A Livable City Study in China
Using Structural Equation Models

Yang Song

Supervisor: Fan Yang-Wallentin

Department of Statistics, Uppsala University

June 2011
Abstract

The concept of *livable city* was put forward naturally and began to gain more attention when people care more about human needs during the economic development. In this paper, we define a livable city as an urban area takes the residents’ demand as first priority. It has a pleasant ecological environment, a mature community with rich public resources such as culture, transportation and medical system, and being economically well developed. Our study first reviews the theory development and literature on the subject. Then we set up a structural equation model (SEM) to verify the theory based on early works and find the dimensions that are important to the development of livable city. Using the data from *China City Yearbook, 2007*, a SEM analysis was performed. The result showed that a well developing economic system has positive influence on a city’s livability.

*Keywords*: livable city, measurement model, structural equation models
1 Introduction

The rapid development in China during the past 30 years has promoted China and Chinese people in many ways. Meanwhile, in the process of urbanization and industrialization, some inevitable problems are also accompanied. To emphasize, rapidly increased city population, shortage of the energy in the urban area, ecological destruction and pollution of environment. All these issues severely deduced the quality of life and this is against the initial intention of economic development. In this paper, we are trying to answer the following questions: what are the factors that should be considered during urban development and construction? What kind of cities are “livable”? How does the economic growths affect the human living standard? The concept of livable city was put forward naturally while “human needs” is the first priority.

1.1 Livable city

The definition of livability and livable city is not unified. Most of the studies included aspects such as ecological environment, economic performance, human environment and public facilities.

In citiesPLUS (2003), livability refers to an urban system that contributes to the physical, social and mental well being and personal development of all its inhabitants. Key principles are equity, dignity, accessibility, conviviality, participation and empowerment. Casellati (1997) defined livability as we experience ourselves as real persons in the city. Hahlweg (1997) considered livable city to be a city for all people. It should be attractive, worthwhile, safe for our children, for our older people. Salzano (1997) pointed out that a livable city is also a sustainable city: a city that satisfies the needs of the present inhabitants without reducing the capacity to satisfy the needs of the future generations.

Based on these definitions, the measurement of livable city should take both sustainable of the ecological system and the living convenience for the current residents into consideration. We define livable city as an urban area takes the residents’ demand as first priority. It has a pleasant ecological environment, a mature community with rich public resources such as culture, transportation and medical system, and being economically well developed.

1.2 Early Work Review

The discussions of the problems associate with the urban development have already started as early as 19th century. Since then, the studies of how to
improve the quality of city development never end.

In the 20th century, the scholars argued about whether the city should be centralized or scattered. The Garden City theory introduced by the British scholar Howard (1902) is the symbol of the scattered city. He planned for moderate decentralization and cooperative socialism. Clearly, it advocates low density of the urban area. On the contrary, Le Corbusier (1933) regarded the centralization as the key point of improving the efficiency.

In the beginning, the scholars only concerned the environment and economics. Later it was agreed that the social construction is also important.

Some scholars considered the residents not only as the beneficiary but also the influence of the city’s construction. Hartz-Karp (2002) pointed out that there should be a “deliberative democracy”, to make an honest, open environment to encourage the residents into the policy establish and implement. Furthermore, the promotion of the concept of sustainable urban development was mentioned. In Planning for Sustainability (2004), Wheeler discussed the recent trends in urban planning and design. He argued the “New Urbanism” is similar to those communities before the age of the automobile. The city should improve the design at building, neighborhood, city and regional levels to create more walkable, livable communities with diverse and mixed-use urban amenities.

The scholars brought out more suggestions for the problems along with urbanization. in Urban environment and infrastructure–toward livable cities (2004), World Bank listed the urban environmental goals: 1), protecting and enhancing environmental health in urban areas; 2), protecting water, soil and air quality in urban areas from contamination and pollution; 3), minimizing the urban impact on natural resources at the regional and global scales; 4), preventing and mitigating urban impacts of natural disasters and climate change. Five aspects should take into consider for reaching the goals: water and sanitation, urban development, environment, energy and transport sectors.

Lee (2008) used structural equation models (SEM) technique to measure the quality of life (QOL). The study used the well-known Detroit Area Study as the basic conceptual structure with modifications to fit the social, cultural and geographical context of Taipei. A field survey of 331 Taipei residents was conducted to give subjective resident assessments of QOL. A SEM analysis was then performed to explore the causal relationships among the QOL variables. QOL study has always been related to the livable life, a livable city can reasonably provide a high quality of life.

Sakamoto and Fukui (2004) did the research based on the system LEES (Livable Environment Evaluation-support System). It aided exploratory evaluating process by: 1), clarifying ambiguous preferences of livability using
a Fuzzy Structure Model; 2), analyzing areas using multi-criteria according to the individual preference structure; 3), visualizing distributions of the results for supporting decision-making.

In the current study, we will set up a model suitable for the livable city base on the theories and former studies, but with a different method and data.

1.3 Measurement of livable city

There are two major methods to measure livable city: subjective evaluation and objective evaluation. The subjective evaluation builds up the evaluation model and the index system by the questionnaires sent to the residents’. The objective evaluation uses the macro economic or the spacial data to build up and analysis the model. Both methods have pros and cons.

When discuss livability, residents’ subjective evaluation of their living, environment is as important as the macro economic index. The subjective evaluation must involve the residents’ feelings of their own life. In this aspect, the subjective way is much suitable for the study of quality of life (QOL) and subjective well-being (SWB). But since people’s attitude is easily influenced by external environment, the subjective evaluation could be biased. In the mean time, we need to realize that the city is objective. The shops, buildings, roads, parks, their status can all be described by data. The objective measurement may present the actual situation better than subjective measurement. In fact, according to Hortulanus (2000), the objective index and the residents’ subjective evaluation should take into consideration at the same time.

Sakamoto and Fukui (2004) proposed a Web-Service that supports the interpretation of the local environment objectively through visualized spatial information, and evaluation of the value subjectively through interactive interfaces. They mainly used Web-GIS (Geographic Information System) in this research. Currently, GIS technique is also widely used in the research of livable city.

In China, the urban hardware facilities’ level is still far below the international standard. We still have long way to reach the urban environmental goals of World Bank. My view is that the satisfaction of residents must rely on the substantial bases. For this reason, we apply the objective method to evaluate city’s livability.
2 Motivation

Most of the studies of livable city used the subjective evaluation to build a index system. This is an efficient way to measure the satisfaction and the expectation of the residents. But it has some significant shortcomings. For example, hard to replicate, expensive, time consuming, and the data is not easy to interpret. On the other hand, the objective evaluation use the macro economic data. It is convenient, inexpensive and easily interpreted. Even though there are many advantages, the objective method is not commonly adopted in livable city study so far. It is more used on the subtopics of livable city or the relative study.

Structural equation model (SEM) as one of the objective method, has increasingly been applied in the research of social science. The primary aim of SEM is to explain the pattern of a series of inter-related dependent relationships simultaneously between a set of latent (unobserved) constructs, each measured by one or more manifest (observed) variables. The procedures emphasize covariances rather than individual cases. Instead of minimizing functions of observed and predicted individual values, we minimize the difference between the sample covariances and the covariances predicted by the model (Bollen, 1989). SEM has following advantages: the first, it could deal with multiple dependent variables simultaneously; the second, it allows measurement errors; last but not least, the latent variables can be indicated with multiple observed variables. It is a useful method to deal with the social studies whose indices are always unmeasurable.

In the study of livable city, the main variables of interest can not be measured directly. For example, the convenience of public transportation, the maturity of the community, the pleasant of the ecological environment. These variables are called latent variables. In order to study the latent variables, we need a set of observable indices. With the advantages mentioned earlier, SEM is the proper method to do the study.

Early works used SEM on minor aspects of livable city, such as QOL. These works focus on the objective evaluation. They designed questionnaires and analyzed the questionnaire-data with SEM. But on subjective evaluation, there are no such studies as far as we can find. This is the reason why we choose SEM methodology to verify the theory based on early works and provide constructive suggestions.
3 Research Hypotheses

There are multiple elements would affect the standard of livable city and no agreements were made so far. We are going to summarize some based on the earlier work and bring out our research hypotheses.

Wheeler (2004) presented the three main elements of creating livable, equitable, and ecological communities, the three “E”s. Namely, Environment, Economics, Equity.

World Bank suggested five elements to achieve the goals toward livable city: water and sanitation, urban development, environment, energy and transport sectors.

Maslow (1943) proposed his famous theory Maslow’s hierarchy of needs in the paper A Theory of Human Motivation. It is often portrayed in the shape of a pyramid, with the largest and fundamental levels of needs at the bottom, the need for self-actualization at the top. They are, in turn: physiological needs, safety needs, love and belonging, esteem, self-actualization. The pyramid is illustrated in Figure 1.

![Figure 1: Maslow’s hierarchy of needs](image)

We summarize the aspects that would affect the livable city into 3 dimensions: ecological environment (Environ), public resources (Public) and level of economic development (Economic). In current study, they are also treated as latent exogenous variables. These three dimensions are suppose to have some impact on livable city (LC), the endogenous variables.

Following are the three hypotheses of the current study based on the former theory and study:
1. According to Evans (2002), ecological sustainable is one of the two main elements of livable city. And most of the researchers have mentioned comfortable environment in the definition of livable city. The ecological environment plays an important role in human life and should be given to the first priority. Also, the ability of dealing with the garbage disposal, sewage and pollution treatment of a city indicates that it can provide a quality life for its residents. Therefore, the first hypothesis is that ecological environment influences the level of a city’s livability in a positive way.

2. Doxiadis (1973) mentioned the theory about the basic principles for human settlement. It includes the maximization of people’s potential contacts with others; minimization of energy, time and cost to reach the goal; fluent contact with the basic elements of life. For example, one may get to the place he wants easily which relies on the convenient transportation; reach anyone when needed which relies on the developed postal or telecommunication system; or easily get knowledge from schools and libraries. These public resources are essential needs in daily life and livable city should have rich resources to provide its residents a colorful and convenient life. Hence, the second hypothesis is that the richness of the public resources has positive impact on the city’s livability.

3. With the development of economics, public facilities always will be improved and the investment on environment protection as well. We can not deny the side-effects with economic development will reduce the satisfaction of residents, such as higher cost of living, busier traffic and maybe higher population density. But the importance of a developed economic system is still can’t be ignored. In fact, economic development level should be a solid background for the public resources or even the ecological environment. The side-effects could be lessen or controlled under proper hands. According to Germany Spatial Development Report, 2005, disposable personal income is an important economic evaluation of quality of life (QOL). In the area of high disposable income, residents tend to be more satisfied with the community. Therefore, the third hypothesis is that a well developing economic system has positive influence on the city’s livability.
4 Data

The data used in this study is from *China city yearbook, 2007*. The yearbook describes 11 general categories of city’s construction, such as education, housing, population density, green land, waste water, public transportation, public health, culture, wage, finance and economics. Each category has several indices, while each index has few variables. It makes 34 indices and 140 variables in total.

The whole data-set contains 354 cases indicate for 354 cities/regions. The data from 69 certain remote areas are absent. Therefore these 69 cases are removed completely. In addition, there is a out-lier for both the variables “IndOut” and “Bus”, and this case is eliminated as well. It makes the sample size 284. All the variables included in the data are continuous.

Based on the discussion in Section 1, 24 variables were chosen for this study. Some variables are rescaled to force all the variables having the same scale. Names in brackets are the variable names used for LISREL input. Measuring unit is shown in parentheses. The complete list of the observed variables are:

\[ x_1: \text{The attainment rate of the industrial wastewater (％) } [\text{RWaste}] \]
\[ x_2: \text{Disposal rate of sewage water (％) } [\text{RSewage}] \]
\[ x_3: \text{Comprehensive utilization rate of industrial solid waste (％) } [\text{RSolid}] \]
\[ x_4: \text{Volumes of passenger traffic (10 million people) } [\text{VolPass}] \]
\[ x_5: \text{Green coverage rate (％) } [\text{Green}] \]
\[ x_6: \text{Postal service Per capita (Yuan) } [\text{Post}] \]
\[ x_7: \text{Local financial expenditure on research and education per capita (100 Yuan) } [\text{ResEdu}] \]
\[ x_8: \text{Domestic water consumption per capita (Cubic meters/person) } [\text{Water}] \]
\[ x_9: \text{Domestic electricity consumption per capita (10 kilowatt-hour) } [\text{Electric}] \]
\[ x_{10}: \text{Proportion of urban construction land (％) } [\text{Construc}] \]
\[ x_{11}: \text{Gross industrial output per capita (Thousand Yuan) } [\text{IndOut}] \]
$x_{12}$: GDP per capita (Thousand Yuan) $[\text{GDP}]$

$x_{13}$: GDP increasing rate (%) $[\text{IncRate}]$

$x_{14}$: Proportion of the tertiary industry in GDP (%) $[\text{PropTer}]$

$x_{15}$: Resident’s deposit balance (savings) per capita by year-end (Thousand Yuan) $[\text{Deposit}]$

$y_1$: Teachers per 100 thousands $[\text{Teachers}]$

$y_2$: Registered unemployment rate in town (%) $[\text{RUnemp}]$

$y_3$: Average wage (Thousand Yuan) $[\text{Wage}]$

$y_4$: Hospital beds per 10 thousands $[\text{HosBeds}]$

$y_5$: Doctors per 10 thousands $[\text{Doctors}]$

$y_6$: Bus per 10 thousands $[\text{Bus}]$

$y_7$: Paving road area per capita (Square meter) $[\text{Road}]$

$y_8$: Public library collections per thousand $[\text{Books}]$

$y_9$: Population density (100/square km) $[\text{PopDen}]$

Furthermore, four latent constructs are used. The exogenous latent variables are:

$\xi_1$: Ecological Environment $[\text{Environ}]$

$\xi_2$: Public Resources $[\text{Public}]$

$\xi_3$: Economic Development $[\text{Economic}]$

The endogenous latent variable is:

$\eta_1$: Livable City $[\text{LC}]$
Table 1: Distribution of Missing Values

<table>
<thead>
<tr>
<th>Variable</th>
<th>No. of M.V.</th>
<th>Variable</th>
<th>No. of M.V.</th>
</tr>
</thead>
<tbody>
<tr>
<td>RWaste</td>
<td>0</td>
<td>RSewage</td>
<td>37</td>
</tr>
<tr>
<td>RSolid</td>
<td>1</td>
<td>VolPass</td>
<td>0</td>
</tr>
<tr>
<td>Green</td>
<td>0</td>
<td>Post</td>
<td>0</td>
</tr>
<tr>
<td>ResEdu</td>
<td>0</td>
<td>Water</td>
<td>0</td>
</tr>
<tr>
<td>Electric</td>
<td>3</td>
<td>Construc</td>
<td>2</td>
</tr>
<tr>
<td>IndOut</td>
<td>0</td>
<td>GDP</td>
<td>0</td>
</tr>
<tr>
<td>IncRate</td>
<td>1</td>
<td>PropTer</td>
<td>2</td>
</tr>
<tr>
<td>Deposit</td>
<td>0</td>
<td>Teachers</td>
<td>0</td>
</tr>
<tr>
<td>RUnemp</td>
<td>0</td>
<td>Wage</td>
<td>0</td>
</tr>
<tr>
<td>HosBeds</td>
<td>0</td>
<td>Doctors</td>
<td>0</td>
</tr>
<tr>
<td>Bus</td>
<td>0</td>
<td>Road</td>
<td>0</td>
</tr>
<tr>
<td>Books</td>
<td>2</td>
<td>PopDen</td>
<td>0</td>
</tr>
</tbody>
</table>

Total Sample Size = 284

<table>
<thead>
<tr>
<th>No. of Missing Value</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Cases</td>
<td>241</td>
<td>38</td>
<td>5</td>
</tr>
</tbody>
</table>

5 Method

5.1 Treatment of the Missing data

Missing value is a common problem for data used in social science studies. There are different approaches to deal with missing values. For example, the listwise deletion, pairwise deletion and the imputation.

First do the data screening in LISREL to reveal the properties of missing data. The result is shown in Table 1.

There are 284 cases for 24 variables in total. 43 cases out of 284 are with missing values. According to the six missing-data pattern (Little, 2002), we have the reason to believe the missing is at random.

Different imputation methods require different conditions of data. Firstly, we check the multivariate normality. Note that as long as there is non-normal variable exists, it would violate the multivariate normality.

According to the Test of Univariate Normality for Continuous Variables, there are only two variables not being rejected when testing both skewness and kurtosis: PropTer and HosBeds with p-values 0.155 and 0.578, respec-
tively. All the other variables are non-normal. Therefore the data is multivariate non-normal.

Because of non-normal properties of the data, the maximum likelihood (ML) estimates of the missing values can not be explicitly implemented. Instead, we consider iterative methods of imputation, in this case, the EM algorithm.

The EM algorithm is a very general iterative algorithm for ML estimation in incomplete-data problems and the range of problems that can be attacked is remarkably broad (Meng and Pedlow, 1992). It formalizes a relatively old ad hoc idea for handling missing data: 1), Replace missing values by estimated values; 2), estimate parameters; 3), re-estimate the missing values assuming the new parameter estimates are correct; 4), re-estimate parameters, and so forth, iterating until convergence. It has two steps, the E step and M step. The M step performs ML estimation of $\theta$ just as if there were no missing data, that is, as if they had been filled in. The E step finds the conditional expectation of the “missing data” given the observed data and current estimated parameters, and then substitutes these expectations for the “missing data”. Let $Y$ denotes the data contains the missing value. Write $Y = (Y_{obs}, Y_{mis})$, where $Y_{obs}$ denotes the observed values and $Y_{mis}$ denotes the missing values. The key idea of EM is that “missing data” are not $Y_{mis}$, but the functions of $Y_{mis}$ appearing in the complete-data log-likelihood $\ell(\theta|Y)$ (Little, 2002).

Specifically, let $\theta^{(t)}$ be the current estimate of the parameter $\theta$. The E step of EM finds the expected complete-data log-likelihood if $\theta$ were $\theta^{(t)}$,

$$Q(\theta|\theta^{(t)}) = \int \ell(\theta|y)f(Y_{mis}|Y_{obs}, \theta = \theta^{(t)})dY_{mis}. \quad (1)$$

The M step of EM determines $\theta^{(t+1)}$ by maximizing this expected complete-data log-likelihood:

$$Q(\theta^{t+1}|\theta^{t}) \geq Q(\theta|\theta^{t}), \quad \text{for all } \theta. \quad (2)$$

In LISREL, it can be easily performed to impute the missing value by EM algorithm. All 43 missing cases were retrieved with imputation.

A new data screening were performed after imputation to test the multivariate normality again. The test results are shown in Table 2.

Note that all the other variables not shown in this table have the p-value 0.000. It is seen that there still are only 2 variables being normal. Others are still non-normal as before. The data-set has the property of multivariate non-normal. The box-plot of these variables are shown in Figure 2.
Table 2: Test of Univariate Normality for Continuous Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Chi-Square</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RWaste</td>
<td>194.501</td>
<td>0.000</td>
</tr>
<tr>
<td>RSewage</td>
<td>11.823</td>
<td>0.003</td>
</tr>
<tr>
<td>PropTer</td>
<td>3.609</td>
<td>0.165</td>
</tr>
<tr>
<td>HosBeds</td>
<td>1.081</td>
<td>0.583</td>
</tr>
<tr>
<td>PopDens</td>
<td>106.951</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Figure 2: Boxplot of the imputed observed variables

5.2 Estimation Method

5.2.1 Maximum Likelihood

The fundamental hypothesis for the structural equation procedures is that the covariance matrix of the observed variables is a function of a set of parameters in the model. As a result, we minimize the difference between the sample covariances and the predicted covariances implied by the model.

For each model, let $S$ be the sample covariance matrix and $\Sigma$ be defined as the implied covariance matrix. The unknown parameters in $\Gamma$, $\Phi$, and $\Psi$ are estimated so that the estimated implied covariance matrix $\hat{\Sigma}$ is close to the sample covariance matrix $S$. The fitting function to be minimized with respect to the free elements of $\Lambda$ and $\Phi$ is,

$$F_{ML} = \log|\Sigma(\Lambda, \Phi)| + tr(S\Sigma(\Lambda, \Phi)^{-1}) - \log|S| - (p + q) .$$

ML method is very widely used, but it is derived based on the assumption that the observed variables have a multi-normal distribution. In our
case, as well as most of the cases with real data, the assumption is violated. The consequence of violation is that the standard errors and Chi-squares will be wrongly estimated. Therefore, robust maximum likelihood (RML) is proposed as a suitable estimation method for the situation.

5.2.2 Robust Maximum likelihood

The fit function of RML is slightly different with ML. According to Jöreskog, Sörbom and du Toit (2000), though the fit function for ML is usually written in the form of Equation (3), but an asymptotically equivalent form is always shown as

\[ F_{ML} = (s - \sigma)'D'(\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1})D(s - \sigma). \]

(4)

where \( s \) is a vector of order \( s \times 1 \) consisting of the nonduplicated elements of \( S \), \( D \) is the duplication matrix which transforms \( s \) to \( \text{vec}(S) \), \( \otimes \) denotes for the Kronecker product. Equation (4) can be interpreted as ML estimated by means of iteratively reweighted least squares in which \( \hat{\Sigma} \) is updated in each iteration. Both of these fit functions have a minimum at the same point in the parameter space, namely at the ML estimates. However, the minimum value of the functions are not the same.

Another important estimate is the standard errors of the parameter estimates. They are obtained from the diagonal elements of Matrix (5).

\[ (\Delta'V\Delta)^{-1}\Delta'VWV\Delta(\Delta'V\Delta)^{-1}. \]

(5)

where \( \Delta = \partial \sigma / \partial \theta \), \( \theta \) be the estimation of the free parameters in the model, \( V = D'(\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1})D \), and \( W = W_{\text{NNT}} \) is the weight matrix under the non-normal condition which will compute by PRELIS if asymptotic covariance matrix is provided. The standard error of \( \hat{\theta}_i \) is \( 1/N \) times the square root of the \( i \)th diagonal element of this matrix.

As for the chi-square for testing structural equation models, few are suggested. Following the notation in Jöreskog, Sörbom, Du Toit (2003), these chi-squares are denoted as \( c_1, c_2, c_3 \) and \( c_4 \). Here \( c_1 \) is \( n \) times the minimum value of the fit function, \( c_2 \) is \( n \) times the minimum of the WLS fit function using a weight matrix estimated under multivariate normality, \( c_3 \) is the Satorra-Bentler scaled chi-square statistic (Satorra & Bentler, 1988, equation 4.1), and \( c_4 \) is equation (2.20a) in Browne (1984) using the asymptotic covariance matrix provided. When dealing with non-normal data, \( c_3 \) is suggested (Yang-Wallentin, Jöreskog, Luo, 2010). Although the asymptotic distribution of \( c_3 \) is not exactly \( \chi^2 \), it is used as a \( \chi^2 \) statistic because the scale factor is estimated such that \( c_3 \) has an asymptotically correct mean
even under non-normality,
\[c_3 = \frac{d}{h_1} c_2.\] (6)

where \(d\) is the degrees of freedom and
\[c_2 = n(s - \hat{\sigma})' \Delta_c (\Delta'_c W_{NT} \Delta_c)^{-1} \Delta'_c (s - \hat{\sigma});\] (7)
\[h_1 = tr[(\Delta'_c W_{NT} \Delta_c)^{-1} (\Delta'_c W_{NNT} \Delta_c)].\] (8)

Here \(\Delta_c\) is an orthogonal complement to \(\Delta\) such that \(\Delta'_c \Delta = 0\), recall \(W_{NNT}\) is the weight matrix under the non-normal condition which will compute by PRELIS if asymptotic covariance matrix is provided and
\[W_{NT} = 2(D'D)^{-1} D'(\hat{\Sigma} \otimes \hat{\Sigma}) D(D'D)^{-1}.\] (9)

6 Model

6.1 Model Specification

Given the former informations, we establish the model for the livable city. Following is the general representation of structural equation models, both the structural model and measurement model.

- **Structural Model**
  \[\eta = \Gamma \xi + \zeta.\] (10)

  Where
  \[\eta = [\eta_1], \quad \xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 \end{bmatrix}, \quad \zeta = [\zeta_1].\]

  \[\Phi = \text{cov}(\xi) = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix}, \quad \Psi = \text{cov}(\zeta) = \begin{bmatrix} \psi_{11} \end{bmatrix}.\]

  Note that \(\eta_1\) represents the independent latent variable \(LC\), \(\xi_1, \xi_2, \xi_3\) represent the dependent latent variables \(Environ, Public, Economic\), respectively.

  For the structural model, we assume that
1. \(E(\eta) = 0, \ E(\xi) = 0, \ E(\zeta) = 0;\)
2. \(\text{corr}(\zeta, \xi) = 0.\)

• Measurement Model
\[
x = \Lambda_x \xi + \delta, \\
y = \Lambda_y \eta + \epsilon. \tag{11}
\]
where
\[
x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{15} \end{bmatrix}, \quad \Lambda_x = \begin{bmatrix} \lambda_1 & 0 & 0 \\ \lambda_2 & 0 & 0 \\ \lambda_3 & 0 & 0 \\ 0 & \lambda_4 & 0 \\ 0 & \lambda_5 & 0 \\ \vdots & \vdots & \vdots \\ 0 & \lambda_{10} & 0 \\ 0 & 0 & \lambda_{11} \\ 0 & 0 & \lambda_{12} \\ \vdots & \vdots & \vdots \\ 0 & 0 & \lambda_{15} \end{bmatrix}, \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_{15} \end{bmatrix},
\]
\[
\Theta_\delta = \text{diag}[\text{var}(\delta_1), \ldots, \text{var}(\delta_{15})].
\]
\[
y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_9 \end{bmatrix}, \quad \Lambda_y = \begin{bmatrix} \lambda_{16} \\ \lambda_{17} \\ \vdots \\ \lambda_{24} \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_9 \end{bmatrix}, \quad \Theta_\epsilon = \text{diag}[\text{var}(\epsilon_1), \ldots, \text{var}(\epsilon_9)].
\]

Note that to scale the latent variables, set the first observed variable of each latent variable as the reference variable. That is to say: \(\lambda_1 = \lambda_5 = \lambda_{12} = 1, \lambda_{16} = 1.\)

For the measurement model, we assume the following:
1. \(E(\eta) = 0, \ E(\xi) = 0, \ E(\epsilon) = 0, \ E(\delta) = 0;\)
2. \(\text{corr}(\epsilon, \eta) = 0, \ \text{corr}(\epsilon, \xi) = 0, \ \text{corr}(\epsilon, \delta) = 0;\)
3. \(\text{corr}(\delta, \xi) = 0, \ \text{corr}(\delta, \eta) = 0.\)
To summarize the free parameters in the model,

$$\theta' = [\Gamma, \Phi, \Psi, \Lambda_x, \Lambda_y, \Theta_\delta, \Theta_\epsilon].$$  \hspace{1cm} (12)

The path diagram of the model is shown in Figure 3.

Figure 3: Path Diagram of the initial Livable city model

6.2 Implied Covariance Matrix

According to Bollen (1989), the estimation is based on choosing the values of the structural parameters to reproduce the covariance matrix. Thus it is important to understand the relation of the covariance elements to the parameters.

$$\Sigma = \begin{bmatrix}
\text{cov}(y, y) & \text{cov}(y, x) \\
\text{cov}(x, y) & \text{cov}(x, x)
\end{bmatrix}. \hspace{1cm} (13)$$
The elements in the implied covariance matrix are:

\[ \Sigma_{yy} = E(yy') = E[(\Lambda_y \eta + \epsilon)(\eta'y + \epsilon')] = \Lambda_y E(\eta\eta')\Lambda_y' + E(\epsilon\epsilon') = \Lambda_y(\Gamma\Phi\Gamma' + \Psi)\Lambda_y' + \Theta_{\epsilon} ; \]

\[ \Sigma_{yx} = E(yx') = E[(\Lambda_y \eta + \epsilon)(\xi'\Lambda_x' + \delta')] = \Lambda_y E(\eta\xi')\Lambda_x' + E(\epsilon\delta') = \Lambda_y\Gamma\Phi\Lambda_x' + \Theta_{\epsilon\delta} ; \]

\[ \Sigma_{xy} = E(xy') = \Sigma_{yx}' = \Lambda_x\Phi\Gamma'\Lambda_y' + \Theta_{\delta\epsilon} ; \]

\[ \Sigma_{xx} = E(xx') = E[(\Lambda_x \xi + \delta)(\xi'\Lambda_x' + \delta')] = \Lambda_x E(\xi\xi')\Lambda_x' + E(\delta\delta') = \Lambda_x\Phi\Lambda_x' + \Theta_{\delta} . \]

They form the following implied covariance matrix based on the free parameters:

\[
\Sigma(\theta) = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} = \begin{bmatrix} \Lambda_y(\Gamma\Phi\Gamma' + \Psi)\Lambda_y' + \Theta_{\epsilon} & \Lambda_y\Gamma\Phi\Lambda_x' + \Theta_{\epsilon\delta} \\ \Lambda_x\Phi\Gamma'\Lambda_y' + \Theta_{\delta\epsilon} & \Lambda_x\Phi\Lambda_x' + \Theta_{\delta} \end{bmatrix} .
\]

Each of the element in Equation (13) matches the corresponding element in Equation (18), the variances and covariances are decomposed into their structural parameters. In order to estimate the model, we must determine whether the model parameters are identified.

6.3 Identification

Model identification is the first step of the estimation procedure. Investigations of identification begin with one or more equations relating known and unknown parameters. Identification is demonstrated by showing that
the unknown parameters are functions only of the identified parameters and that these functions lead to unique solutions. If this can be done, the unknown parameters are identified (Bollen, 1989). The easiest test to apply the identification is $t$-Rule. The $t$-Rule for identification is that the number of non-redundant elements in the covariance matrix of the observed variables must be greater than or equal to the number of unknown parameters in $\theta$ (See Equation 12), i.e.,

$$t \leq \left(\frac{1}{2}\right)(p + q)(p + q + 1).$$ (19)

where $p$ is the number of the dependent observed variables, $q$ is the number of independent observed variables and $t$ is the number of free parameters in $\theta$. While in this specific case, $p$ is 9, $q$ is 15 and $t$ is 54. The $t$-Rule is met so that the model is identified.

7 Result

We performed a prior analysis of measurement model before running the full model. In this way, we can ensure that all the measures are valid for measuring the latent factors. Note that the model fits the data reasonably is the base of further analysis of the structural model. The result shows that the measurement model fits approximately. Therefore, we go on with the estimation of the full model.

7.1 Result of the initial model

The hypothesis of the model is $\Sigma = \Sigma(\theta)$. Based on the discussion in Section 5.2, RML method is adapted to estimate the unknown parameters.

LISREL is used to estimate the 54 parameters. The goodness of fit indices are shown in Table 3. Clearly, this model does not fit. None of the GOF indices reach the suggested limit. Furthermore, the 3 paths of the structural model are all non-significant. The model has to be improved, otherwise, there is no meaning to explain the parameter estimates at all.

7.2 Model Modification

Two possible problems can lead to the bad fit of structural equation models. One is misspecification and the other is the data does not fit the hypothetical distribution. In our case, the reason could be from the model specification. It could be lacking of some observed variables or relative latent variables, or
Table 3: Goodness of Fit Indices of the Initial *Livable City* Model

<table>
<thead>
<tr>
<th>Index</th>
<th>Suggested limit</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satorra-Bentler Scaled $\chi^2$</td>
<td>As small as possible</td>
<td>1266.20</td>
</tr>
<tr>
<td>df</td>
<td>As small as possible</td>
<td>246</td>
</tr>
<tr>
<td>RMSEA</td>
<td>$&lt; 0.08$</td>
<td>0.12</td>
</tr>
<tr>
<td>GFI</td>
<td>$&gt; 0.90$</td>
<td>0.66</td>
</tr>
<tr>
<td>NFI</td>
<td>$&gt; 0.90$</td>
<td>0.83</td>
</tr>
<tr>
<td>CFI</td>
<td>$&gt; 0.90$</td>
<td>0.86</td>
</tr>
</tbody>
</table>

lacking some paths between certain variables. Under the circumstance, we could use the significant test result and the suggested model modify index given by LISREL to improve the model and make it fits the data better.

Based on the *modification indices* given in LISREL, a few error correlations are suggested. Adding the correlations could have great impact on the model improvement. Error correlations means these two variables have some in common in the left-overs. They are reasonable, especially in economic and social data since the social phenomenon is complicated, any variable could have relations with others. The variables have correlated errors are:

“Doctors” and “HosBeds”; “Water” and “Electric”; “PopDen” and “Construc”; “IndOut” and “ResEdu”; “GDP” and “Wage”; “Electric” and “GDP”; “IndOut” and “GDP”; “PropTer” and “IndOut”; “PropTer” and “GDP”; “Deposit” and “PropTer”.

Doctors and hospital beds represent the status of the city’s medical system, it is no doubt that there will be common effects in their left-overs. The consumption of water and electricity could have common in the residents’ habits and daily life. Since construction land per capita reflects the status for each resident, the population density could have impact on it. The input of research and education must rely on the economic status of a city, which highly relies on the industrial output. Clearly, GDP reflects the economic condition from the macro way, it is easy to understand that it has some common in left-overs with wage (which reflects the economic condition from the micro way), electricity consumption (which would be effected by the financial status), proportion of the tertiary (very important composition of GDP), and industrial output (one of the very important index in macro economics). The performance of tertiary industry is part of the industrial output, the proportion of it should have something in common with industrial output’s left-overs; savings are left of wage after the investment and consumption, since the consumption relates to the tertiary industry closely, it is reasonable
to have common in the left-overs for these two variables.

Furthermore, the estimation of the initial model show that the coefficients of “RSolid” to \textit{Environ} and “PropTer” to \textit{Economic} are not significant and very close to zero. Therefore, these two paths are removed from the model.

The modified path diagram is shown in Figure 4 (note that the error correlations of “PopDen” and “Construc”; “GDP” and “Wage” are not marked on the path diagram). The result of the modified model estimation is shown in Table 4.

![Figure 4: Path Diagram of the modified Livable City model](image)

The main indices of GOF statistics is represented in Table 5. Though the model is rejected for the close fit test since p-value is still low, we could conclude that the model fits the data approximately according to the other fit indices. RMSEA is below the limit of 0.08, NFI and CFI are larger than 0.9. Detailed explanations of a result are in the proceeding sections.
7.3 Result of the modified model

The first, we analysis the measurement models. Table 4 reveals the factor loadings. All of the coefficients are statistically significant and match the theories and common sense.

For *Environ*, when “RWaste” increasing by one unit, it will improve 0.52 unit.

For *Public*, “Electric” has the most influence, following by “VolPass” and “Water”. We could conclude that the basic needs like domestic electricity and water consumption take a great part as public resources measurement.

For *Economic*, “IndOut” has the most impact. When “IndOut” increase by one unit, *Economic* will improve 2.61 unit. It is clear that gross industrial output is still take the leading role when measure the status of the economical status.

For *LC*, “HosBeds” has the largest impact. “Doctors” follows. It is clear that the medical status takes the most important role to measure the livable city. “RUnemp” is a special variable, since it represents the rate of the unemployment, we would like it to be as small as possible. That makes the factor loading of “RUnemp” reasonably minus. When it decrease 1 unit, *LC* will improve by 0.34 unit.

Secondly, for the structural model. The results are not very satisfactory. The path from *Economic* to *LC* is significant. The estimate is fairly small (0.06), shows that when the status of economics improve by 1 unit, the measure of livable city will improve by 0.06 unit. It evidenced the theory and hypothesis 3, that a well developing economic system has positive influence on a city’s livability. Small coefficient imply that even a small step forward would need huge effort on relevant aspects. But both the coefficients from *Environ* and *Public* to *LC* are not significant. The hypothesis 1 and 2 are not solidly demonstrated.

8 Discussion

From the analysis of the model, it turned out that the data does not fit the model well. By goodness of fit test, RMSEA, NFI, CFI are much better than the critical values, but the chi-square test of the model is highly significant which means to reject the null hypothesis $\Sigma = \Sigma(\theta)$. Though the measurement models are acceptable, but the structural model has two main coefficients non-significant. The possible reasons are discussed as following.

First, the quality of the data is not good enough. For a empirical study, the sample size 284 is relatively small. The problem could be solved by
using another data. We could choose the data from another year or another country. If it still does not work well, we could isolate the problem from the data.

Furthermore, the model may be misspecified. Since the social and economic phenomenons are complex and interrelated, it is very likely that more variables should be included. It could be either the latent variable or the observed variable. In fact, the choice of the observed variables are limited by the data-set. Other than that, there may be some ignorance of the error correlations as well.

In addition, it could be insufficient to only use objective method to analysis such topic. As discussed before in Section 1.3, the objective evaluation uses the macro economic data to establish and analysis the livable city model while the subjective evaluation relies on questionnaires to do the evaluation. Our result shows that the measurement of the livable city can not only take the objective aspect into consideration. If we could combine subjective questionnaire data as well as the city data from macro economic year book, the model would be more adequate and we have greater opportunity to get a better result.

The purpose of this study is to build up a structural equation model that could confirm the theories that sociologist established in the area of livable city in general terms. We applied the SEM model and fit it with data from City Year Book, 2007. The results show that the latent variable Economic is a significant influence on livable city. At the meantime, all the factor loadings of the observed variables to measure the latent variables are significant. Though the model does not verify the hypothesis 1 and 2, but we still regard it as a model with practically significance.

With the increasingly attention on the human needs along the economic development, future studies are needed for this subject. The estimation of the model with panel data is one of the area that can be of interest.
References


Appendix

The variance-covariance matrix of the data-set is in the text file `city.cm` and the asymptotic covariance matrix if in the text file `city.acc`.

Initial model can be estimated by RML using the following PRELIS syntax file:

```
Observed variables: RWaste RSewage RSolid VolPass Green Post
                   ResEdu Water Electric Construc IndOut GDP
                   IncRate PropTer Deposit Teachers RUNemp Wage
                   HosBeds Doctors Bus Road Books PopDen

Covariance matrix from File city.cm
Asymptotic Covariance Matrix from File city.acc

Latent Variables: Environ Public Economic LC
Sample Size 284
Relationships:
RWaste = 1*Environ
RSewage = Environ
RSolid = Environ

IndOut = 1*Economic
GDP = Economic
IncRate = Economic
PropTer = Economic
Deposit = Economic

VolPass = 1*Public
Green = Public
Post = Public
ResEdu = Public
Water = Public
Electric = Public
Construc = Public

Teachers = 1*LC
RUNemp = LC
Wage = LC
HosBeds = LC
```
Doctors = LC
Bus = LC
Road = LC
Books = LC
PopDen = LC

LC = Economic Public Environ

Options: AD=OFF IT = 2000
Path Diagram
LISREL output
End of Problem

Modified model can be estimated by RML using the following PRELIS syntax file:


Covariance matrix from File city.cm
Asymptotic Covariance Matrix from File city.acc
Latent Variables: Environ Public Economic LC
Sample Size 284
Relationships:

RWaste = Environ
RSeWage =1* Environ
!RSolid = Environ

IndOut = Economic
GDP = 1*Economic
IncRate = Economic
!PropTer = Economic
Deposit = Economic

VolPass = 1* Public
Green = Public
Post = Public
ResEdu = Public
Water = Public
Electric = Public
Construc = Public
Teachers = 1* LC
RUnemp = LC
Wage = LC
HosBeds = LC
Doctors = LC
Bus = LC
Road = LC
Books = LC
PopDen= LC

LC = Economic Public Environ

Set the errors of Doctors and HosBeds correlate
Set the errors of Water and Electric correlate
Set the errors of PopDen and Construc correlate
Set the errors of IndOut and ResEdu correlate
Set the errors of GDP and Wage correlate
Set the errors of Electric and GDP correlate
Set the errors of IndOut and GDP correlate
Set the errors of PropTer and IndOut correlate
Set the errors of PropTer and GDP correlate
Set the errors of Deposit and PropTer correlate
Set the errors of Public and Economic correlate

Options: AD=OFF IT = 2000
Path Diagram
LISREL output
End of Problem
Table 4: The RML Estimates for the Modified *Livable City* Model.

The Measurement Equations

<table>
<thead>
<tr>
<th>Observed Variables</th>
<th>Parameters $\Lambda_x$</th>
<th>RML Estimate (T-value)</th>
<th>Observed Variables</th>
<th>Parameters $\Lambda_y$</th>
<th>RML Estimate (T-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RWaste</td>
<td>$\lambda_1$</td>
<td>0.52</td>
<td>Teachers</td>
<td>$\lambda_{16}$</td>
<td>1.00$^c$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.80)</td>
<td></td>
<td></td>
<td>(-)</td>
</tr>
<tr>
<td>RSewage</td>
<td>$\lambda_2$</td>
<td>1.00$^c$</td>
<td>RUnemp</td>
<td>$\lambda_{17}$</td>
<td>-0.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-)</td>
<td></td>
<td></td>
<td>(-2.81)</td>
</tr>
<tr>
<td>VolPass</td>
<td>$\lambda_4$</td>
<td>1.00$^c$</td>
<td>Wage</td>
<td>$\lambda_{18}$</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-)</td>
<td></td>
<td></td>
<td>(6.30)</td>
</tr>
<tr>
<td>Green</td>
<td>$\lambda_5$</td>
<td>1.01</td>
<td>HosBeds</td>
<td>$\lambda_{19}$</td>
<td>8.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.20)</td>
<td></td>
<td></td>
<td>(8.78)</td>
</tr>
<tr>
<td>Post</td>
<td>$\lambda_6$</td>
<td>1.19</td>
<td>Doctors</td>
<td>$\lambda_{20}$</td>
<td>5.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.22)</td>
<td></td>
<td></td>
<td>(10.16)</td>
</tr>
<tr>
<td>ResEdu</td>
<td>$\lambda_7$</td>
<td>1.08</td>
<td>Bus</td>
<td>$\lambda_{21}$</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.04)</td>
<td></td>
<td></td>
<td>(5.79)</td>
</tr>
<tr>
<td>Water</td>
<td>$\lambda_8$</td>
<td>1.47</td>
<td>Road</td>
<td>$\lambda_{22}$</td>
<td>2.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.50)</td>
<td></td>
<td></td>
<td>(6.83)</td>
</tr>
<tr>
<td>Electric</td>
<td>$\lambda_9$</td>
<td>6.69</td>
<td>Books</td>
<td>$\lambda_{23}$</td>
<td>2.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.48)</td>
<td></td>
<td></td>
<td>(6.53)</td>
</tr>
<tr>
<td>Construc</td>
<td>$\lambda_{10}$</td>
<td>1.05</td>
<td>PopDen</td>
<td>$\lambda_{24}$</td>
<td>1.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.95)</td>
<td></td>
<td></td>
<td>(3.97)</td>
</tr>
<tr>
<td>IndOut</td>
<td>$\lambda_{11}$</td>
<td>2.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(6.64)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP</td>
<td>$\lambda_{12}$</td>
<td>1.00$^c$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IncRate</td>
<td>$\lambda_{13}$</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.91)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deposit</td>
<td>$\lambda_{15}$</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.88)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: c = constrained to equal 1.*

The Structural Equations

\[
\text{LC} = 0.01 \text{ Environ} + 0.14 \text{ Public} + 0.06 \text{ Economic.}
\]

(0.73) (0.89) (2.01)
Table 5: Goodness of Fit Indices of the Modified *Livable City* Model

<table>
<thead>
<tr>
<th>Index</th>
<th>Suggested limit</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satorra-Bentler Scaled $\chi^2$</td>
<td>As small as possible</td>
<td>465.29</td>
</tr>
<tr>
<td>df</td>
<td>215</td>
<td></td>
</tr>
<tr>
<td>RMSEA</td>
<td>&lt; 0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>GFI</td>
<td>&gt; 0.90</td>
<td>0.82</td>
</tr>
<tr>
<td>NFI</td>
<td>&gt; 0.9</td>
<td>0.93</td>
</tr>
<tr>
<td>CFI</td>
<td>&gt; 0.9</td>
<td>0.96</td>
</tr>
</tbody>
</table>