p. 473.
48. Gastwirth, J.L.; Gel, Y.R.; Miao, W. The impact of Levene’s test of equality of variances on statistical theory and practice. *Stat. Sci.* **2009**, *24*, 343–360. [[CrossRef](#)]
49. Breusch, T.S.; Pagan, A.R. A simple test for heteroscedasticity and random coefficient variation. *Econometrica* **1979**, *47*, 1287–1294. [[CrossRef](#)]
50. Pinheiro, J.C.; Bates, D.M. *Mixed-Effects Models in S and S-Plus, corrected third printing*; Springer: Berlin/Heidelberg, Germany, 2000.
51. Taylor, L. Aggregation, variance and the mean. *Nature* **1961**, *189*, 732–735. [[CrossRef](#)]
52. Chen, Y.; Jiang, B. Hierarchical scaling in systems of natural cities. *Entropy* **2018**, *20*, 432. [[CrossRef](#)]
53. Chen, Y. The mathematical relationship between Zipf’s law and the hierarchical scaling law. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 3285–3299. [[CrossRef](#)]
54. Wen, H.; Wei, D. *Urban Economics*; Tsinghua University Press: Beijing, China, 2008.
55. See e.G. Available online: [https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/fedreg_T1\textbar{}2010/06282010_metro_standards-complete.Pdf](https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/fedreg/T1\textbar{}2010/06282010_metro_standards-complete.Pdf) (accessed on 17 November 2018).

56. See e.G. Available online: <http://www.Census.Gov/population/metro/> (accessed on 17 November 2018).
57. Kohavi, R. In A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the IJCAI, Montreal, QC, Canada, 20–25 August 1995; pp. 1137–1145.
58. Cawley, G.C.; Talbot, N.L. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.
59. Greene, W.H. *Econometric Analysis*; Pearson Education India: Noida, India, 2003.
60. Rice, W.R. Analyzing tables of statistical tests. *Evolution* **1989**, *43*, 223–225. [[CrossRef](#)] [[PubMed](#)]
61. Gomez-Lievano, A.; Patterson-Lomba, O.; Hausmann, R. Explaining the prevalence, scaling and variance of urban phenomena. *Nat. Hum. Behav.* **2016**, *1*, 0012. [[CrossRef](#)]
62. Watts, D.J.; Strogatz, S.H. Collective dynamics of ‘small-world’ networks. *Nature* **1998**, *393*, 440. [[CrossRef](#)] [[PubMed](#)]
63. Yan, G.; Jing, L. Derivation of relations between urbanization level and velocity from logistic growth model. *Geogr. Res.* **2006**, *25*, 1063–1072.
64. West, G. *Scale: The Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies*; Penguin: London, UK, 2017; p. 26.
65. Petri, G.; Expert, P.; Jensen, H.J.; Polak, J.W. Entangled communities and spatial synchronization lead to criticality in urban traffic. *Sci. Rep.* **2013**, *3*, 1798. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).