Alma mater studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

Economia e Statistica Agroalimentare

Ciclo XXVI

Settore Concorsuale di afferenza: SECS-P/02 POLITICA ECONOMICA

Settore Scientifico disciplinare: 13/A2- POLITICA ECONOMICA

Spatial Disparity and Dynamic Trajectory of Convergence in China: Construction and Application of a Composite Index of Regional Development

Presentata da: **Peng Bin**

Coordinatore Dottorato

Roberto Fanfani

Relatore

Roberto Fanfani

Esame finale anno 2014

TABLE OF CONTENTS

ABSTRACT

China is a large country characterized by remarkable growth and distinct regional diversity. Spatial disparity has always been a hot issue since China has been struggling to follow a balanced growth path but still confronting with unprecedented pressures and challenges. To better understand the inequality level benchmarking spatial distributions of Chinese provinces and municipalities and estimate dynamic trajectory of sustainable development in China, I constructed the Composite Index of Regional Development (CIRD) with five sub pillars/dimensions involving Macroeconomic Index (MEI), Science and Innovation Index (SCI), Environmental Sustainability Index (ESI), Human Capital Index (HCI) and Public Facilities Index (PFI), endeavoring to cover various fields of regional socioeconomic development. Ranking reports on the five sub dimensions and aggregated CIRD were provided in order to better measure the developmental degrees of 31 or 30 Chinese provinces and municipalities over 13 years from 1998 to 2010 as the time interval of three "Five-year Plans". Further empirical applications of this CIRD focused on clustering and convergence estimation, attempting to fill up the gap in quantifying the developmental levels of regional comprehensive socioeconomics and estimating the dynamic trajectory of regional sustainable development in a long run. Four clusters were benchmarked geographically-oriented in the map on the basis of cluster analysis, and club-convergence was observed in the Chinese provinces and municipalities based on stochastic kernel density estimation.

Keywords: *Composite index, Regional sustainable development, Comprehensive socioeconomics, Spatial disparity, Convergence, Principal Component Analysis, Cluster Analysis, Stochastic Kernel Density, Dynamic trajectory*

1. Introduction

Uneven development exists as a common phenomenon in the development of human society, referring to the uncoordinated and mismatching relationships in the development process. It is widespread internationally, with only differences in the unequal degrees or phases. China, as an important economy of the world, has been plagued by this problem in its remarkable growth in recent years.

1.1. Research background

The well-known reform and opening up in China initiated its fast growth and created remarkable achievements in the past decades, whereas also brought side effects of inequality that has been widening up along with the rapid development. Even though the GDP of China is claimed to increase in an average rate of 8% during the last decade, there were still around 128 million Chinese people living in the poverty as reported, occupying 10% of the whole population approximately[1]. The Pareto Principle (The 80/20 Rule)[2] can vividly picture this unequal wealth distribution as 80% of national wealth is held by 20% of the population. Gini coefficient[3] could clearly picture the wide gap in the whole nation. It is officially published that the Gini coefficient of China has been altering in the range of 0.473 and 0.491 during the last decade (2003-2013). Though these data were already way higher than the official alert line (0.4), many scholars and economists still argued they did not accord with the fact and proposed even higher Gini coefficients by their studies, most of which were around 0.6 to 0.7 and more acceptable to the public. If extended to a macro way, regional disparity has always been a hot topic in both research and empirical fields. There has been, on one hand, a large gap between urban and rural areas, and on the other, uneven development between coastal and inland areas. The Pareto Principle (The 80/20

[1] The data was reported in China News (www.chinanews.com), and the poverty line here is the new one defined by the Chinese government, with the standard of income per capita as $1.8 dollars per day, higher than the international standard of $1.25 dollars per day. If calculated in the international standard, then the poverty population was 26.88 million in 2011.

[2] The Pareto principle (also known as the 80–20 rule), as Wikipedia defined, states that, for many events, roughly 80% of the effects come from 20% of the causes.

[3] The Gini coefficient (also known as the Gini index or Gini ratio), as Wikipedia defined, is a measure of statistical dispersion intended to represent the income distribution of a nation's wealth.

Rule)[4] is also applicable to depict this phenomenon in China: a relatively large portion of the national wealth has been created and owned by the high-developed regions while the rest only take a small proportion.

The regional disparity in China reflects in many aspects in economics, social development, urbanization, and etc. As reported, the top three regions with the highest per capita GDP was Shanghai (RMB80, 345 Yuan), Beijing (RMB76, 544 Yuan) and Tianjin (RMB71, 723 Yuan) locating in the east coast, where, geographically, was also the top class of China in 2010; the bottom three were Yunnan (RMB15, 628 Yuan), Gansu (RMB 14,404 Yuan) and Guizhou (RMB 11, 749 Yuan), all of which located in the western area. It is clear that the per capita GDP of Shanghai is almost seven times of Guizhou's, and this gap was even larger (nearly 13 times) in 2004. If extended to the geographic dimension, the average per capita income of western area (RMB18, 090 Yuan) was only 46.88% of eastern (RMB38,587 Yuan) with a gap of RMB20, 497 Yuan in 2010, and a majority (nearly 90%) of poverty population mentioned above was distributed in the western area[5]. This unequal spatial distribution also reflected in the social aspect. There were 16 regions reported as with lower level of social development than national average, all belonging to western area, whereas the top ten regions with higher level of social development all located in eastern area.[6]. The household expenditure in education in western area was only 73.5% of eastern; more than 70% of superior medical and health resources were concentrated in the urban, whereas the mass rural area with nearly half of national population only took less than 30%[7]. Moreover, the urbanization level was also lower in the western area than eastern area, half of the cities (administrative level) located in the east and only 20% was distributed in the west. According to the official statistical data published in 2010, the density of city distribution was 330 units per million of sq.km in the east, whereas this number was reduced to 30 units in the west. Meanwhile, the

---

[4] The Pareto principle (also known as the 80–20 rule), as Wikipedia defined, states that, for many events, roughly 80% of the effects come from 20% of the causes.

[5] All the data and calculation in this paragraph are on the basis of the statistical reports published by National Bureau of Statistics in PRC.

[6] This is concluded from the social development indices reported by National Bureau of Statistics in PRC, which evaluated various aspects such as education, medical health, sanitation and etc.

[7] According to the statistical report, the average educational level of medical staffs in some underdeveloped rural regions is low, with only 1% with bachelor degree.

population density in eastern urban area was 11 times of western. Half of the metropolitans with million(s) population were located in the east, whereas the western metropolitans only took 19.23%[8] of national total.

Many studies focus on grouping China through the uneven socioeconomic conditions: some were according to the migration and earnings (Y. Zhao, 1999), some based on the correlation of average growth rate and education level (F. Cai, et al., 2002), and some specialized researches focused on the rural classification (R. Fanfani and C. Brasili, 2003). An interesting conception of "one China, four classes" was emerging in the last decade[9]. China was classified into four groups by the income level: the first class was compose by the superior developed regions such as Shanghai and Beijing, the second class was constituted with majority of east coastal provinces and municipalities such as Tianjin, Guangdong, Zhejiang, Jiangsu, Fujian, et.al., the third class was composed by the northeast and some central regions such as Hebei and Shanxi, et.al., and the fourth class was composed by the peripheral regions such as Guizhou and Yunnan, et.al.. The gap between the first and the fourth is enormous, as the latter lagged behind half a century to the former in its socioeconomic development.

Many researchers focus on whether the inequality narrowed down or widened up and what types of factors influenced its change (Chen, J., 1996; Cai, F. 2002). Possible reasons as political, historical, geographic factors are examined to estimate the causes of the inequality and provide further suggestions to policy makers for social stability (Wang, 1999; Cai, 2002). Since it is complicated to explain perfectly why the deep differences existing in Chinese regions and how they changed in the past decades, the disparities of regional development in China are certainly a significant angle to study. The high regionalism of Chinese policies in the early stage of its marketization reform started from 1978 has certainly provided a head-start for east coastal regions, with the initiatives of benefiting the whole country (Fan, 1997) from uniform wealth via Deng Xiaoping's guideline of permitting some regions to get rich first. The design of Special Economic

[8] The latest published data reported there are 52 metropolitans in China, in which the eastern took 26, the central took 16 and the western took only 10.

[9] This conception was first proposed by Hu, A.G. (2001), based on the World Bank's criterion of income group, with the application of GDP per capita ($ PPP).

Zones (SEZ) and opening of 14 coastal cities for the foreign investors in 1984 promoted the market liberalization and Foreign Direct Investment (FDI) policies, in one aspect (Démurger et.al, 2002; Yang, 1997), whereas also led to unbalanced development and regional inequality in contrast. The central and western regions have been left far behind in their socioeconomics in spite of series of supporting policies and programs launched later by national and local governments to balance the uneven growth. The first movement was "Western Region Development Strategy" in 2000 by the Chinese government orientated to promote local socioeconomic development such as infrastructure and resource and further strengthen national defense. However, the effects of the supporting policy for undeveloped regions are not well paid off in eliminating the polarization, even though it has indeed provided remarkable achievements in poverty reduction and wealth creation (reducing the proportion of population living in poverty from 81.6% to 10.4%, and wealth creation remained uneven among provinces) (Chen, 2008). This Chinese governmental reform-design of regional inclination shifting from east to central and west initiates a pattern of growth clubs, where coastal regions get priorities to develop ahead and other inland regions start to catch up in the later phase (Andersson, et.al, 2013).

The Chinese central government started to realize the series of consequences coming up with the biased policy and conduct structural adjustments of regional development. The "Ninth Five-year Plan" started in 1996, firstly pinpointed to adopt "adhering to the coordinated development of regional economy and gradually narrowing down the regional gaps" as one of the most important guidelines to be implemented in the following fifteen years. Variety of subsequent plans and supporting policies and programs had been brought up. The 15th Party Congress in 1997 further emphasized to improve the reasonable structure of regional layout and coordinated growth, and also analyzed the regional disparities shown then. In the "Tenth Five-year Plan" started in 2001, a new supporting program of "Western Development (Go-West)" was commenced; following-up supporting programs aimed at revitalization of the traditional northeast industry base and spring up central China ("Go-Central") were implemented in succession at 2003 and 2004 respectively. The "Eleventh Five-year Plan" pointed out "…strengthening regional coordination and interaction mechanisms…" and "…encouraging the eastern region to take the lead in development, so as to form a pattern in which the eastern, central and western regions act together, take advantage of each other's strengths, promote each other and enjoy common development". Therefrom, I target this study in regional development over the time interval covering the three "Five-year Plans"[10], hence to analyze the Chinese spatial disparities and dynamic trajectory of regional development under this macro background.

1.2. Research significance and objective

Coordinating the regional development is the key fundamental to guarantee the stable and efficient growth of national economics in China. The central and west areas possess majority of land and over half of population, representing massive space to develop potential market and pull up gross national consumption. To diminish the gap among the east and the central and west, not only can improve the life quality of the local inhabitants in the central and west, but also can benefit the eastern area with the development of potential market and resource input-output. Moreover, the central and west are mostly mountainous area with favorable natural resources but fragile ecological system, also are traced as headstreams of the main national river in China, and hence perform a substantial role in China's environmental protection and sustainable development. Given the significance of coordinated regional development, it is noteworthy to address the research on spatial disparity and regional dynamic trajectory as one major topic in the long run. In this study, I attempt to provide a scientific and systematic measure for the evaluation of comprehensive regional socioeconomic development in China, and therefrom quantify the developmental levels of different regions and depict the heterogeneous or homogeneous features among their development paths.

However, due to regional characteristics of Chinese policies and its massive administrative area, and also the availability of data, it is complicated to capture the exact pattern of growth trend in a long term. A certain amount of researches focus on whether

[10] My initial intention was to set up the time interval from 1996 to 2010, with a full coverage of the three "Five-year Plan". However, there were massive missing data in environment, science and technology, and public facility in Chinese yearbook 1996 and 1997. The official Chinese yearbook changed its structure and statistical calendar in 1998. Therefore, in order to keep consistence of the data, I finally decided to fix the time interval from 1998 to 2010. I think it still make sense to fix the initial time two years later, as there are always lagging-behind effects of policy implantation.

there is a convergence existing in China and what type of convergence if there is. However, from another point of view, it implies that significant attentions are paid on the regional disparities yet few focuses on the aggregate change over time. It has also been criticized that the high governmental interventions and large share of state-owned sectors lead to short-term fluctuations of regional growth, which may be overlooked in estimation of long-run growth trend (Andersson et.al, 2013). Another argument is the small sample size of provincial-level data in China (Islam, 2003), which may result in failure of remove of short-term variations.

The regional development contains multiple dimensions such as economic, social, environmental, et al, and hence it is significant for the policy makers to learn about the comprehensive regional condition and individual effect of different factors in the overall development. From previous studies, there had been a huge amount of application of GDP as the indicator to estimate regional disparities and development, whereas caused plenty of problems in the accuracy of evaluation. Therefore, alternative approaches have been proposed containing more aspects such as well-being, life quality, etc., which initiates the desire of a composite index involving various aspects and domains that has been omitted in the GDP measurement (Kaufman, et al., 2007). Also, the vast diversities in geographic, economic, social, cultural, environmental, and et.al performances observed in different Chinese provinces and municipalities constitute as another motivation to learn more about the overall impact of local policies and identify the key factors to influence the higher-developed or lower-developed economies. Since the local policies might have spreading influences in different domains, and the different regions might have complex interactions among each other, it is highly required an evaluation method to aggregate all of those impacts.

1.3. Research structure

The major objective of this study is to comprehensively evaluate the socioeconomic development Chinese provinces and municipalities measured by a Composite Index of Regional Development (CIRD), taking the five different sub dimensions into consideration, including macroeconomic conditions, science and innovation, human capital, environmental sustainability and public facility. In the following chapters, the construction and application of CIRD will be elaborated. In the Chapter 2, I make the literature review to provide theoretical background for the study, previous studies on construction of a composite index, evaluation of regional development and estimation of convergence are summarized here. Then a step-by step framework of methodology is established in Chapter 3, in order to prepare the technique supports for the analysis. It introduces the principles in the whole procedure in sequence from data selection, normalization, reduction and aggregation, to cluster analysis and convergence estimation. With the theoretical and technical assistances, the final CIRD is constructed with the application of a large set of panel data on provincial level over 1998-2010 in Chapter 4. Each pillar as *Macroeconomic Index* (MEI), *Science and Innovation Index* (SII), *Environmental Sustainability Index* (ESI), *Human Capital Index* (HCI) and *Public Facility Index* (PFI) is set up with one principle component respectively and ranking reports are provided pillar by pillar with further analysis, and then the five sub dimensions are combined to provide the comprehensive ranking report of CIRD. After completing the CIRD, I further apply it into empirical analysis. Chapter 5 and 6 investigate the regional disparity and dynamic trajectory in China with the application of CIRD. In Chapter 5, cluster analysis is applied to group the Chinese municipalities and provinces by CIRD, cluster performances are evaluated pillar by pillar, indicator by indicator, and maps are also provided to clearly depict the grouping scenario. In Chapter 6, convergence trend is estimated by different tests in each pillar and CIRD, and dynamic trajectory of convergence is also analyzed by stochastic kernel density estimation respectively. Chapter 7 draws the conclusion. The references, acknowledgement and appendix are attached in the last chapters.

2. Literature review

There is a vast literature attempting to measure indicators from different socioeconomic dimensions with the application of a composite index by looking at a number of independent indices. From a methodological point of view, there are plenty of composite indices are constructed with the application of non-parametric methodology, where the experts define the weights of indicators or sub-indices subjectively according to their experiences and knowledge of the relative significance of the indicator or sub-index. An opposite manner to build up a composite index is parametric methodology, that is, the weights of indicator or sub-indices are determined by the relative variation among those indicators or sub-indices. Principal Component Analysis (PCA) and Factor Analysis (CFA) are two of the most common parametric methods that frequently applied to estimate the weights of indicators or sub-indices[11]. Dreher (2006) and Heshmati (2003, 2006), for example, applied PCA to estimate the weights of sub-indices of a composite globalization index. Andersen and Herbertsson (2003) used factor analysis to combine several indicators of economic integration and international transactions into a single measure or index of globalization.

This study mainly aims to build up a multi-dimensional index to measure the regional socioeconomic development in individual regions of China on provincial level. Referring to the construction of the composite index, the work on my composite index of regional development (CIRD) can be traced from the studies on development indices[12], which could be resulted in the construction of the Human Development Index (HDI)[13], and in the research focusing on linking the measurement of a quality of life with welfare

[11] PCA and factor analysis are both variable reduction techniques. If communalities are large, close to 1.00, results could be similar. PCA minimizes the sum of the squared perpendicular distances to the axis of the principal component while least squares regression minimizes the sum of the squared distances perpendicular to the x axis (not perpendicular to the fitted line) (Truxillo, 2003). Factor analysis is traditionally has been used to explore the possible underlying factor structure of a set of measured variables without imposing any preconceived structure on the outcome (Child, 1990). PCA includes correlated variables with the purpose of reducing the numbers of variables and explaining the same amount of variance with fewer variables (principal components). Factor analysis estimates factors, underlying constructs that cannot be measured directly.

[12] E.g. Nordhaus and Tobin, 1972; Amartya Sen, 1987

[13] E.g. OECD, 2006; Douglas and Wall, 1999, 1997 and 2000; Deller et al. 2001; Rudzitis, 1999; Nord and Cromartie, 1997

and regional indicators[14]. The most popular index "Human Development Index" (HDI) produced by the United Nations Development Programme (UNDP) is reported periodically to estimate the different international economies. Another famous index in the European Union (EU) is the Lisbon Strategy Indices (LSI), which is similar to HDI, reported to measure the comprehensive development level of EU member economies. There are some other indices set up by individual or private agencies, such as the composite globalization index created by the consulting firm, A.T. Kearney (2002, 2003) on the basis of variety of indicators on economics, technology, demography, and politics. The indicator selection has authenticated the alternative of various composite indices, some of which reflect macro development insofar with the combination of a range of social, economic and political factors to measure the level of development (Todaro, 1989). These studies are based on the conception that a single benchmark cannot be precise to measure development just as a single set of objectives cannot describe adequately the variety of development patterns among the world. (Wilson and Woods, 1982; Sainz, 1989). Moreover, it is possible that the integration of either composite index or special indicators is created somehow under the requirement of measuring the increasingly heterogeneous socioeconomic conditions, whereas the motivations behind actually are different. Some studies might aim at the separate evaluations of different dimensions such as social situation with independent indices, whereas some might interpret them as indicators compared with economic indicators and combined different aspects of development in measurement (Sainz, 1989). In recent years, besides social or political factors, the environmental factors have become a hot issue to consider as a significant part in the construction of a composite index. Many studies made plenty efforts to link up the economic growth with environmental sustainability. The OECD, World Bank, the United Nations, and etc. have all put a special focus on this issue, such as the OECD's environmental performance review and UN's sustainable developmental index. In a recent study of Chinese urban sustainability index, environmental sustainability was also considered as an important contribution to measure the Chinese cities (Xiao, etc., 2010).

The main focus of the studies is the valuable work of the construction of theoretical

[14] E.g. Midgley, Hodge and Monk, 2003; Hagerty et al. 2001; Noll, 2002; Bryden, 2003

structure to measure development, so the considerable overlap cannot be avoided. Moreover, not all these efforts moved from theory to practice, and most of the empirical application paid more attentions on the national-level measurement or the developed economies. The measurements on developing economies, especially with consideration of environmental factors, are not very common, and usually take directly the indicators from the measurement of developed economies, which are actually not suitable in some degree since their developmental patterns, are totally different. Therefore, it is not quite available to measure the regional development and differences comprehensively in China with the current measurements, which makes the requirement of a composite indicator most acute.

The application of a composite index has been widely documented. Normally it is used to classify regions with their similarities into several heterogeneous groups. Cluster analysis has become one of the most popular approaches to benchmark regional development for the researchers and practitioners. It has been widely used in social science such as marketing, industry, economic and even scientific aspects. Since the theory of cluster has become increasingly prominent in spatial studies, it has been mainly adopted at national, regional or sub-national level. The supply-oriented methods have been commonly used at national or regional level since 1970s (OECD, 1999), while a new trend is showing up in infrastructure and technology aspects at regional level. It has been widely used in European Union regions, many states in US, Asia-Pacific countries, etc., such as an estimation of the technical change and innovation in OECD countries and United States, dividing them into knowledge-based economies and learning economies (OECD, 1999. Boosting innovation: the cluster approach, OECD proceedings, Published: France); specific studies such as the estimation of regional competitiveness in a case study of East German (Kronthaler, 2003), evaluation of agricultural and rural changes in EU members with case studies in Serbia and Hungary (Monasterolo, 2010, 2011). Fanfani R. and Brasili C. (2003) used the cluster analysis to group rural China on provincial level by their socioeconomic conditions from the first agricultural census of 1996.

The concept of development has for long been associated to that of economic growth. In this perspective the variable used for comparative analysis has been the GDP per capita. One of the crucial debates in the economic growth literature has been whether or not countries with lower initial per capita level exhibit higher growth rates than the richer counterpart. Convergence theory is one of the hottest issues in the macroeconomic empirical studies. Theoretically, there were different estimations of convergence from neoclassic economic growth theory and endogenous growth theory. Empirically, the convergence study could contribute to the elimination of polarization and narrow down the gap between the rich and poor. The convergence theory focuses to three basic conceptions: absolute convergence, conditional convergence and club convergence. These three conceptions can be concluded by the following questions: the first is whether the income/economic gap is permanent between the observations, and the second is whether it is caused by the differences of economic systems or initial conditions of the observations if the permanence exists. If the answer to the first question is negative, there is absolute convergence; if the answer is positive, then the reason to cause this permanent gap is the key point. We claim the conditional convergence when the reason is the heterogeneous economic structures, whereas the club convergence exists when it is caused by the combined factors of heterogeneous economic structures and initial conditions, or else to say, when two regions have the same economic structures, they may also converge to different equivalences due to their different initial levels, which we claimed as club convergence.

Three common methodologies are applied to study convergence: the first one is regression, referring to cross-sectional regression (Baumol, 1986; De Long, 1988; Barro, 1991; Mankiw, et al., 1992) and panel data regression (Isalm, 1995; Caselli, et al., 1996; Lee, et al., 1998). The parameter of initial level smaller than zero signifies absolute convergence when there is only initial level in the regression without other controlling variables, whereas it is conditional convergence if taking other controlling variables into consideration. This methodology however cannot depict the convergence trend very precisely because of Galton’s Fallacy. The second methodology is dynamic distribution estimation (Quah, 1996; Bianchi, 1997; Paap and van Dijk, 1998; Anderson, 2004; Juessen, 2007; Maasoumi et al., 2007). Stochastic kernel density is usually applied in this methodology to depict how the different observations perform in different times. It analyzes the distribution shape from an integral aspect, as whether how the distribution evolves dynamically and whether it is a uni-mode shape or multi-mode shape distribution. The third methodology is the time-series method (Bernard and Durlauf, 1995, 1996; Evans, 1998; Hobijn and Franses, 2000), mainly focusing on the study that whether the gap of average income between different regions will disappear. The common methods are panel unit root test, co-integration test, and etc.

The distribution methodology is used more commonly EU. Tsionas (2002) applied Markov chain to analyze the regional economic development in Greece based on the NUTS-3 data, indicating a clear tendency of economic polarization in Greece during 1971 to 1993; Gallo (2001) also applied Markov chain to estimate the club convergence in the 138 European regions in the time interval 1980—1995. Plenty of researches on the convergence estimation can also be documented in China. It is commonly accepted that there is no absolute convergence in China with the regression methodology, but there are still conflicting arguments about the existence of conditional convergence and club convergence (Chen, 1996; Weim 1997; Cai and Du, 2000, Lin and Ma, 2002). Cai et al. (2000, 2001) concluded no national-level convergence in China during 1978 to 1998, whereas the club convergence was formed among different economic zones. Liu (2001) pointed out there was a club convergence inside east-coast, inland and western regions, but it showed a divergence tendency inside south and north regions. Shen and Ma (2002) also certified the existence of club convergence among east-coast, inland and western regions. Lin (2003) indicated it showed a conditional convergence during 1981-1999 with an annual converging rate of 5% to 7%. Tan (2004) infers an annual converging rate of over 2.2% during 1978-1990, whereas a divergence tendency between 1991 and 1999, and the club convergence only existed inside the low-level income group and high-level income group. Dong (2004) discovered an obvious conditional convergence with an annual rate of 9.6% during 1985-2002. Xu (2004) observed both σ-convergence and β-convergence in the prefectural level data. Peng (2005) assumed that there was conditional convergence in western regions with an annual rate of 7.3%, whereas club convergence showed in east coast with an annual rate of 1.12%. Xu and Li (2006) also confirmed the existence of conditional convergence with an annual rate of 7.6%. He (2008) developed this conclusion with four convergence clubs for GDP per capita and three convergence clubs for labor productivity. Tan (2008) estimated a clear club convergence with three different time intervals, five provinces located in east coast were included in the highly developed group but excluded Beijing and Shanghai, which was different from other studies. Liao (2011) indicated β-convergence was not obvious but club convergence became increasingly noticeable with a long time-series data of 1952-2009. However, Ma (2003) doubted the existence of convergence and considered a divergence tendency with an annual rate of 1.2%-2.1% during 1981-1999. Liu et al. (2004) questioned the validity of club convergence in eastern, middle and western regions, and Wang (2004) suspected the conclusion of conditional convergence among national level in China. From the previous studies, it is clear the club convergence is becoming more and more significant in recent convergence research of China. However, various methodologies were applied and different time intervals were selected with rare explanations. The Chinese scholars are more apt to apply the traditional methodology to analyze the convergence trend; instead the distribution methodology is not very common used.

3. Methodology: a step-by-step framework

When referring to the construction of composite index or indicator, there are enormous methodological difficulties, not only restraining the construction of such an index, but also preventing its further application. Here I applied a step-by-step framework to explain the construction of CIRD theoretically, statistically and empirically. This part includes the methodologies of indicator selection, data computerization, the extraction of composite index, cluster analysis and convergence estimation. First, I collected a large data pool of indicators, referring as many aspects of socioeconomics as possible to reflect the 30 Chinese provinces and municipalities[15]; then I tested the correlation significance of those indicators and selected the important ones. After setting up the sample dataset, I normalized the data with strong skew in order to better process the following steps of indicator extraction. I applied Principle Component Analysis (PCA) to extract the component from the selected indicators by each dimension, and then used simple standardized statistical methods to get the reports of CIRD. When the construction of CIRD is completed, I applied it to further empirical analysis. Firstly, cluster analysis was used to group the different Chinese provinces and municipalities and benchmark them in the map, and then convergence estimation was processed to depict the dynamic trajectory of socioeconomic development in China.

3.1. Indicator selection and data collection: theoretical basis and statistical method

OECD defines indicators as "… data or combination of data collected and processed for a clearly defined analytical or policy purpose" (Le Gallic, 2002). Since observing the main determinants in regional socioeconomic conditions has always been very comprehensive and complicated, it always requires a fully understanding of the main policy issues and the complexity of local interactions. Before setting up a composite index, several basic questions should be raised up. The first is whether it is possible to

[15] Tibet was excluded because of the missing data (there are no data reported in the pillars of SII and ESI over the whole time interval).

objectively measure the comprehensive socioeconomic conditions beyond GDP in different regions of China, and make comparison across individual regions. If the answer is affirmative, how large the regional disparities are across different regions in China; whether it is possible to observe the general performance of individual changes over years; which regions are the leaders and which are the laggings in terms of the comprehensive socioeconomic conditions; and what factors contribute more to the local development. A good composite index could give reasonable answers to the questions above, that is the key point to be kept in mind when building up the composite index of regional development (CIRD) in China. Moreover, several main characteristics have to be considered in the construction of CIRD, containing comprehensiveness, multi-dimensionality, availability to indicator reduction and extraction of a synthetic measure (Saisana and Tarantola, 2002; OECD 2005). It is documented that a composite index should satisfy the following conditions (Hagerty, 2001; OECD, 2005):

- A solid theoretical framework should be set up as the basis of a composite index;
- An empirical analysis with the investigation of whole structure of selected indicators is required, two dimensions of sub-indicators and regional units in the dataset are needed at least;
- A single dimension can be reported from the index, which can also be divided into individual domains or components with significant but discrete portions and high relevance to the majority, and their weights and aggregation should be valued;
- The tests of robustness and sensitivity of the index should be accessible;
- The index should be transparent and noticeably related to other measures;
- Timeliness is necessary to admit periodic observation and aggregation;
- Data normalization is required to solidify the comparability.

Supplementary criteria applicable to policy analysis was proposed by Kaufmann, et al. (2007), including frequency which is possible to calculate in line with the project requirement; objectivity that means observing the index with the least subjectivity; transparency as the measurement could be applied by other studies; simplicity that it had to be easily understood; comparability that it could be compared across regions and countries; and dynamics that measured changes over time. Generally speaking, an ideal

CIRD should measure multiple domains with a complete coverage of all significant dimensions in regional development seldom captured by partial indicators, such as economic conditions, investment, trade, employment, human resources, infrastructures, environment protection, science and technology, et.al., meanwhile, all of the above domains could be aggregate into an unified index with solid weights or parameters. Based on this general theoretical framework and previous literature, and according to the understanding of regional development in China, I set up five pillars (sub-indices) to demonstrate the CIRD, referring to *Macroeconomic Index (MEI), Science and Innovation Index (SII), Environmental Sustainability Index (ESI), Human Capital Index (HCI) and Public Facility Index (PFI)* to evaluate the policy objectives of Five-year Plan. Before going into the indicator selection, I decided to define the panel data ranged from 1998 to 2010 under these five dimensions. The reason I chose the time period from 1998 to 2010 is mainly based on the governmental guidelines “Five-year plan”[16]. As discussed above, the socioeconomic development vastly depended on the policies. Therefore, it is reasonable to divide the whole time interval into three sub-periods 1998-2000, 2001-2005 and 2006-2010 according to the three “Five-year plans” enacted in those years.

Besides the theoretical framework discussed above, statistical method is also applied to build up a more practicable CIRD. An essential prerequisite is the high quality of the dataset. The variable availability, significance, analytical soundness, and timeliness should be considered as basis to select the indicators. There are different sources that provide territorial data on China’s provinces. In order to guarantee the consistency and authorization, data have been collected from annual series Chinese statistical yearbook that published by the National Bureau of Statistics of P.R.C., so it is reliable and comparable. In my study, I firstly collected enormous indicators as many as available during the fixed time interval (1998-2000)[17], trying to cover as more socioeconomic

[16] The “9th Five-year plan” was from 1996 to 2000, the “10th Five-year plan” was from 2001-2006, and the “11th Five-year plan” was from 2006 to 2010. However, I here started from 1998 due to the missing data reported in the pillar of environmental sustainability in the first two years 1996 and 1997 in the China national statistic yearbooks.

[17] A large indicator pool was set up initially, 22 original indicators were collected in MEI pillar, 21 original indicators were collected in SII pillar, 18 original indicators were collected in ESI pillar, 14 original indicators were collected in HCI pillar, 15 original indicators were collected in PFI pillar. Some of those indicators were calculated primarily from an even larger database. Here I also used PCA in the original datasets of each pillar. In MEI, SII and HCI, four principle components were extracted; in ESI and PFI, five components were

aspects as possible. An outlier (Tibet) was disregarded in two sub pillars and the aggregated CIRD. After the data collection, I checked the correlation and significance of each indicator with GDP growth rate and disregarded the non-significant indicators. Another requirement for the following analysis is to keep the positive relations of each indicator to easily interpret, for example, the larger value of one indicator in CIRD the higher level of local socioeconomic development. In that case, potential reversion might be applied if necessary[18]. In order to run the following-up multivariate analysis, here it is better to keep one latent extracted from each pillar, hence the indicators disturbing the common latent dimension will be removed. I chose the first five indicators that are most significantly correlated to GDP growth rate in each dimension. From Figure 1, it is clear that the average CIRD shows a significant correlation with the average GDP growth rate, hence we can certify this index is valid to proxy for the socioeconomic development of the Chinese provinces and municipalities.

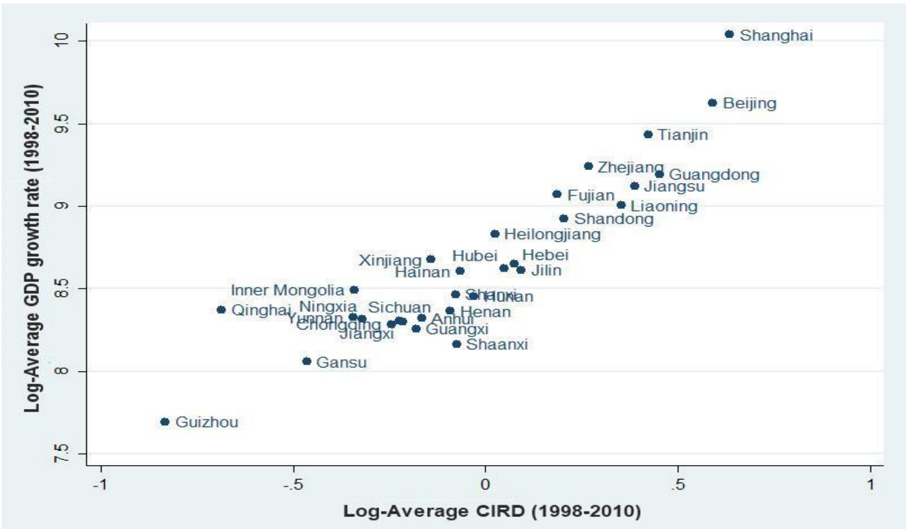


Figure 1 *- Correlation of Average CIRD and GDP Growth Rate (1998-2010)*
Source: Author's calculation based on the data from Chinese National Statistical Bureau

Generally speaking, the CIRD is designed to estimate the local development over time in Chinese municipalities and provinces by a set of socioeconomic categories. The five-dimension (five-pillar) of CIRD constitutes 25 individual indictors that are critical to

extracted. I also observed the indicators which played significant parts in the components of each pillar and finally decided the five indicators for each dimension.

[18] In the final dataset of the indicators, the indicators of final consumption rate, employ compensation portion and discharge of living waste per capita are reversed to run the following analysis, as they all have negative relations with the GDP growth rate.

measure comprehensive situations in local socioeconomics.

Table 1 *- The indicators of five-dimensional CIRD*

Macroeconomic Index (MEI)	
GRP per capita	GRP per capita (deflated) (10,000 Yuan)
FDI	proportion of foreign direct investment in GRP (%)
trade balance	trade difference of export-import over total trade value (%)
consumption	final consumption rate (%)
compensation	employ compensation proportion over GRP (%)
Science and Innovation Index (SII)	
government expenditure SII	proportion of government expenditure in science and innovation over total government expenditure (%)
labor productivity in high S&T firms	average labor productivity of high science and technology enterprises (deflated) (10,000 Yuan per worker)
income per capita in high S&T firms	average income per capita in high science and technology enterprises (deflated) (10,000 Yuan per worker)
R&D expenditure level	R&D expenditure proportion over GRP (%)
trade level in S&T market	trading value in science and technology market over GRP (%)
Environmental Sustainability Index (ESI)	
industry waste water cleaned	proportion of discharge of industrial waste water meeting standards (%)
utilization industry waste	utilized ratio of industrial waste (%)
output from waste	output value of products made from waste gas, water and solid wastes per worker (deflated) (10,000 Yuan)
living waste	volume of living waste discharged per capita (ton)
industry anti-pollution	investment of industrial enterprises on anti-pollution projects per worker (deflated) (10,000 Yuan)
Human Capital Index (HCI)	
literate proportion	proportion of literate population (%)
employ 3rd sector	employ ratio of 3rd sector over 1st sector (%)
working age distribution	working age population distribution (%)
life expectancy	life expectancy (estimated by the population census)[19]
urban proportion	urban population proportion (%)
Public Facility Index (PFI)	
government expenditure in public service	governmental expenditure proportion in public services over total government expenditure (%)
public vehicles	public transportation vehicles per 10,000 persons (unit)
paved road	area of paved road per 10,000 persons (sq.m)
education	public schools per 10,000 persons (unit)
medical health	public medical services per 10,000 persons (unit)

Source: author's collection and analysis

Five indicators are allocated to describe each dimension, which is a

[19] The life expectancy is calculated by the national population census reports 1990, 2000 and 2010.

pre-determination for the possibility to apply the equal-weighting method in the following aggregation step (The detailed reason and theory of application of this equal-weighting method will be elaborated in the following part aggregation methodology (3.4). The detailed indicators are showed in the Table 1.

3.2. Data normalization: Box-Cox transformation

Box-Cox is a parametric transformation technique, aiming at incongruity reduction. The dataset that is non-additive or non-normalized is required to be transformed in order to get a normal distribution with a common variance. The initial data transformation can be traced back to the studies of Tukey (1957), which provided a monotonic function of observations and expressed as

$$\mathcal{Y}_i^\lambda = \begin{cases} y_i^\lambda & (\lambda \neq 0) \\ log\, y_i & (\lambda = 0) \end{cases} \qquad (3.2.1)$$

It has been modified by Box and Cox (1964), who considered the incoherence when $\lambda = 0$ and represented as

$$\mathcal{Y}_i^\lambda = \begin{cases} (y_i^\lambda - 1)/\lambda & (\lambda \neq 0) \\ log\, y_i & (\lambda = 0) \end{cases}$$

$$where\ \mathcal{Y}_i^\lambda = \left(y_1^\lambda, y_2^\lambda\ \ldots y_n^\lambda\right)' = Y\theta + \varepsilon \qquad (3.2.2)$$

here Y is the matrix of known constants, θ is the vector of the unknown parameters of the transformation values, and ε is the vector of random errors. However, the second equation is only sound when $y_i > 0$, so it has to be modified if the values of the observations are negative. A substitute measure had been proposed by Manly (1976) to deal with negative observations. It is declared as effective in symmetrizing the skew distributions and expressed as:

$$\mathcal{Y}_i^\lambda = \begin{cases} (exp(\lambda y_i) - 1)/\lambda & (\lambda \neq 0) \\ y_i & (\lambda = 0) \end{cases} \qquad (3.2.3)$$

The maximum likelihood and Bayesian methodology was proposed by Box and Cox (1964) in order to estimate the parameter λ. The Box-Cox transformation has been used commonly in empirical practices to make the data linear and normalized. It is a hot topic of testing the consistence of the transformation parameter of Box-Cox and the hypothesized value. Even though there has been plenty of debates and arguments, it is clear that the Box-Cox transformation is more practical in the empirical determination of functional relationships in a variety of fields, especially in econometrics (Sakia, 1992).

Since normalization of the data is necessary for the comparability as discussed above, in my study, I first checked the skew of the indicators, and then applied Box-Cox transformation to normalize the data as following function:

$$\begin{cases} x' = \ln(x) & \text{when it is positively skewed} \\ x' = \dfrac{x^{\lambda} - 1}{\lambda}, (\lambda \geq 2) & \text{when it is negatively skewed} \end{cases} \tag{3.2.4}$$

In order to avoid the extreme values, Tibet was excluded in the following analysis due to the missing data and extreme low values of the data that could be available.

3.3. Extraction of CIRD: PCA

After the data processing, Principal Component Analysis (PCA) was applied to extract the component scores, with the combination of the samples in each time intervals (1998-2000, 2001-2005, and 2006-2010). PCA is a data reduction method, which aims to summarize and order the information in a large data set, and hence to avoid double counting. It shows the information content carried by each of the selected variables and the amount of information shared among the variables. What has to be paid attention to is that the higher the correlation among the initial variables/indicators, the smaller the number of artificial variables we need to collect to precisely describe the objectives. If the correlation is significant, we can directly run PCA and comment the results on a smaller dataset of new artificial (latent) variables; otherwise, we have to count for a loss of information and the introduction of an error term, due to the exclusion of some variables from the analysis if the correlation is not significant. The reasons of choosing PCA are as following: first, it helps us to synthesize the information content of several variables in order to get a reduced number of indicators available to give an interpretation and hierarchy the main characteristics of a selected area, at a disaggregated level; second, it helps us to evaluate the impact of policy measures on a pre-determined and limited area, however, sometimes the sample data quality is not high enough to run the regressions; third, it helps us to pursue the main features of the structural changes occurred on a selected area and show results easy to be understood and commented by policy makers.

Hotelling (1993) introduced the most famous formulation of PCA, while the methodology itself comes from Pearson (1901). Since the objective of PCA is to maximize the variability explained by the components, the total variability of the p extracted components equals the total variability of the p original variables. It permits easier selection of a sub-set of components (Mario Mazzocchi, 2008). The premise of running PCA is that the original variables/indicators applied in the dataset are highly correlated, hence it is possible to identify a reduced set of uncorrelated linear combinations from these initial correlated variables, and the principal components are able to describe the most of the initial variability. It is reasonable to compute the PCA both on the covariance and on the correlation matrix. The latter is usually chosen in order to avoid the distorting influence of the indicators with higher variance in the extraction[20]. Once the values of components are obtained from the correlation matrix, the scores of each statistical unit (municipalities, provinces or etc.) for each component can be calculated. The principal components come from the following linear combinations, expressed as a matrix:

$$Y = XA \qquad (3.3.1)$$

where Y is a $n \times m$ matrix (m is number of principle components), and the scores are calculated by the n statistical units in the m components; A is the vector matrix $p \times m$ of the normalized coefficients; X is the $n \times p$ matrix of the standardized data. The eigenvalue of each principle component equals to the variance of its component score. The standardized scores will result in that all principle components will have the same variance, in other words, they will carry the same level of information content and also the same weight.

From the mathematical point of view, PCA is a mathematical technique instead of a statistical method. The calculation for principle components is consecutive, which means it is necessary to find the dimension containing the most of the variance. For example, in the linear combination with the highest variance, the first principle component can be expressed as:

[20] The correlation matrix is used when we the original variables we have to deal with have different measurement units. Therefore, a 'standardization' of the original indicators takes place.

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \tag{3.3.2}$$

here a is coefficient vectors, X is original variables, Y_1 is the first principle component. After obtaining first principle component, the search of the second principle component Y_2 will focus on the defined orthogonal sub-space of the linear combination with the definition in that direction carrying the most of variance, which can be expressed as:

$$Y_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \tag{3.3.3}$$

and so forth. The vector indicates to maximize the related variance of the principle component, for example, the a_1 vector maximizes the variance of the first principle component Y_1. The function can be stated as a formula of a_1 and M matrix:

$$\begin{aligned} &Var\,(Y_1) = a_1'\,\mathcal{M}\,a_1 \\ &s.v. \parallel a_1 \parallel = a_1'a_1 = 1 \end{aligned} \tag{3.3.4}$$

here M is the matrix of variance-covariance of the original variables, a_1 is the coefficient vector of Y_1, and a_1' is the one transposed from a_1. If the Lagrange multipliers are used to find the max and min of the function, it follows as:

$$\begin{aligned} &max\ [a_1'\mathcal{M}a_1 - \mathcal{L}_1(a_1'\,a - 1)] \\ &s.v.\ a_1'a_1 - 1 = 0 \end{aligned} \tag{3.3.5}$$

here L_1 is the Lagrange multiplier, M is the matrix of variance-covariance of the original variables and a is the vector of coefficients. If we determine the partial derivations as 0 for the elements of a_1, it follows:

$$\mathcal{M}\,a_1 = \mathcal{L}_1\,a_1 \tag{3.3.6}$$

If the system is homogeneous, then the equation can be expressed as:

$$(\mathcal{M} - \mathcal{L}_1\ \mathcal{I})a_1 = 0 \tag{3.3.7}$$

here $\mathcal{I}$ is the identity matrix. The solution $a_1 = 0$, which indicates that the first principal component is 0, cannot be allowed in the system, hence it is required the determinant of the matrix as 0.

$$\mathcal{D}\,(\mathcal{M} - \mathcal{L}_1\ \mathcal{I}) = 0 \tag{3.3.8}$$

The equation of PCA can be expressed as follows:

$$C = XA' = \begin{cases} c_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p \\ c_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p \\ \cdots \\ c_p = a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_p \end{cases} \tag{3.3.9}$$

where C is the $n \times p$ matrix of principle component scores, X is the data matrix, and A is the $p \times p$ matrix of component loadings. Once the matrix A has been computed, the component scores can be calculated as follows:

$$\hat{a}_{ij} = \frac{a_{ij}}{\sqrt{\lambda_j}} \tag{3.3.10}$$

here λ_j is Eigen values. The number of principle components extracted from the original dataset is usually determined by the researcher, and the common methodologies are on the basis of the variance level explained and eigenvalues. Generally speaking, it is more valid to take the components with an explanation of over 60% of the total variability[21], or we can use the Kaiser rural to consider the principal components with eigenvalues higher than 1[22]. Another pictorial method to determine the number of principle components is relied on the scree-plot, where the number of principle components is reported on the axis of x and the eigenvalues are represented in the order of extraction on the axis of Y. It is valid to choose the number of components at which reached the knee point of the curve.

Two approaches can be applied in running PCA. After insertion of all the selected indicators and variables, those variables and indicators are gathered in homogeneous blocks such as economic, agricultural, demographic, etc. PCA can also be applied to each block. These two approaches give back similar results if the correlation of variables in different blocks is not significant, which has to be tested carefully and, in general, is difficult to obtain. Generally the first option is commonly used. Considering the theoretical requirement discussed before, that weighting and aggregation should be paid special attention to, I will apply both methods in this study to construct CIRD and combine their results to better measuring China specifically and comprehensively. First, I

[21] This threshold is merely indicative, as to determine the number of components, a good comprehension of the examining objective is also significantly required.

[22] Note: the eigenvalues corresponding to each component represents the amount of variance they explain; the sum of eigenvalues equals the original number of variables, therefore a lower level of eigenvalues would explain less than the standardized variance which should be equal to 1.

will apply PCA in each separate dimension that I selected and defined before in the three time periods[23], hence to give a specific description for each dimension. Then I will gather all the indicators together and run PCA in general, hence to get a comprehensive picture of regional development in China, but here I will not discuss about the implication of the component since I have already divide the initial variables into different blocks, which is reasonable and can be documented (OECD 1994, 1996; FAO 2008; Fanfani et al. 1999). In this study, the components are computed both on covariance and correlation matrices. In order to describe the regional development more specifically, the first component was extracted in each pillar (here we considered the first component score should be over 50% when we extracted it). From the scree plot of PCA, it is also judicious to extract the first components in each dimension.

3.4. Aggregation methodology: multivariate analysis and parametric method

The aggregation of the multidimensional indictors to a unique dimension has always been a common issue in the construction of a composite index. In this study, I first apply an equal weighting method to the five pillars, used in many previous analysis, which can be documented in various literature (Huovari, et al, 2001; Huggings and Izushi, 2008; UNDP, HDI index). This step here is under the multivariate analysis. Then I follow a parametric method trying to assign the weights to the principle components, in order overcome the potential subjectivity in the previous aggregation, based on the methodology from the study of a composite index of economic integration in Asia-Pacific (Chen, 2008).

Multivariate analysis is applied to authenticate the uniformity of the interior data inside each dimension. As discussed above, the composite index is set up with the consideration that each pillar is invented to picture a special latent aspect combined with different indicators in that dimension. Therefore, there have to be high correlation inside each dimension to correspond with the expected characteristic of the composite structure, which suggested a unique latent component lying beneath each dimension. The data reduction technique PCA is applied as described before to complete this requirement, clearly picturing a unique latent aspect in each dimension. Generally speaking, the core of

[23] I assume that the component score is consistent under the macro policy background (Five-year plan) in each time intervals.

multivariate analysis is the existence of the unique single latent dimension extracted from the indicators contributed in an equal portion in each sub pillar (OECD, Annoni and Kozovaska, 2010). It verifies the simple choice in my study to choose equal weights for the five pillars to calculate the CIRD in a linear combination of their component scores if a sole principle component is extracted in each dimension, since I selected equal number of indicators in each sub pillar.

However, this multivariate analysis might be argued as non-parametric, which directly determined weights for the indicators or component according to the researcher's subjective opinion for their relative significance. In order to avoid this subjectivity, I then applied the parametric method to assign weights for individual indicators to get the final CIRD report, in the case of running all the indicators without division by pillars in PCA. In this case, the PCA is processed under the 25 indicators all together, without the division of sub dimensions. One parametric method proposed in the previous documents assumed that a frame existed beneath the variation of indicators, so the weights of each individual indicator could be determined by the covariance among them. The calculation of the weights was relied on the previous analysis of PCA (Chen, 2008), and expressed as:

$$\mathcal{W}_j = \frac{\sum_{i=1}^{i=p} \lambda_i a_{ij}}{\sum_{i=1}^{i=p} \lambda_i} \qquad (3.4.1)$$

here W_j represents the weights of individual indicators, λ_i represents the i^{th} eigenvalue, that is the variance the principal component; and a_{ij} represents the i^{th} eigenvector of the correlation matrix. However, this method was argued by the author that the weights might be biased if some indicators were highly inter-correlated. Other methods are also applied variously in different studies. For example, the weights are assigned on the basis of the significance by the researcher, such as EU regional competitiveness index (Annoni and Kozovaska, 2010), or the global competitiveness index by the World Economic Forum, whereas it is more relied on the researchers' experiences. Recently, the econometric methodology on the basis of regression has been emerged in the aggregation procedure, such as the construction and application of rural development index in Poland and Slovakia (Michalek and Zarnekow, 2012). This method seems to be very appealing, however all under a large sample number to guarantee the robustness of the regression, which cannot be satisfied with the small sample size on provincial level of data in China. Due to all the limitations of the parametric methods above, in this study, I will assign the contribution rates of the principle components as their weights, which is a basic and common technique in the PCA theory to calculate the CIRD, and the function can be expressed as:

$$\mathcal{Z} = \sum_{j=1\ldots n}^{i=1\ldots p} w_i \cdot (PC)_{ij} \tag{3.4.2}$$

here Z is the total score of principle components, w_i is the weight getting from the contribution proportion of the principle components, PC_{ij} is the component score.

3.5. Ranking reports: Min-Max method

Min-Max is a normalization technique that can be documented in previous study of construction of composite index. In the EU's regional competitiveness index, it is used to report the rankings of EU's cities on NUTS-2 level (OECD, Annoni and Kozovaska, 2010). After the principal component analysis (PCA) used to extract the first component score, with remain of all the first components over 50%, I applied Min-Max to standardize the component scores and get the ranking.

$$x^* = \frac{x - \min}{\max - \min} \tag{3.5.1}$$

All these methodologies are introduced step-by-step in the order of practical application. To summarize, I first selected the indicators based on the theoretical framework and statistical significance and positive correlation with the GDP growth rate. Five sub-pillars were assigned and meanwhile, the time range was determined according to the policy background and sub-periods were divided according to the "Five-year plan" enacted by Chinese government in every five years. Then Box-Cox transformation was used to normalize the data. After data normalization, PCA was applied both in the sub-pillar level and the whole indicator dataset to extract the latent dimension beneath the various indicators. To better interpret the comprehensive socioeconomic conditions, I aggregated the PCA results from two different methods: one is based on the multivariate analysis to give a balancing weight to each dimension, while the other is based on parametric method to calculate the weights of individual indicators. The former will be applied in the PCA on sub-pillar level, and the latter will be applied in the PCA of the whole indicators together. After getting the unique index from the previous analysis, Min-Max transformation was employed to get the ranking reports of the 30 Chinese provinces and municipalities, both on sub-dimensional level and the integrated five-dimensional level.

3.6. Grouping and mapping China: cluster analysis

Cluster analysis is a statistical method to classify observations into groups by their similarity. The reasonability of grouping is determined by several technical selections which are the complicated part in the application of cluster analysis. The first and fundamental step in the application of cluster analysis is the variable selection. Three basic rules are required to detect whether the variables are appropriate, as inductive, deductive and cognitive (Ketchen et al., 1993). The inductive rule is sort of based on a pre-consideration of a wide range of variables to cover as complete as possible (McKelvey's, 1975, 1978) in order to find out better exploratory classification of observations, which means that a maximum likelihood to investigate the significant differences is expected from the application of the variables. Deductive rule actually implies the necessary of relevant variables, since cluster analysis aims to extract the most internally stable groups across all variables. The irrelevant variables will decline the validity (Punj and Stewart, 1983). The cognitive rule involves a theoretical background to define the variable conceptions and requires the researcher’s own experiences and knowledge. Another problem that has to been confronted in the application of cluster analysis is the multi linearity. If a correlation of variables is too high, there might be a problem to overweight one latent dimension. In this case, an approach of Mahalanobis distance was suggested (Hair, et al. 1992) to both standardize variables and correct their high correlations. This problem is solved in my study since I applied the PCA before to reduce the original variables and create uncorrelated factors. It is also a classic way suggested by the previous researches (Punj and Stewart, 1983). Here it has to be mentioned that the principal components analysis has to be orthogonal rotated for the following application of cluster analysis. The rotation function is expressed as:

$$(PC) = \frac{a_{ij}}{\sqrt{\lambda_i}} \qquad (3.6.1)$$

here a_{ij} is the eigenvectors and λ_i is the eigenvalues of the principle component (PC). In the process of cluster analysis, the choice of clustering algorithms is very essential. Two basic algorithms of cluster analysis are hierarchical and nonhierarchical respectively. The hierarchical algorithm is based on the construction of a tree-like structure by either adding individual elements to the clusters, such as agglomerative algorithm, or deleting them from clusters, such as divisive algorithm (Ketchen, et al., 1996). In the agglomerative algorithm, the most common methods are single linkage, complete linkage, average linkage, centroid method, and Ward's method (Hair et al., 1992), that are different in their mathematical procedures of the distance calculation by their own systematic biases for grouping members. For example, the centroid method has a tendency of creating irregularly-shaped clusters, and hence it is only applied when the data is interval or ratio (Hair et al., 1992); Ward's method, however, has a tendency to make a balanced member size among clusters (SAS Institute, 1990) and its results usually tend to be biased by extreme values or outliers (Milligan, 1980). Ketchen (1996) proposed a match table between the algorithm selection and the underlying data structure, such as sample size, distribution of observations, and types of variables (nominal, ordinal, ratio, or interval). He suggested the centroid method should be applied only under two conditions, the first as the data were interval or ratio scales and the second as clusters were expected to be heterogeneous among each other; Ward's method should be applied also under two premises as when an approximately equal member size of clusters was assumed and no outliers in the original dataset. Nevertheless, the divisive method is not very commonly applied in social science. There are also two types of divisive techniques, the monothetic technique with the application of binary variables and polythetic with the opposite (Everitt, 1980). Nonhierarchical algorithms usually referred to K-means or iterative methods, with a division of the data set into clusters. It is normally used in the large sample data, which cannot be satisfied in my study. Therefore, I chose the hierarchical algorithm.

In the application of hierarchical algorithm, several problems have to be confronted usually by the researchers (Ketchen, 1996). The first is the selection of the most reasonable algorithm, the second is the modification of poor cluster assignments, and the third is the unstable results when dropping some cases. To better deal with the above problems, I will explain the procedures of the hierarchical methodology step by step. Assuming that there are *n* samples and *p* variables, the data matrix is:

$$X = \begin{array}{c} \\ X_1 \\ X_2 \\ \vdots \\ X_n \end{array} \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & \cdots & x_p \end{array} \\ \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \end{array} \tag{3.6.2}$$

where $x_{ij}\,(i=1,\cdots,n;\ j=1,\cdots,p)$ is the data of variable *j* that observed in the sample *i*. The similarity between two samples can be described by *distance*, and there are several basic methods to calculate the *distance*:

(i) Minkowski distance

$$d_{ij}(q) = \left(\sum_{a=1}^{p} \left| x_{ia} - x_{ja} \right|^q \right)^{1/q} \tag{3.6.3}$$

When q=1, $d_{ij}\,(1)$ is absolute distance; when q=2, $d_{ij}\,(2)$ is *Euclidean distance*; when q=∞, $d_{ij}\,(\infty)$ is *Chebyshev distance*. Here I use *Euclidean distance* as:

$$d_{ij}(2) = \left(\sum_{a=1}^{p} (x_{ia} - x_{ja})^2 \right)^{1/2} \tag{3.6.4}$$

for a generic number of co-ordinates, where is the measurement of the *j-th* variable on the *i-th* observation. All of the data are processed by the statistic software SPSS.

The *Minkowski distance*, especially the *Euclidean distance* is the most commonly used technique. However, there is a premise to use *Minkowski distance*. If the observations are significantly different, the data has to be standardized first to authorize the application of *Minkowski distance*. Also, from statistical point of view, the application of *Euclidean distance* requires uncorrelated component in one vector, and they contributed equally in the *Minkowski distance*. A technique of weighting coordinate can be used to approve this application, for example, the coordinate setting of Q is fixed, while P is unfixed with independent changes, then their statistical distance can be expressed as:

$$d(P,Q) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}} + \cdots + \frac{(x_p - y_p)^2}{s_{pp}}} \tag{3.6.5}$$

here s_{11}, s_{22}, s_{pp} are the variances, and the weights are $k_1 = \frac{1}{s_{11}}, k_2 = \frac{1}{s_{22}}, \cdots, k_p = \frac{1}{s_{pp}}$,

so when $y_1 = y_2 = \cdots = y_p = 0$, the distance is from P to origin O, and when $s_{11} = s_{22} = \cdots = s_{pp}$, the distance is *Euclidean distance*.

(ii) Mahalanobis distance

If Σ is the correlation matrix of the variable, then:

$$\Sigma = (\sigma_{ij})_{p \times p} \tag{3.6.6}$$

where: $$\sigma_{ij} = \frac{1}{n-1} \sum_{a=1}^{n} (x_{ai} - \bar{x}_i)(x_{aj} - \bar{x}_j) \quad \text{i, j} = 1, \cdots, \text{p},$$

$$\bar{x}_i = \frac{1}{n} \sum_{a=1}^{n} x_{ai} \qquad \bar{\mathrm{x}}_j = \frac{1}{n} \sum_{a=1}^{n} x_{aj} \tag{3.6.7}$$

If Σ^{-1} exists, then the *Mahalanobis distance* between two samples can be expressed as:

$$d_{ij}^2(M) = (X_i - X_j)' \Sigma^{-1} (X_i - X_j) \tag{3.6.8}$$

here X_i is the vector composed by the p variables (components) of sample X_i, the same to X_j. The *Mahalanobis distance* from X to overall G is:

$$d^2(X, G) = (X - \mu)' \Sigma^{-1} (X - \mu) \tag{3.6.9}$$

here μ is the mean vector of the total and Σ is the correlation matrix.

(iii) Canberra distance

$$d_{ij}(L) = \frac{1}{p} \sum_{a=1}^{p} \frac{|x_{ia} - x_{ja}|}{x_{ia} + x_{ja}} \qquad \text{i, j} = 1, \cdots, \text{n} \tag{3.6.10}$$

It is only applied if $x_{ij} > 0$, however, this distance ignores the correlation among variables.

What has to be kept in mind is the distance d_{ij} between two samples x_i and x_j reflects their similarity. The large the distance is, the more heterogeneous the two samples are. The matrix of distances among any two individual samples can be expressed as:

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & & & \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix} \tag{3.6.11}$$

here $d_{11} = d_{22} = \cdots = d_{nn} = 0$. The samples can be classified by D, as the ones near to each other are grouped together, whereas far away from each will be grouped into different clusters.

As the different techniques to estimate the distance between samples, there are also different methods to estimate the distance between clusters, where various hierarchal algorithms are initiated. The most common used ones are *single-linkage method, complete-linkage method, median-linkage method, centroid method, average-linkage method, flexible-beta method, McQuitty method and Ward method*. Even though the definitions are different, the clustering procedure is the same. If $\mathcal{D}_{ij}$ is the distance between two clusters G_i and G_j, then:

(i) Single-linkage method

It determines the distance between G_i and G_j as the nearest distance between two samples:

$$D_{ij} = \min_{G_i \in G_i, G_J \in G_j} d_{ij} \tag{3.6.12}$$

If the clusters G_p and G_q can be combined into a new cluster G_r, then its distance with another cluster G_k is:

$$\begin{aligned} D_{kr} &= \min_{X_i \in G_i, X_j \in G_j} d_{ij} \\ &= \min\left\{ \min_{X_i \in G_k, X_j \in G_p} d_{ij}, \min_{X_i \in G_k, X_j \in G_q} d_{ij} \right\} \\ &= \min\{D_{kp}, D_{kq}\} \end{aligned} \tag{3.6.13}$$

This step is repeated until it reaches the required number of clusters.

(ii) Complete-linkage method

It is the opposite way of the previous method, determined by the longest distance of G_i and G_j as:

$$D_{pq} = \max_{X_i \in G_p, X_j \in G_q} d_{ij} \tag{3.6.14}$$

It also has the same procedure with …, which classified each individual sample as one cluster firstly, then combined the two samples (clusters) according to its definition. Here if the clusters G_p and G_q can be combined into a new cluster G_r, then its distance with another cluster G_k is:

$$D_{kr} = \max_{X_i \in G_k, X_j \in G_r} d_{ij} \tag{3.6.15}$$

$$= \max\left\{ \max_{X_i \in G_k, X_j \in G_p} d_{ij}, \max_{X_i \in G_k, X_j \in G_q} d_{ij} \right\}$$

$$= \max\{D_{kp}, D_{kq}\}$$

(iii) Median-linkage method

It determines the distance between G_i and G_j as their median distance:

$$D_{kr}^2 = \frac{1}{2} D_{kp}^2 + \frac{1}{2} D_{kq}^2 + \beta D_{pq}^2 \qquad -\frac{1}{4} \le \beta \le 0 \tag{3.6.16}$$

when $\beta = -\frac{1}{4}$, this D_{kr} is the central line of the triangle in the Figure 2.

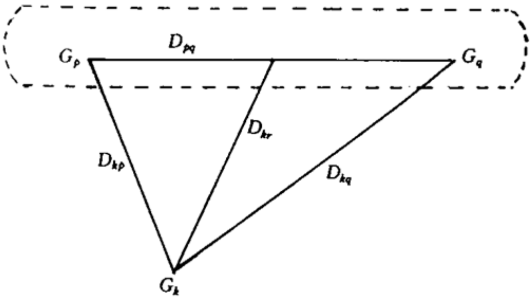


Figure 2 *- Geometric Expression of Median Distance*
Source: author's own document

(iv) Centroid method

It is usually used to reflect the number of members in each cluster, determined by the distance of median points between two clusters. If the mean values of clusters G_p and G_q are vectors $\overline{X}_p$ and $\overline{X}_q$, and if the *Euclidean distance* is applied, then:

$$D_{kr}^2 = d_{X_k X_r}^2 = (\overline{X}_k - \overline{X}_r)'(\overline{X}_k - \overline{X}_r) \tag{3.6.17}$$

(v) Average-linkage method

It assumed that the square distances between two clusters as the mean value of the square distances between any two individual samples insider these two clusters, expressed as:

$$D_{pq}^2 = \frac{1}{n_p n_q} \sum_{X_i \in G_p} \sum_{X_j \in G_j} d_{ij}^2 \tag{3.6.18}$$

(vi) Flexible-beta method

It is processed in the same steps with the previous one, only determined the distance between one cluster G_k and the new cluster G_r as:

$$D_{kr}^2 = \frac{n_p}{n_r}(1-\beta)D_{kp}^2 + \frac{n_p}{n_r}(1-\beta)D_{kq}^2 + \beta D_{pq}^2 \tag{3.6.19}$$

here β is changeable and β>1.

(vii) McQuitty method

It is also in the same procedure with the previous ones, but determined the distance between one cluster G_k and the new cluster G_r as:

$$D_{kr}^2 = \frac{1-\beta}{2}(D_{kp}^2 + D_{kq}^2) + \beta D_{pq}^2 \tag{3.6.20}$$

here β is changeable and β>1.

(viii) Ward method

If the *n* samples can be classified as *k* clusters: G_1, G_2, … , G_k, the vector of sample *i* in cluster G_t is expressed as $X_i^{(t)}$, n_t is the number of samples in G_t, $\overline{X}^{(t)}$ is the median point of G_t, then:

$$S_t = \sum_{i=1}^{n_t}(X_i^{(t)} - \overline{X}^{(t)})'(X_i^{(t)} - \overline{X}^{(t)}) \tag{3.6.21}$$

So the sum of deviation square within of *k* clusters can be expressed as:

$$S = \sum_{t=1}^{k} S_t = \sum_{t=1}^{k}\sum_{i=1}^{n_t}(X_i^{(t)} - \overline{X}^{(t)})'(X_i^{(t)} - \overline{X}^{(t)}) \tag{3.6.22}$$

The *Ward method* can be traced by analysis of variance (ANOVA). The sum of deviation square of individual samples inside one cluster should be slight, whereas the sum of deviation square of different clusters should be significant, if the result is valid. It follows the procedures as: firstly it classifies each individual sample as one single cluster as usual, and then the S will increase with the combination of clusters in the following steps, till the required number of clusters is reached. To better interpret its relation with the previous methods, here I define the distance of G_p, G_q as:

$$D_{pq}^2 = S_r - S_p - S_q \tag{3.6.23}$$

Then the *Ward method* can be expressed as:

$$D_{kr}^2 = \frac{n_k + n_p}{n_r + n_k}D_{kp}^2 + \frac{n_k + n_q}{n_r + n_k}D_{kq}^2 - \frac{n_k}{n_r + n_k}D_{pq}^2 \tag{3.6.24}$$

This methodology also considers the differences among individuals, so I applied this method in my study.

The matrix of the component scores is usually used in the following-up the cluster analysis[24] to group the homogeneous regions together and divide heterogeneous regions separately. This approach allows us to identify areas which show similar structural problems and to describe their peculiarities. The data are obtained from the previous PCA of CIRD dataset, and processed with statistical software (SPSS).

3.7. Dynamic trajectory: convergence theory

Convergence, as one of the most popular growth theories, can be traced back to the neoclassic growth model (Solow, Swan, 1956). The fundamental conception of "*the lower the starting level of real per capita gross domestic product (GDP) the higher is the predicted growth rate*" (Barro, 1996) initiate the prosperity of further studies in convergence theory. There have been various conceptions in the convergence theory, such as absolute convergence, conditional convergence, club convergence and etc. The difference among these concepts mainly focused on the supposition of common homogeneous characteristics that the economies have. For example, absolute convergence assumes that all the economies will converge together with their income per capita in the long run with no presupposition of common characteristics among these economies, whereas conditional convergence and club convergence predict the economies will converge all together or form into two or several groups if there are some homogenous characteristics among the economies (Galor, 1996; Barro and Sala-i-Martin, 1992; Quah, 1996). In the meantime, variety of quantitative techniques is developed such as σ-convergence, β-convergence, and etc. σ-convergence signifies the differences of income per capita among economies are reducing in a long run, whereas β-convergence reflects the correlation between initial level and growth rate of income per capita as the poorer economies will grow faster. Another new methodology to estimate convergence trend has been emerging in recent decades, which is kernel density estimation. Unlike the traditional

[24] Some software, as SPSS (used here), automatically provide standardized values, which are then used in the cluster analysis. Then, all the components are supposed to share the same variance equal to 1, and therefore the same weight in the mapping, with the possible distortive effects.

techniques, it concentrates more on the dynamic transformation along with time changes. All the above three techniques will be applied in the study, in order to give a more complete picture of convergence estimation in China. In the following, I will elaborate the methodological expressions of these measurements of convergence. The basic hypothesis can be expressed as:

$$\frac{d\,(var\,(ln\,Y_t))}{dt} > 0 \qquad (3.7.1)$$

here $\operatorname{var}(Y_t)$ is the variance of productivity or income per capita at time t across countries or regions. If consider two years, then the hypothesis will be:

$$\frac{Var\,(ln\,Y_0)}{Var\,(ln\,Y_1)} > 1 \qquad (3.7.2)$$

here the productivity or income per capita in the initial time is Y_0, and the one in the observed year is Y_1. The empirical estimation of convergence is normally based on the regression equation as following:

$$\ln(Y_1) = (1+\beta)\ln(Y_0) + u = \pi \ln(Y_0) + u \qquad (3.7.3)$$

It assumes that $-1 \leq \beta \leq 0$ and $\pi < 1$ to satisfy the hypothesis of convergence, which implies the countries or regions with lower initial level of productivity or income per capita will incline to grow faster in the subsequent time. That is the well-known mean reversion hypothesis. However, it had been argued the validity to deduce the convergence hypothesis, as a negative β might not necessarily indicate the dispersion, or else to say, the variance of productivity or income per capita reduced over time (Barro, Sala-i-Martin, 1991). Lichtenburg (1994) further improved this argument and proposed the most well-known measurement of convergence as *Lichtenburg Test*. In his point of view, the σ-convergence indicates the reduction of variance of productivity cross-section over time, and should be valued more since it is a necessary and sufficient condition for β-convergence (Lichtenburg, 1994). This *Lichtenburg test* can be expressed as:

$$\frac{\sigma_1^2}{\sigma_0^2} = (1+\beta)^2 + \frac{\sigma_u^2}{\sigma_0^2} = \pi^2 + \frac{\sigma_u^2}{\sigma_0^2} \qquad (3.7.4)$$

here the $\sigma_1^2, \sigma_0^2, \sigma_u^2$ are the variances of $\ln(Y_1), \ln(Y_0)$ and the disturbance of u. It is an inferential function from the expression (3.7.3), with the consideration that the convergence is not only depended on the mean reversion or the parameter β, but also the dispersion of disturbance. It is also pointed out that $\frac{\sigma_1^2}{\sigma_0^2}$ may still over than 1 even if β<0 (π<1) if the disturbance is large enough, or else to say, the σ_1^2 is determined by the regression slope π and the variance of disturbance σ_u^2. Therefore, the convergence test cannot be only relied on t-test on the significance of β, and then F-test is employed with an additional statistic R^2 which is defined as:

$$\frac{\sigma_u^2}{\sigma_1^2} = 1 - R^2 \qquad (3.7.5)$$

If deducing σ_u^2 by this equation and inferring it to the equation (3.7.4), then it can be rearranged and deduced as:

$$\frac{\sigma_0^2}{\sigma_1^2} = \frac{R^2}{\pi^2} \qquad (3.7.6)$$

here it is in a F-distribution with (n-2, n-2) degrees of freedom and we name it as T_1. If we define $\frac{1}{\pi^2} > 1$, it is actually the same with the mean reversion hypothesis that $\beta < 1$, which may cause overestimation of convergence level. Therefore, the additional R^2 is added here to modify the disturbance of the original regression. To summarize, the *Lichtenburg test* assumes that $\frac{\sigma_0^2}{\sigma_t^2}$ will be in the F-distribution with degree of freedom (n-2, n-2) if there is no convergence, which is suggested by the acceptance of null hypothesis. However, this test had been criticized as the biased test procedure (Carree and Klomp, 1995). They argued that Lichtenburg was incorrect and might commit a type II error in assuming F-distribution in the common sense of $\pi > 0$ since the variance at initial time and ending time would not be independently distributed if $\pi \neq 0$. Hence, they proposed further tests expressed as:

$$T_2 = (N - 2.5)ln\left[1 + \frac{1}{4}\frac{\hat{\sigma}_0^2 - \hat{\sigma}_t^2}{\hat{\sigma}_0^2\hat{\sigma}_t^2 - \hat{\sigma}_{0t}^2}\right] \qquad (3.7.7a)$$

$$T_3 = \left[\frac{\sqrt{N}(\hat{\sigma}_0^2/\hat{\sigma}_t^2 - 1)}{2\sqrt{1-\hat{\pi}^2}}\right] \qquad (3.7.7b)$$

here $\hat{\sigma}_{0t}^2$ is the covariance of the initial time 0 and the final time observed t. From equation (3.7.7), it is clear that $\hat{\pi}^2 < 1$ is a premise to estimate T_3. In this case, then if T_1, T_2 and T_3 have a value over the critical value, it is solid to conclude the trend of convergence. These three tests can estimate *σ-convergence* and *β-convergence* simultaneously, however, as Quah (1995) and Frideman (1992) both valued the *σ-convergence* as it could clearly depict whether the economies developed in a balanced manner, they always are used as the measurements of *σ-convergence*. Here what has to be paid special attentions to is the relationship of *σ-convergence* and *β-convergence*, *σ-convergence* might not be accompanied with *β-convergence* since the latter is not a sufficient but necessary condition to the former (Higgins and Levy, 2004).

The measurements above are used in many empirical studies such as the OECD countries or U.S. county level data. However, these measurements more focused on the two signal time points of the starting year and ending year, which might distort the conclusion with the neglect of the middle years across the time interval. In my study, I find it is really complex to depict clearly the convergence tendency clearly only with these three tests: if the variances of the beginning year and the ending year happened to be similar, which means $\pi \cong 1$, then this problem blurs the estimation as the changes in the middle years are all omitted. This problem brings the following technique of *Kernel density estimation*, which aims to estimate the dynamic transformation along with times.

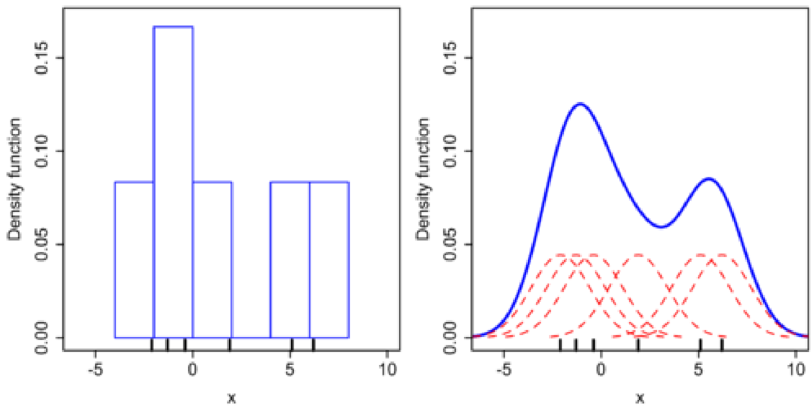


Figure 3 *- Histogram and Kernel Density Estimation Curves*
Source: from the official website of Wikipedia

Quah (1995) proposed this method on the basis of a model of growth and imperfect capital mobility, with the availability to depict the convergence trajectory features, trying to answer the tricky unsolved questions such as what type of trail is that a poor economy had been going through to catch up with the rich ones in the case of convergence, or whether the gap between poor economies and rich economies would remain stubborn or deteriorated as a trend of polarization. This *nonparametric stochastic kernel density estimation* was documented in the previous studies of Quah (1993). Simply interpreted, the principle of *kernel density* is similar to the histogram to estimate the frequency or density of single points in one area, or more precisely to say, how many single points there are around one point. As Figure 3 shows, same data is described by the left histogram and the right kernel density estimate[25].

If we assume the setting of the single points as $X = \{x_1, x_2, \dots, x_n\}$, then the kernel density estimation can be expressed as:

$$\tilde{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-x_i}{h}) \tag{3.7.8}$$

here h is the bandwidth[26] determining the smoothness of the curve (the larger of h the smoother of the curve), x_i is one random point used to how it effects the kernel density estimation of x by it distance to x, *K(.)* is the kernel function with symmetry as $\int K(x)dx = 1$. The most commonly used *K* functions are as following:

$$Uniform: \frac{1}{2}\mathcal{I}(|t| \leq 1)$$

$$Epanechikov: \frac{3}{4}(1-t^2)\mathcal{I}(|t| \leq 1)$$

$$Quatric: \frac{15}{16}(1-t^2)^2\mathcal{I}(|t| \leq 1)$$

$$Gaussian: \frac{1}{\sqrt{2\pi}}exp\left\{-\frac{1}{2}t^2\right\} \tag{3.7.9}$$

The quality of *kernel density* estimation depends directly on the *K(.)* and *h*, for example, if chosen Gaussian function, then the density value $\Phi(\frac{x-x_i}{h})$ will depend on the distance between x_i and *x*, or else to say, the nearer of x_i and *x* are, the larger the density value is. On the other hand, the value of *h* will determine the function of x_i to *x* in the density estimation, as an increase of the bandwidth will initiate one further x_i to function on *x*. If Uniform function is applied, then only the points of x_i with the smaller distance to *x* than the bandwidth *h* can be involved into the density estimation. If Epanechikov is used, then not only the points outranged the bandwidth will be disregarded, but the rest ones with a smaller distance to *x* than the bandwidth will also be weighted according to their distances to *x*. Since the Epanechikov and Gaussian are the two most commonly used functions in

---

[25] The six red-dashed curves describe the individual kernel functions, while the blue-solid curve estimates the kernel density.

[26] In order to avoid the kernel curve shows in a too-sharp or too-smooth shape to blur the exact distribution, it is significant to choose a suitable h. Several ways were proposed in the determination of h such as rule of thumb, plug-in and cross-validation, etc.

the previous studies, in this study I will apply them in the kernel density estimation of CIRD and sub five dimensions (MEI, SII, ESI, HCI, and PFI), and the choice between these two functions will rely on the shapes of kernel density curves with more straightforward interpretation.

In order to estimate the stochastic and dynamic process of evolvement from one time to another, contour plot and kernel surface had been proposed by Quah (1995) with the application of a panel data about cross-country productivity over fifteen years. He interpreted the Figure 4 as following: if the kernel ridge gathered around the positive-sloped diagonal, then it implied the existence of persistence where the countries or regions stayed almost as the same positions as they were; if the kernel ridge moved parallel to the horizontal axis, then the persistence level is low as one's initial status would determine its further location; if the kernel ridge moved parallel to the vertical axis, then a tendency of convergence was assumed to exist as the poor grew faster while the rich grew slower; if the kernel ridge moved totally counterclockwise to accumulate around the negative-sloped diagonal, then a complete reversion happened as the poor became the rich while the rich became the poor in the last time period. He also interpreted the Figure 4 specifically as a persistence shown up with the accumulation along with the positive-sloped diagonal and a club convergence existed with the two peaks investigated in the surface.

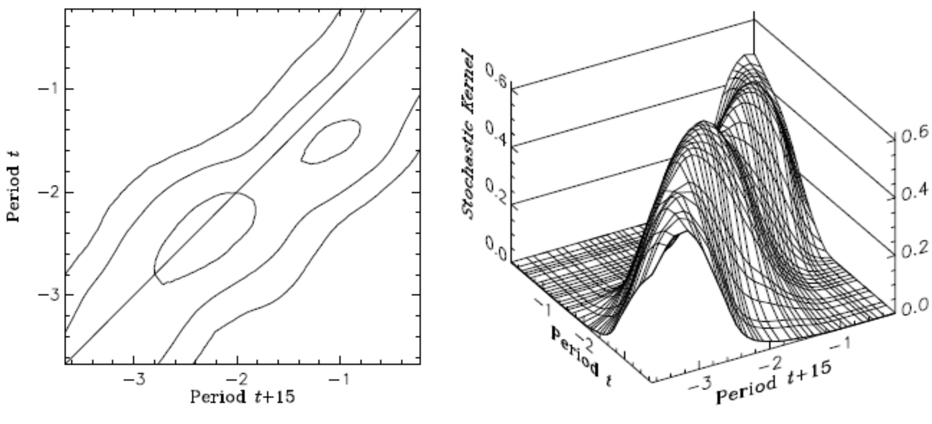


Figure 4 *- Contour Plot and Kernel Surface: Cross-country productivity over 15 years*
Source: Quah, 1995

4. The construction of composite index of regional development (CIRD)

After the step-by-step elaboration of the methodology, I expound the construction of CIRD indicator by indicator, dimension by dimension. PCA results are illustrated in each dimension, and the significance of each single indicator is interpreted on the sub-dimensional level in five pillars. Then the overall socioeconomic conditions in different Chinese regions are reported according to the combination of the five pillars and the aggregation of all the indicators respectively. Ranking reports under these two aggregations are stated and compared.

4.1. Macroeconomic index (MEI)

In the MEI pillar, five indicators were fixed from the initial indicator selection as: GRP per capita, FDI level, international trade balance, final consumption rate, and compensation level. GRP per capita is a very common indicator to estimate the general economic situations in different regions of China. FDI level is applied to observe the development degree through its openness and attraction to the foreign investments. International trade balance is employed here to estimate the openness and activity of the market in different Chinese regions. Final consumption rate is used to evaluate the wealth of local inhabitants and whether the individual needs can be satisfied by local supply. However, the share of consumption in GDP keeps decreasing after the 1978 reform in China and bears a negative correlation with investment (Xu, 2010)[27]. Employment compensation level is used to assess its portion in the GRP distributed, which is also negative and needed to be reversed.

4.1.1. Comments on the PCA result of MEI

First I applied Box-Cox to normalize the skew data in each sub time interval (Appendix 1.1). Then PCA is applied in the three time intervals and one principle component is extracted from the five indicators in each time interval. In the Appendix 1.2,

[27] According to Xu, the reasons of this negative correlation are fourfold: decrease of consumption propensity, adjustment of national income distribution structure, reform of urban housing system and slow growth of rural income.

it is reported the PCA results in each time period in this MEI dimension. In the first time period (1998-2000), the principle component represents 60.59% of total, in which GRP per capita is extracted the most to contribute in this component. In the second period (2001-2005), the principle component represents 59.14% of total; the GRP per capita still is the indicator extracted the most in the component, whereas the extraction from the trade balance decreases. In the third period (2006-2010), the principle component represents 52.02% of total; the indicator contributed the most becomes final consumption rate, whereas the GRP per capita decreases to the second; the extraction from trade balance keeps reducing.

From the PCA result, it is clear the one unique latent dimension lies underneath the five indicators in MEI over the whole time interval. However, the representativeness decreased a little bit along with time, which can be interpreted that the factors that influence macroeconomic conditions become increasingly complex. More and more extra issues that did not reflected in the five indicators are added in the interaction of macroeconomic factors. GRP per capita once occupied as the first significant variable with the largest contribution in the first two time periods, whereas final consumption rate replaced its position, indicating the essentiality to increase the income and promote the consumption for the macroeconomic development. In the government report 2004, Chinese premier Wen proposed that "Balancing investment and consumption is an important aspect of this year's work for macro-control. Consumption accounts to a too small proportion of China's GDP. This is not conducive to ensuring a stable increase in domestic demand, sustaining rapid economic growth and establishing a benign economic cycle", and he also put forward to "…keep working on changing the situation that the rate of investment is high while the consumption rate is low" (Wen, 2004). This change of policy inclination may also cause the higher significance of consumption in the third period[28]. It also verified the significance of final consumption rate in the sustainable and stable development of MEI in China.

4.1.2. Ranking reports of MEI

[28] Even though this report is published in 2004, I here considered the lagging effects of policy on macroeconomic development. Therefore, the influences are more significant in the last period.

After the PCA, I standardized the component score by Min-Max and get the ranking report of Chinese provinces and municipalities. Table 2 reflected the average Min-Max score and change over the whole time interval 1998-2010 (a detailed report table was presented in Appendix 1.3).

Table 2 *- MEI ranking report of China (1998-2010)*

MEI	Average Score	MEI	Average Change
Shanghai	100	Inner Mongolia	3.56
Beijing	89	Tianjin	2.70
Tianjin	74	Jiangsu	1.46
Zhejiang	47	Shaanxi	1.12
Jiangsu	46	Ningxia	0.93
Guangdong	43	Shanxi	0.88
Liaoning	34	Zhejiang	0.85
Shandong	33	Jilin	0.82
Fujian	33	Liaoning	0.73
Inner Mongolia	27	Beijing	0.67
Heilongjiang	22	Qinghai	0.62
Jilin	20	Hunan	0.61
Hebei	19	Shandong	0.59
Hainan	16	Jiangxi	0.26
Hubei	15	Chongqing	0.13
Xinjiang	14	Henan	0.13
Shanxi	10	Guangxi	0.12
Chongqing	10	Guizhou	0.09
Henan	8	Hubei	-0.02
Shaanxi	8	Hebei	-0.05
Sichuan	5	Guangdong	-0.11
Hunan	5	Anhui	-0.20
Anhui	4	Gansu	-0.34
Jiangxi	4	Sichuan	-0.36
Ningxia	3	Yunnan	-0.37
Gansu	3	Fujian	-0.45
Tibet	2	Shanghai	-0.47
Qinghai	2	Hainan	-0.74
Yunnan	2	Tibet	-1.09
Guizhou	2	Xinjiang	-1.11
Guangxi	1	Heilongjiang	-1.39

Source: the author's calculation

From the ranking according to the average Min-Max score, which estimates the general conditions of regional economic development, it is clear that geographical locations are proxy to this grouping result. The regions in the upper level all belong to the east coast, inland regions stay as the mediate level, whereas the underdeveloped level are mostly composed by the peripheral regions. What has to be paid attention to is the dramatic gap between the highly economic developed regions and the underdeveloped regions, reflected as the huge difference in the Min-Max scores in MEI. Also, the macroeconomic differences among the east coastal regions inside the upper level are larger than the internal differences in other groups. The regions in middle economic level, if estimated from the Min-Max scores, are more contiguous to the regions in low economic level rather than the regions in upper economic level as the scores are more continuous in the middle and low levels. The dash lines show the classification according to the score values, based on the national average which is 23. Three groups are formed gradually from high, average to low level of MEI. The member size of the middle group is the smallest among the three groups, while the bottom group is the largest. It implies the middle class around national average level in MEI is missing, while a tendency of polarization is showing up.

However, the ranking report based on the average change of the Min-Max scores is totally different. In the fast-growing level, only several east regions are included. Main portion of the members in this fast-growing level constitutes as inland regions belonging to the middle economic level and some western regions belonging to the below-average or low economic level in the previous ranking. This phenomenon, as I assumed, cannot be separate with the governmental policy propensity in recent years. Plenty of supporting projects are located to the western regions, such as the well-known Western Development Project initiated in 2000. On the other hand, the inland regions bear lots of spreading benefits from developed east coast, and theoretically, their marginal effects are higher than the east coast. Beijing is assigned as the level of middle or upper-middle economic growth, and Shanghai, ranked as the top of economic level in the whole nation, shows a negative average change during these years. The one in top 3 of average-score ranking still remains in the top in the growth ranking is Tianjin, which is reasonable since the government intensifies the supports to this municipality in recent years. The slowest growing province is Heilongjiang, a traditional region of heavy industry. This province was once the core of the national industry in the 1950s-1970s. However, it also suffers a backward transformation in production structure. No special policy inclination had been enacted in that region, which should be paid attention to for the policy makers. Other negatively growing regions are mainly the peripheral regions, however, this backward, in the author's opinion, is more caused by the extreme low level of economic conditions (as reported in the previous ranking), so the benefits from the policy support are limit.

4.2. Science and innovation index (SII)

In the SII dimension, five indicators[29] were selected on the basis of preferred methodology discussed before, containing the governmental expenditure proportion in SII, labor productivity in high science and technology enterprises, income per capita in high science and technology enterprises, R&D expenditure level, and trade in the science and technology market. The governmental expenditure portion in SII is employed to estimate the degree of importance that local governmental attached to SII; the labor productivity and income per capita in high science and technology enterprises are two indicators to investigate the production and development of individual enterprises and firms in SII; R&D expenditure proportion is applied to estimate the sustainability of research and development; trading value in science and technology market is used to estimate the level of market activeness and the output of science and technology. All of the indicators are positively correlated with GDP growth rate so there is no need to reverse.

4.2.1. Comments on the PCA result of SII

Before running PCA, the distribution of the data has to be checked and normalized by the Box-Cox as discussed. The skew data is reported in the Appendix 2.1. One principle component is extracted from the five selected indicators in the each of three time periods. Appendix 2.2 reported the PCA result of SII in three time periods. In the first period (1998-2000), the principle component explained 56.16% of total variance, and the indicators that had been extracted most are the two indicators referring to the high scientific technology enterprises and firms, whereas the governmental expenditure in SII contributed the least. In the second period (2001-2005), the principle component

[29] The five indicators are all calculated by the author from the original variables available from the Chinese Statistical Yearbooks.

represents 54.92% of total variance, and still the two indicators on high scientific technology enterprises and firms contributed most in the component, meanwhile the contribution from governmental expenditure in SII increased, but the trading value in scientific market decreased its proportion. In the third period (2006-2010), the principle component represents 55.32% of total, where the most contributed indicator changed to R&D expenditure level, and the indicators referring to the high scientific technology enterprises decreased their portions dramatically, governmental expenditure in SII overturned from the last to the second.

The PCA result clearly indicates only one unique latent dimension can be extracted from the five indicators in SII over 1998-2010, and its variance explanation keeps stable in the three sub periods, reflecting the combined influence of the five indicators to SII is stable on some level. However, the distribution of the five indicators' contributions to the SII principle component changed over different time periods. From the Component Matrix in the Appendix 2.4-2.6, it is obvious that all the indicators' contributions tend to distribute in a more balanced manner, indicating that all the five indicators are becoming similarly significant to the SII development. The indicators of high scientific technology enterprises took the most extractions in the first two periods, reflecting the SII development was more relied on the individual enterprises. However, the extraction from these two indicators reduced dramatically in the last period, and the R&D expenditure level and governmental expenditure proportion became the first two, indicating the SII development in recent years was more relied on public and governmental encouragements, and also reflecting somehow the importance of capital injection in the improvement of SII. Actually, this tendency has already been implied a little in the second period, even though the two indicators of high scientific technology firms still remained high, their significances decreased and meanwhile, there have been the increases in the contribution of governmental expenditure in SII. It cannot be analyzed without the governmental inclination in the investment of science, technology and innovation in the guidelines. It is also instructive for the policy makers to attach more attentions to the investment in SII.

4.2.2. Ranking reports of SII

After running PCA, I applied Min-Max to obtain the ranking scores of SII for 30

Chinese municipalities and provinces. Here Tibet is excluded due to the missing data in most of the years. Table 3 shows the average Min-Max score and change over the whole time interval 1998-2010 (the ranking report of each single year is presented in the Appendix 2.3).

Table 3 *- SII ranking report of China (1998-2010)*

SII	Average Score	SII	Average Change
Shanghai	94	Sichuan	3.05
Beijing	85	Hebei	2.10
Guangdong	84	Beijing	1.80
Fujian	75	Liaoning	1.63
Jiangsu	74	Gansu	1.59
Zhejiang	58	Ningxia	1.39
Tianjin	47	Inner Mongolia	1.26
Shandong	44	Shaanxi	1.23
Chongqing	40	Chongqing	1.21
Sichuan	36	Shanxi	1.19
Jilin	32	Heilongjiang	1.17
Hunan	31	Henan	1.06
Anhui	30	Hubei	1.05
Hubei	24	Hainan	0.99
Shaanxi	22	Anhui	0.93
Yunnan	22	Tianjin	0.85
Heilongjiang	22	Xinjiang	0.51
Hebei	21	Jilin	0.33
Liaoning	21	Yunnan	0.17
Hainan	18	Qinghai	0.03
Henan	16	Shandong	-0.01
Jiangxi	15	Guizhou	-0.04
Ningxia	10	Fujian	-0.28
Xinjiang	8	Jiangxi	-0.40
Inner Mongolia	8	Guangdong	-0.42
Gansu	8	Guangxi	-0.73
Shanxi	7	Shanghai	-1.50
Guangxi	6	Hunan	-2.41
Qinghai	1	Jiangsu	-3.93
Guizhou	0	Zhejiang	-6.11

Source: the author's calculation

The average Min-Max score estimates the SII developmental level in each province or municipality. Similar to the MEI pillar, the ranking of SII average score is proxy to the geographical locations. This ranking is almost following with the common opinion that there is a descending classification from east to inland to west in China. The upper group (top 10) is composed as mainly east coast regions, however, it is different that the southwest regions Chongqing and Sichuan jumps into the top 10, where their MEI belong to the middle even below-middle levels. If we take a look at the ranking report of average change, it happens to show that Sichuan and Chongqing are the top two regions increasing fast in SII, which somehow make their relative high positions in SII performance more reliable. As I calculated, the governmental expenditure in SII of Sichuan and Chongqing increased dramatically during these years (103.42% and 97.61% respectively). Also, the local governments put forward plenty of supporting policies in fostering academy, innovation and scientific research (it can be reflected somehow from the high level of patent application and patent authorization, which are two original indicators disregarded in the selection procedure due to their low significances). These issues might explain the dramatic growths and better performances of these two regions in SII. The average scores also show the slip in SII between the upper developed regions and middle level regions since the scores are discrete. We can see a clear discontinuing between the fifth (Jiangsu) and the sixth (Zhejiang). The gaps inside the super developed regions in SII (top 5) are not as large as in MEI. If we look at the dash lines in the row of average score that classify the provinces into three parts, it is clear that these three groups are nearly balanced in the number of their member sizes. The gaps inside the first group are larger than the other two groups, indicating a split-up tendency of SII inside the developed regions.

The ranking of average change in SII is somehow out of the common opinion. The last six with negative changes included four east coastal regions. Shanghai, Guangdong, Jiangsu, and Zhejiang, which are commonly accepted as highly developed regions here show negative changes in SII. As I assumed, it might be caused by the slight increase of governmental expenditure in SII or R&D expenditure. But it does not imply the capital investment is low, on the contrary, the governmental expenditure in SII and R&D are all in a high level in these provinces, which might be the reason of negative change since they are already very high. Also, it shows a tendency of convergence in SII, since the regions with higher (initial) level changed slowly.

4.3. Environmental sustainability index (ESI)

The ESI pillar also contains five indicators which are selected from the initial indicator selection: proportion of industry waste water discharged meeting required standards, utilized ratio of industry waste (waste water, waste solid and waste gas), output from the products made of industry waste, discharged living waste per capita and investment of anti-pollution in industry firms. This part focused more on pollution since it is the key problem that China has been confronting in recent years[30]. The proportion of industry waste water discharged under the qualified standards is used to estimate the capacity of facilities to clean industry waste water. Utilization ratio and output value of industry waste are both for estimation of the waste recycling conditions, the former is more focused on the industry recycling ability, while the latter is more concentrated on the economic benefits created by recycling, or in other words, it is to assess the potential sustainability of reprocess from production angle. The discharge of living waste per capita is a “negative” indicator to the objective of ESI pillar, which has to be reversed in the practical PCA. It evaluates the pollutant produced by the daily life of local inhabitants. Investment for anti-pollution project in industry firms is used to gauge the control and recover the pollution from producer’s aspect.

4.3.1. Comments on the PCA result of ESI

After employing Box-Cox to normalize the skew data (Appendix 3.1), PCA is applied in the three time periods separately to extract the component. In the ESI dimension, two principle components were extracted from the selected five indicators with the eigenvalues over than 1, however, the first principle component occupied all over than 50%, hence it is reasonable that I here just extract the first component for the following calculation. The PCA results of ESI are reported in Appendix 3.2. In the first time interval (1998-2000), the first principle component explained 57.04% of total variance, where the indicator of utilization ratio of industry waste contributed the most and proportion of industry waste water discharged meeting standards were the second, slightly less than the former. In the second time period (2001-2005), the first principle

[30] More original variables were selected referring to public green area, population density, facility capacity for waste water, etc. however showing a non-significant correlation with other variables.

component occupied 52.97% of total, and the most contributed indicator was still the utilization ratio of industry waste, whereas discharge of living waste per capita was the least significant one. In the third period (2006-2010), the first principle component occupied as 55.87% of total, where the utilization ratio of industry waste still stayed as the first contribution, and discharge of living waste per capita still remained as the lowest.

The PCA results signified not only one unique latent indicator beneath the five indicators in ESI. However, the first principle components explained the majority of the total. The indicator of utilization ratio of industry waste was always occupied the most contributed one, implying the importance of industry recycling process to the environmental sustainability, which might be owing to the majority of the pollutants came from the heavy industry. It is also verified by the least importance of discharge amount per capita of living waste. It is reported that the industrial emission keeps declining in China, even though it still remains well above the standards when compared with developed countries; meanwhile, the waste treatment has made great improvement thanks to the new technology and policy initiates (Xiao, 2010). As I concerned, the core power for environmental sustainability focus on the technological progress, including the production technology and waste-treatment technology. Another key point is the government and public support. Special policies and regulations should be encouraged for supporting the green technology and green energy source. On other hand, we cannot slacken the attention on living waste discharge even though it is showed the least significant among the five indicators. It is showed that the discharge amount of living waste water, solid and dust are increasing over time, which might due to the urbanization and the growing population density.

4.3.2. Ranking reports of ESI

In this step, Min-Max was applied to rank the 30 Chinese provinces and municipalities[31]. Table 4 shows the Min-Max score of average and change over 1998-2010 (the ranking report of each year is in Appendix 3.3). The ranking in this dimension is different with the previous two pillars, and also different with the common opinion according to the geographical locations. If reading the ranking report on the average

[31] Here Tibet is excluded because of the missing data.

scores, the province performed above average in ESI either belongs to the peripheral regions, such as Hainan, Xinjiang, Ningxia, Yunnan, or the south regions, such as Jiangsu, Zhejiang, and Guangdong. The north regions all place at the bottom of the ranking. It shows a lengthwise classification from south to north, rather than the commonly accepted lateral classification from east to west. It is reasonable considering the south regions in China have better natural and climate superiority. Another reason to be considered is the industrial structure. The peripheral regions, such as Yunnan, Hainan, performed better in this pillar, mainly due to the underdeveloped heavy industry in these regions, so the pollutants are less than the developed regions. Instead, if we take a look at the last provinces in the bottom of the ranking, they all belong to the north area, and their heavy industries are the main sector, occupying a large portion in their GRP. Another common of these regions is the high population density; hence the discharge of pollutant coming from daily life cannot be omitted. Moreover, these regions are neither developed in the science and technology, which can be reflected from their SII ranks. It implies they reserve less strength in supporting the sustainable development. The dash lines in the column of average score divided the provinces into three groups by their score values, as the national average ESI is 39. The first group is almost overlapped with the top quintile. The second group takes the majority of the provinces and municipalities, indicating more regions are gathering together around national average level of ESI. The third group is with the least number of members and totally overlapped with the bottom quintile. There is also a slice existing among members inside this group that cannot be neglected.

The ranking report according to the average change of Min-Max scores presents a scattered distribution geographically. If we observe the traditional developed east coast, none of them belongs to the top quartile. This phenomenon, as I concerned, may due to that some of them already showed a high performance in the ESI average score, such as Jiangsu, Zhejiang, and Guangdong, and some of them are so developed and urbanized that bear more pollutants from both industrial production and daily life. The common-accepted underdeveloped regions, such as Yunnan, Ningxia and Guizhou, ranked in the upper level in this part, is mainly because their relative ranks raised up since some inland regions slipped back in recent years because of their development in heavy industry. This upgrade, as I concerned, more comes from the change of relative position instead of the self-improvement, which can be further verified if looking detailed data of the original indicators.

Table 4 - *ESI ranking report of China (1998-2010)*

ESI	Average Score	ESI	Average Change
Jiangsu	96	Hunan	4.39
Zhejiang	81	Yunnan	2.75
Guangdong	78	Inner Mongolia	2.27
Hainan	61	Shaanxi	2.01
Shandong	61	Shanxi	1.85
Shanghai	54	Shanghai	1.80
Ningxia	52	Guizhou	1.72
Hubei	50	Jiangxi	1.13
Hebei	45	Hubei	0.85
Sichuan	44	Tianjin	0.79
Beijing	44	Qinghai	0.43
Hunan	40	Anhui	0.35
Shanxi	40	Jiangsu	-0.00
Tianjin	40	Guangdong	-0.03
Xinjiang	37	Gansu	-0.04
Liaoning	35	Ningxia	-0.05
Henan	35	Chongqing	-0.05
Inner Mongolia	34	Hebei	-0.10
Yunnan	30	Hainan	-0.23
Guangxi	29	Xinjiang	-0.24
Qinghai	28	Fujian	-0.58
Heilongjiang	27	Henan	-0.67
Jiangxi	26	Guangxi	-0.71
Fujian	26	Zhejiang	-1.00
Anhui	22	Jilin	-1.13
Guizhou	18	Shandong	-1.37
Jilin	14	Beijing	-1.77
Chongqing	10	Sichuan	-2.57
Shaanxi	4	Liaoning	-2.94
Gansu	3	Heilongjiang	-4.64

Source: the author's calculation

4.4. Human capital index (HCI)

In this pillar, five indicators were selected from the original dataset, including proportion of literate population, ratio of employment in the third sector over the first sector, proportion of population at working age, life expectancy, urban population. It is actually developed from the conception of the well-known HDI with three dimensional of longevity (life expectancy), education (knowledge) and resource (standard of living)[32]. In this HCI, the proportion of literate population is used to estimate the education level of local population; ratio of employment in the third sector over in the first sector is used to picture the employment structure and the productive activities of the local inhabitants; proportion of the population at working age is used to assess the distribution of available labor force in the region; life expectancy is estimated from China's population censuses, depicting the average number of years a person can expect to live; proportion of urban population is used to estimate the urban-rural population distribution structure, since China has a very distinct administrative classification of its population into two groups as rural and urban population, of which the urban population can have better infrastructure for health care, education and employment.

4.4.1. Comments on the PCA result of HCI

After the Box-Cox normalization of the skew data (Appendix 4.1), PCA is applied in the three sub periods and one principle component is extracted from the five indicators in each sub period. In the Appendix 4.2, the PCA results of HCI are reported by the time intervals. In the first sub period from 1998 to 2000, the principle component explained 71.56% of total variance, and the ratio of employment in the third sector over the first sector contributed the most. In the second sub period from 2001 to 2005, the principle component explained 75.45% of total, where the urban population and life expectancy contributed as the highest top two. In the third sub period from 2006 to 2010, the principle component explained 76.11% of total, where urban population and life expectancy still remained as the top two.

From the PCA result, it is clear the one unique latent dimension lies underneath the five indicators in HCI over the whole time interval, and explains majority of the five indicators, which might be caused by the high correlation of the indicators. All the indicators' contributions are not significantly different, implying they wholly attach

[32] This standard of living is actually reflected from the previous pillar MEI, so here the HCI is actually expended from the other two dimensions.

importance to the HCI dimension. The urban population in the second and third periods, replaced the first position of the employment ratio of the third sector over the first sector in the first sub period, indicating the social attribute have great consequence in the human capital resources, as it described above, people living in the cities in China enjoy relatively higher level of standard of living. It may also be caused by the speeding-up process of urbanization. The statistical data tells us that, the urban population has grown from 172 million to 665 million from 1978 to 2010, increasing at an average rate of 4.2% a year. Today 49.7% of the Chinese population is living in cities and towns, compared with just 17.9% thirty years ago. For another, the rural population fell from 790 million (82.1% of the population in 1978) to 674 million (50.3% of the population), whereas still has a relatively higher natural birth rate compared with urban area. The enormous number of immigrant workers should be considered when explicating migration: there is 131 million of labor immigration from rural to urban areas, but most are still rural residents in the “HUKOU” system even though they work for long periods in the industry, construction and service sectors in urban areas. If we take a look at the three sectors in China, it is an obvious trend of the transformation from the first sector (agriculture) to the second and third sectors (industry and service). The proportion of employment in primary sector declines dramatically from 70.5% in 1978 to 36.7% in 2010, while the proportions of employment in secondary industry and tertiary sectors increase from 17.3% and 12.2% in 1978 to 28.7% and 34.6% in 2010 respectively. On the other hand, the GDP composition has also changed: even though the secondary industry has always been taking the main part in total (all over 40% from 1978 to 2010, 46.8% in 2010), the tertiary sector has increased its part significantly (from 23.9% in 1978 to 43.1% in 2010), while there is a decrease of agricultural proportion (from the highest 33.4% in 1982 to 10.1% in 2010).

4.4.2. Ranking reports of HCI

The Min-Max standardization is applied here to calculate the ranking scores of the 31 Chinese provinces and municipalities (Tibet is included here). Table 5 showed the ranking report according to average scores and changes over the years (1998-2010) respectively (a detailed report table was presented in Appendix 4.3).

Table 5 *- HCI ranking report of China (1998-2010)*

HCI	Average Score	HCI	Average Change
Beijing	100	Guangdong	2.05
Shanghai	90	Jiangsu	0.42
Tianjin	44	Tibet	0.07
Jiangsu	30	Beijing	0.00
Guangdong	29	Yunnan	-0.12
Zhejiang	24	Anhui	-0.27
Liaoning	23	Shanghai	-0.28
Jilin	18	Guizhou	-0.28
Heilongjiang	15	Chongqing	-0.32
Hubei	15	Hebei	-0.34
Fujian	15	Qinghai	-0.35
Shanxi	13	Zhejiang	-0.45
Inner Mongolia	12	Ningxia	-0.56
Chongqing	12	Gansu	-0.72
Shandong	12	Sichuan	-0.74
Hebei	11	Shandong	-0.82
Hainan	11	Henan	-0.84
Jiangxi	11	Shaanxi	-0.85
Xinjiang	10	Jiangxi	-0.88
Shaanxi	9	Hubei	-0.93
Anhui	9	Guangxi	-0.95
Hunan	9	Fujian	-0.97
Ningxia	8	Hunan	-0.97
Sichuan	8	Xinjiang	-1.02
Guangxi	7	Hainan	-1.03
Qinghai	7	Jilin	-1.11
Henan	5	Inner Mongolia	-1.23
Gansu	5	Shanxi	-1.23
Guizhou	4	Liaoning	-1.61
Yunnan	2	Heilongjiang	-1.63
Tibet	0	Tianjin	-2.93

Source: the author's calculation

The ranking on average Min-Max score pictured the general condition of HCI in each region. The ranking here is similar to the geographic division, the east coastal regions are mostly classified as the top in the HCI dimension, middle class are mainly the inland regions, and the peripheral regions generally distributed at the bottom. The top quintile is composed by the six east coastal regions commonly accepted as the most developed regions in China. However, we can still see a very large fracture inside this group. The top two Beijing and Shanghai are supper developed in HCI with high scores and small difference between each other, whereas the other four regions dropped dramatically in their ranking scores. Actually, if we divided by the score value, there is a missing group in the middle level. All the other regions are below 50 except Beijing and Shanghai. There are 24 provinces with the ranking scores equal or less than the average score (18), occupying as 77.42% of the total 31 provinces and municipalities. The dash lines in the column of average score divided the provinces into three groups by their score values, as the national average HCI is 18. The first group here is overlapped with the top quintile, with deep slices inside it. The second group as the national average level and the third group as the below national average level of HCI, are more close to each other from the score values.

The ranking report of average change of HCI shows there are a lot of regions retrogressed in HCI over these years, only four regions increase or remain the same level. Even though the deduction of HCI in some regions can be attributed to their relatively better performance in the initial time (such as Shanghai), the large-scale of decrease should also be paid attention to. It may indicate a polarization of HCI among the regions, and it seems like the regions tend to group into two different clusters, and these two clusters are getting further away from each other.

4.5. Public facility index (PFI)

In this PFI dimension, five indicators were also selected from the original variables, containing governmental expenditure portion in public service, public transportation vehicles per capita, paved road per capita, public schools per capita, hospitals and clinics per capita. The governmental expenditure proportion in public service is used to explore the investment extent and local government attention in the construction of public facilities. The following four indicators are designed to assess the public facilities possessed by local inhabitants from four different yet basic aspects: transportation facility, transportation infrastructure, education and medical health.

4.5.1. Comments on the PCA result of PFI

First I used Box-Cox to normalize the skew data in each sub periods (Appendix 5.1).

After that I applied PCA in the three sub time intervals and two components are extracted in each period, where the first component all explain over half of the total variance, hence I consider it is reasonable to extract the first component here. In the Appendix 5.2, it is reported the PCA results in each time interval in the PFI dimension. In the first time period (1998-2000), the first principle component represents 54.08% of total, in which the unit public vehicles is extracted the most to contribute in this component. In the second period (2001-2005), the principle component represents 51.07% of total, in which the unit public schools is extracted the most to contribute in this component, and the unit public vehicles decreased slightly to be the second. In the third period (2006-2010), the principle component represents 54.27% of total; the indicator contributed the most remains as the unit public vehicles, whereas the governmental expenditure increases its contribution to the first principle component.

From the PCA result, we can assume that the correlation among the five selected indicators is not as strong as the previous dimensions, which is rational since they reflect separate individual aspects of public facilities. It is clear that the public transportation vehicles occupy a significant place in the development of public facilities. Also, the contribution of government expenditure in public services increases. Improving the infrastructure construction is proposed by all the recent “Five-year plans”, showing the importance that the Chinese government attached to the public facilities. According to incomplete statistics, the governmental expenditure in public services is six-fold over 1998 to 2010. The governmental attention produces remarkable effects on the development of public facilities. The unit public vehicles increased 14.55%, paved road per capita increased 45.70%, unit public schools increased 20.33%, and unit medical services increased 8.31% during 1998-2010.

4.5.2. Ranking reports of PFI

After the PCA, I standardized the first component scores of PFI by Min-Max and get the ranking report of 31 Chinese provinces and municipalities (Tibet is included here). Table 6 reported the average score and change over the whole time interval 1998-2010 (a detailed report table was presented in Appendix 5.3).

Table 6 - PFI ranking report of China (1998-2010)

PFI	Average Score	PFI	Average Change
Shanghai	100	Hunan	1.56
Beijing	94	Guangdong	1.32
Jiangsu	81	Hubei	1.23
Zhejiang	77	Tianjin	1.22
Guangdong	75	Liaoning	0.90
Tianjin	63	Hebei	0.82
Fujian	56	Fujian	0.68
Liaoning	48	Qinghai	0.65
Hunan	47	Guangxi	0.58
Hubei	44	Xinjiang	0.55
Shandong	40	Gansu	0.55
Jilin	33	Henan	0.54
Shaanxi	28	Ningxia	0.53
Yunnan	28	Shandong	0.50
Heilongjiang	26	Shanxi	0.36
Hebei	23	Jiangsu	0.21
Xinjiang	23	Chongqing	0.09
Sichuan	22	Shanghai	0.00
Chongqing	22	Tibet	0.00
Ningxia	22	Beijing	-0.01
Hainan	21	Zhejiang	-0.14
Shanxi	20	Sichuan	-0.15
Anhui	19	Heilongjiang	-0.21
Guangxi	15	Guizhou	-0.32
Henan	15	Anhui	-0.39
Jiangxi	14	Jiangxi	-0.43
Inner Mongolia	12	Hainan	-0.62
Guizhou	11	Jilin	-0.87
Gansu	11	Inner Mongolia	-1.03
Qinghai	7	Shaanxi	-1.45
Tibet	0	Yunnan	-1.90

Source: the author's calculation

From the ranking according to the average scores, which pictures the overall conditions of development of public facilities, if classified the provinces by the ranking number, it is clear that all the super developed regions belong to the east coast. The above average group locates in the northeast and middle inland. The rest groups which are average level, below average and underdeveloped are more scattered geographically, but none of them states in the east coast. Another issue that has to be paid attention to is the gradually decrease of the scores along with the classification. The differences inside each group are also not very dramatic. It implies that there might be a tendency of convergence existing in the PFI pillar. If classified by the ranking scores, as the average is 35, two dash lines in the score column divide the provinces into three groups from developed to underdeveloped. The gap inside the developed group is dramatic, whereas inside the middle and underdeveloped groups, the gaps are not so large.

Nevertheless, the ranking report according to the average change of the PFI Min-Max scores is more complicated to comment as always. In the super fast-growing group, some inland regions such as Hunan and Hubei are grouped in, indicating the dramatic progress of PFI in these regions. The above-average growing group contains more regions belonging to peripheral regions such as Qinghai, Guangxi, Xinjiang, and Gansu, where are the regions covered by the China's Western Development Program. We can assume the progress in these regions to the governmental investment, which can be verified by the statistical data. The governmental expenditure in public services increased 3.96%, 9.34%, 7.06%, and 5.67% in Qinghai, Guangxi, Xinjiang, and Gansu respectively, all higher than the average growth level. Therefore, the investment to the construction of public facilities must be a high level. As usual, the commonly accepted developed regions such as Shanghai, Beijing, and Zhejiang show a zero or negative change over these years, may be ascribed as they are already in a high level of PFI so there is not very large spare space for them to increase.

4.6. The composite index of regional development (CIRD)

After analyzing the CIRD pillar by pillar, I will combine all the five dimensions together to investigate the comprehensive situation of regional development in China. In this part, I first simply combined the average scores of five dimensions together, which means to give an equal weight to the five dimensions, and rank the 30 Chinese provinces and municipalities[33] by this dataset of combined average scores. Then I run the PCA again with all the 25 indicators together in each sub period, and get the principle component scores. These principle components scores will be aggregated with the weights assigned as their contributions to the

[33] Tibet is excluded because of the missing parts in SII and ESI.

total.

4.6.1. *Ranking reports of CIRD with equal-weighting scores*

In this part, I gave equal weight to each dimension and combined the five together to get the CIRD scores and classify the 30 Chinese provinces and municipalities according to the total score and its change. The ranking report is shown in Table 7.

Table 7 *- CIRD ranking report of China (1998-2010)*

CIRD	Total Average Score	CIRD	Total Average Change
Shanghai	435	Inner Mongolia	4.83
Beijing	411	Hunan	3.17
Jiangsu	327	Hubei	3.04
Guangdong	309	Guangdong	2.82
Zhejiang	286	Tianjin	2.63
Tianjin	269	Hebei	2.44
Fujian	205	Ningxia	2.25
Shandong	189	Shaanxi	2.06
Liaoning	161	Shanxi	1.49
Hubei	149	Qinghai	1.39
Hunan	132	Guizhou	1.17
Hainan	126	Chongqing	1.06
Hebei	119	Gansu	1.02
Jilin	116	Henan	0.89
Sichuan	116	Beijing	0.69
Heilongjiang	111	Yunnan	0.53
Ningxia	98	Anhui	0.41
Chongqing	93	Jiangxi	-0.31
Inner Mongolia	93	Shanghai	-0.45
Xinjiang	92	Sichuan	-0.77
Shanxi	89	Shandong	-1.10
Yunnan	85	Liaoning	-1.29
Anhui	85	Xinjiang	-1.31
Henan	79	Fujian	-1.59
Shaanxi	71	Hainan	-1.65
Jiangxi	71	Guangxi	-1.69
Guangxi	58	Jiangsu	-1.84
Qinghai	45	Jilin	-1.96
Guizhou	36	Heilongjiang	-6.70
Gansu	29	Zhejiang	-6.85

Source: the author's calculation

From the ranking report in Table 7, I also divided the 30 provinces and municipalities into five classes, from the best performance to the worst performance. We can see the best performance group all constitutes by the east coastal regions, which are commonly accepted as the highly developed regions. However, the gaps among these regions inside this group are not very slight, showing a tendency of polarization inside the group. The above-average group is composed by the rest of east coastal regions besides the ones in the first group and two inland regions Hubei and Hunan, which locate in the central of China. The members in the average group are more speckled in their geographical locations, and a slight gap exists among the members inside this group. The below average group are either inland regions or peripheral regions, and the gaps among themselves are little. The underdeveloped group is also composed by inland and peripheral regions, but the gaps inside this group are not slight. If I classified the groups by their scores into three groups as high, average and low level of growth, as the national average is 150, it is vividly read from the dash lines dividing the score row into three parts that the first group contains all the east coast regions as a belt locating along the coastal area. The second group is composed by the inland regions adjacent to east coast, northeast and southwest, but the gaps among themselves are not large. The third group is composed by some inland regions in the north and the peripheral regions. Therefore, it is reasonable to conclude the east coastal regions are clustering together, however with some fault lines inside themselves; the inland regions sort of split up vertically from north to south; while the peripheral regions also show a tend to group together.

Referring to the ranking report according to the average change of the CIRD Min-Max scores, it is more complex to comment as always. In the first group, inland regions such as Hunan, Hubei and Hebei and peripheral inland such as Inner Mongolia are grouped in, indicating the dramatic growth rate of CIRD in these regions. The second group, uncommonly, is mainly composed by peripheral regions which all show a below-average level of CIRD. Some of them, such as Qinghai and Guizhou enjoy plenty of supports and policy propensities from the central government in recent years. The commonly accepted developed regions, such as Beijing, Shanghai, Zhejiang and Jiangsu ranked in the lower groups, show a near zero or negative change over these years, may be ascribed as they are already in a high level of PFI so there is not very large spare space for them to increase. However, one of the developed regions, Guangdong shows a better performance in both rankings of general condition and growth rate.

The equal weights might be argued as subjective. As discussed above, it is an essential premise that the unique single latent dimension is extracted from in each dimension, and hence the indicators can be allocated as contributing in an equal weight. However, in the previous PCA pillar by pillar, there are two principle components extracted from the dimensions of ESI and PFI, which cannot satisfy the required premise of equal-weighting technique. This problem leads me to the following analysis of integral PCA.

4.6.2. Integral PCA results and its ranking reports of CIRD

In this part, the PCA is processed with all the 25 indicators together in each sub time periods, and the results are reported in the Appendix 6.1-6.3. According to the Kaiser rule to keep the eigenvalues over than 1, I extracted five principle components in the first and second time periods (1998-2000 and 2001-2005) respectively, and six principle components in the last time period (2006-2010). In the first time period (1998-2000), the extracted five principle components explained 74.54% of total; in the second period (2001-2005), the extracted five principle components explained 75.72% of total; in the second period (2001-2005), the extracted five principle components explained 77.61% of total. Therefore, it is valid to use these principle component scores. Then their proportions of contribution were applied as the weights to aggregate and a final total score of CIRD was obtained. The ranking report is calculated as before in Table 8.

If looking at the ranking by the total average score, the dash lines divided the 30 Chinese provinces and municipalities into three groups as high, middle and low level of CIRD. In the high level group, all the members belong to the east coast, commonly accepted as the developed belt because of its superior nature location. However, the gap inside this group is also large. Slices existing among the members cannot be omitted. The middle level group contains majority of the provinces and municipalities, indicating a tendency that a great proportion of regions are gathering around the national average level of CIRD, which is 23. The low level group is in the same size of the first group in their member numbers. The differences between and inside these two groups are not so dramatic, which is sort of an implication of tendency to convergence.

Table 8 - *CIRD ranking report of China on integral PCA results (1998-2010)*

CIRD*	Total Average Score*	CIRD*	Total Average Change*
Shanghai	100	Shandong	100
Beijing	87	Shaanxi	88
Tianjin	79	Jiangsu	87
Jiangsu	69	Inner Mongolia	83
Zhejiang	68	Hebei	80
Guangdong	59	Hunan	77
Liaoning	59	Hubei	75
Shandong	58	Jilin	74
Fujian	55	Shanxi	70
Heilongjiang	46	Chongqing	69
Hubei	45	Sichuan	69
Jilin	44	Anhui	67
Shaanxi	42	Jiangxi	66
Hebei	40	Zhejiang	66
Shanxi	40	Henan	63
Hainan	38	Yunnan	62
Xinjiang	37	Guangxi	61
Hunan	34	Guangdong	57
Inner Mongolia	33	Heilongjiang	56
Anhui	32	Fujian	50
Qinghai	32	Gansu	50
Chongqing	31	Liaoning	49
Sichuan	29	Ningxia	45
Henan	28	Guizhou	45
Ningxia	28	Xinjiang	42
Yunnan	27	Tianjin	35
Jiangxi	27	Qinghai	20
Gansu	25	Hainan	13
Guangxi	17	Shanghai	3
Guizhou	0	Beijing	0

Source: the author's calculation

If we take a look at the ranking of change[34], the geographical classification is no more functioned here. The dash lines also divided the 30 Chinese provinces and municipalities into three groups as fast, middle and slow level of CIRD growth. The fast-growing group is overlapped with the top quintile, most of which are inland regions. The middle-growing group is still with the largest member size, covering the second to

[34] Here I also used Min-Max standardization to better depict the growth performances because some of the factor scores obtained from integral PCA is negative.

fourth quintile provinces. It implies the majority of regions in China developed in a rate around the national average growth rate. The low-growing group has the smallest member size, which contains only five regions, whereas three of them are east coastal regions. Beijing and Shanghai, the top two in the ranking of CIRD, are located in this group. As I emphasized before, these regions performing in a low growth rate might be attributed to that they had already been in a superior level of general CIRD, so there was little spare space for them to grow more. Two maps according to the average and change were reported in the Appendix 8.

If compared the two ranking reports on equal weighting and aggregated PCA, the difference mainly focused on the group distribution. If given equal weights and then trisection, the low level group is constituted by the largest member size and the high level group as the smallest, but the difference among the member sizes of three groups is not quite dramatic. However, if aggregated the principle components, the majority is concentrated in the middle level group while the other two groups on both sides are in a similar size that is relatively smaller. It is more or less a normal distribution, which, as I concerned, is more suitable to describe the current reality in China. Besides the distribution, the ranking orders of some provinces also changed inside the group, but the general classification remained consistently. However, the ranking on total average change showed some alterations, such as Beijing and Shanghai performed at the bottom in the aggregated PCA result, whereas located in the middle-lower part of the ranking with equal weights. I assume that all these differences might be caused by the different influences and significances of different indicators/dimensions in different time periods. For example, environmental sustainability was put higher emphasis in recent years, and science and innovation, human capital also were attached with increasingly importance and gradually took the primary task of economic growth. Therefore, even though the ranking report based on equal weighting could somehow provide a reasonable result, the application with aggregated PCA, as I concerned, depicted the Chinese regional socioeconomic development more precisely, and hence the following applications were also on the basis of this method.

5. Application of CIRD: clustering and mapping China

After the separate analysis in each dimension, I collected all the indicators in the five dimensions together to get a general picture of the regional socioeconomic development. First, I ran the principle component analysis (PCA)[35] again in all the 25 indicators in the three time intervals (the results are reported in the Appendix 6.1-6.3), and then I used the component scores to run cluster analysis to see how the different Chinese provinces and municipalities perform among each other in the three time intervals[36]. Here I choose the beginning year and ending year in each time interval, which are the years 1998, 2000, 2001, 2005, 2006 and 2010, to estimate how the cluster process evolves. Six maps are provided to directly perceive the different groups of Chinese provinces and municipalities. In the cluster analysis, I grouped the 30 Chinese provinces and municipalities into four clusters. From the maps in Figure 5, it clearly showed that the four clusters generated by the geographic locations. Even though it may show some hierarchy grading in the four clusters, we should keep in mind the cluster number does not signify the supper or lower levels, or else to say, the clusters are actually defined according to the similarity or common characteristics among the members in one cluster.

[35] Here the application of principle component analysis is only aimed for data reduction, therefore I did not use the rotated component matrix since it is used for better interpret the component characteristics.

[36] In the principle component analysis and cluster analysis of all 25 indicators, Tibet was excluded in the dataset since there were missing data in the single dimensions of science and innovation (SII) and environmental sustainability (ESI). It is reasonable since Tibet is so different in social ideology with the other main land regions. Many previous studies on regional performances of China also separated Tibet considering both data availability and its special social ideology.

***Table 9a** - The cluster reports on five dimensions 1998*

1998	Cluster 1	Cluster 2	Cluster 3	Cluster 4
MEI	***68***	***30***	***17***	***0***
GRP per capita	182.98	74.60	50.91	51.62
FDI level	0.98	0.41	0.11	0.02
trade balance	(7.45)	(21.71)	(15.56)	(26.75)
Consumption	54.68	44.33	37.58	37.90
Compensation	52.62	44.69	40.49	37.57
SII	***82***	***34***	***11***	***0***
gov_exp_SII	2.01	1.78	1.50	0.97
labor productivity in high S&T firms	35.33	22.69	13.92	13.58
income per capita in high S&T firms	29.51	21.88	11.85	9.00
R&D expenditure	1.43	0.52	0.29	0.27
trade value proportion in S&T market	1.27	0.46	0.30	0.13
ESI	***58***	***46***	***36***	***35***
discharge of industrial waste water meeting standards	74.07	59.59	43.09	70.64
utilized ratio of industrial solid waste	79.00	44.26	33.34	19.02
output value of products made from industry waste	68448.80	117717.44	56004.88	1345.00
volume of living waste water discharged per capita[37]	41.30	14.30	9.96	18.18
investment of industrial enterprises on anti-pollution projects	48049.80	50178.56	22069.00	10804.00
HCI	***65***	***21***	***12***	***8***
literate proportion	90.45	87.55	82.18	60.18
employ 3rd sector	239.32	54.55	38.08	38.40
working age distribution	70.89	68.68	66.73	67.05
life expectancy	74.79	70.06	68.02	64.67
urban proportion	67.38	35.85	27.80	33.86
PFI	***80***	***45***	***19***	***3***
gov_exp_public_service	11.27	9.82	9.03	10.99
public vehicles per capita	13.51	8.84	7.88	8.57
paved road per capita	8.72	8.36	7.41	6.34
Education	0.18	0.17	0.09	0.12
medical health services per capita	35.53	30.18	25.93	35.58
Total	***71***	***35***	***19***	***9***

Source: data from Chinese statistical yearbook and calculation by the author

[37] The indicator of volume of living waste water discharged per capita is reversed in the PCA to get the positive component matrix, here it is the original non-reversed data, hence it is negative in the ESI dimension, which means the more discharged, the less the environmental sustainability is.

Table 9b *- The cluster reports on five dimensions 2010*

2010	Cluster 1	Cluster 2	Cluster 3	Cluster 4
MEI	***75***	***42***	***22***	***3***
GRP per capita	459.56	274.21	207.34	174.97
FDI level	0.95	0.33	0.19	0.09
trade balance	(5.84)	(1.25)	(11.69)	(7.18)
Consumption	52.15	49.61	38.52	29.47
Compensation	54.96	55.64	48.74	49.25
SII	***87***	***48***	***31***	***8***
gov_exp_SII	3.19	1.13	1.23	0.77
labor productivity in high S&T firms	93.25	93.28	91.13	51.01
income per capita in high S&T firms	119.19	106.02	105.58	62.31
R&D expenditure	2.10	1.03	0.96	0.80
trade value proportion in S&T market	1.63	0.40	0.19	0.19
ESI	***53***	***28***	***33***	***41***
discharge of industrial waste water meeting standards	96.92	91.11	92.40	73.51
Utilized ratio of industrial solid waste	82.66	67.30	63.43	47.05
output value of products made from industry waste	96.58	40.12	45.81	15.28
volume of living waste water discharged per capita	40.97	24.67	22.81	17.19
investment of industrial enterprises on anti-pollution projects	17.26	13.42	74.11	74.77
HCI	***66***	***36***	***34***	***22***
literate proportion	97.04	94.67	94.25	90.78
employ 3rd sector	395.38	86.09	67.25	76.11
working age distribution	77.49	74.97	71.34	70.87
life expectancy	77.13	74.94	73.57	71.10
urban proportion	73.63	48.76	41.16	38.22
PFI	***90***	***83***	***33***	***14***
gov_exp_public_service	14.33	15.23	13.70	14.05
public vehicles per capita	13.62	9.93	9.17	10.62
paved road per capita	13.77	12.34	13.08	10.09
Education	0.22	0.21	0.11	0.14
medical health services per capita	38.59	40.23	25.48	23.98
Total	***74***	***48***	***31***	***18***

Source: data from Chinese statistical yearbook and calculation by the author

5.1. The first cluster

The first cluster is characterized as brightly light color in the map, locating in the east coast area. It developed from several single points to a long belt along all the east coast and even spreading into inland regions. The members in the first cluster, if summarized the common, have all been in a relatively developed condition and in the head of urbanization. Industry and service contributed more to their economic growth rather than the primary sector. From Table 9a and 9b, the MEI index performs better than the other clusters in all the time intervals, the GRP per capita increased dramatically (151.15%). However, other indicators such as consumption and compensation kept stable. The trade balance indicator even decreased, indicating the transformation of this cluster that the economic growth relies less on the export. Moreover, the average ranking score of MEI decreased from 68 to 55 in this cluster even though it still performed as the best in this dimension, which somehow signified the gap between this cluster and others was decreasing. The SII index performs the highest with all the best level of the indicators in this dimension. During 1998 and 2010, the governmental expenditure in science and innovation increased 58.30%, the labor productivity of high science and technology enterprises almost tripled, the income level of high science and technology enterprise of 2010 was fourfold than in the initial time, the R&D expenditure increased 46.89% and the trading value in science and technology market increased 28.45% We can clearly see the huge improvement in high science and technology enterprises, indicating this cluster is becoming the scientific core. The ESI index however is not so dramatically distinguished with other clusters, especially in the first year. The gaps among four clusters are not as large as other dimensions, but increased from the initial time to the final year. As far as I concerned, it might due to the structure transformation. In the first periods, this cluster still occupied as the center of industry development of the whole nation. After that, some parts of heavy industry were transferred to inland regions. On the other hand, the local industry enterprises put lots of efforts on the pollution management, which could be signified as the highest increasing rate of anti-pollution investment (259.21%). However, it still required to be pay special attention on the decrease of the total score on ESI index, which might indicate the deterioration of the integral environment. The HCI index of this cluster almost stayed stable between the first year and the last year, whereas the gap between this cluster and others are enlarging. It demonstrated the highest level of employment activities in the third sector, which indicates that this cluster is developing an internal system to occupy the surplus labors from rural areas or agriculture, such as township business, and agricultural production is no longer the main force of economic growth to rely on. Also, it is the indicator with the largest increase (65.21%). It showed the highest proportion of urban population signifying the speeding-up process of urbanization. The PFI index of this cluster is still the highest. However, the gap between this cluster and the second cluster is decreasing more than fivefold. Moreover, the government expenditure on public services and the medical services per capita in 2010 of this cluster are exceeded by the second cluster, indicating these two clusters tend to converge with each other. To summarize, this cluster can be defined as the core area of high-tech and superior human capitals. The FDI level and governmental expenditure in science and innovation are also the highest, indicating the advanced productive forces. The proportion of educated population is the largest, indicating the high level of labor force. Also, it shows a super level in public facilities such as road, vehicles and education. Compared the mean value with other clusters, this cluster performs better in high quality labor force, science and technology, infrastructure construction and etc., which exactly prove the high developed level and urbanization of this cluster.

5.2. The second cluster

The second cluster always has the largest group of members even though the number has decreased along with time. In the first period, most of the members were inland regions and also with several northern and eastern regions. In the end of the second period, this cluster has been divided geographically as north and south part since some of its previous members such as Shaanxi have been clustered into the third group. Also, the east regions that used to belong to this cluster have been cluster into the first cluster. This might be a sign of non-convergence. In the last period, it members further decreased even though it is still the largest group. Some inland regions such as Hunan and Hebei were excluded from this group since the influence of the first cluster has been spread. Some peripheral region was left out and became the member of the third cluster such as Yunnan, and some inland region was included into this cluster such as Henan. Therefore, this cluster eventually gathered as one part geographically and moved a little bit north of its aggregate location. This cluster is certainly not developed as the first one in all dimensions. The MEI ranking score is less than half of the report of first cluster, whereas this gap decreased slightly in the last year. The consumption level and compensation level are almost at the same level with the first cluster, and the compensation level even exceed in the last year (0.68%), indicating a labor-intense manner of this cluster. In the SII dimension, its difference to the first cluster decreases 10 points, with the dramatic improvements in the high science and technology firms, and R&D expenditure level. However, the governmental expenditure in science and innovation and the trading value proportion in science and technology market decrease. It might be due to the expanding of the market ratio of the first cluster in science and technology market on the one hand, the changes of membership in the second cluster, especially the previous members locating in the east coast might also be one the of reasons. In the ESI dimension, it is the cluster deteriorated the most in the ranking report (18 points). However, all the indicators seemed to keep increase positively in this dimension, its increasing rate is much less than the third and fourth cluster. This cluster bears a lot of heavy industries that has been speared out from the first cluster. Also, it has been influenced negatively by the pollution caused in the developing process of the first cluster owing to their adjacency. The HCI report has narrowed down its gap with the first cluster. The literature proportion increased 8.14% and its difference with the first cluster slightly decreased from 0.54%. The employment ration of the third sector over the first sector changed the most (57.82%), signifying the production structure change in this cluster. The second large change in this cluster in HCI is the urban population (36.01%), indicating the process of urbanization. The labor distribution and life expectancy also increased 9.15% and 6.96%. The PFI increased most in all dimensions (38 points), the gap with the first cluster was cut down six fold. The governmental expenditure in public services increased the most (123.15%), the public vehicles per capita increased 13.92%, the paved road per capita increased 47.62%, the public schools per capita increased 101.28%, and the medical services per capita increased 33.28%. As the aggregate score increases the most among all the clusters (12 points), this cluster somehow can be defined as the fast developing group with all its members have just started or been in the process of urbanization, so the third sector still composed as an average proportion in the local economic developments. Also, some of them are still in the initial transformation stage of urbanization, which could be confirmed by a relative average level of urban population. However, the labor force in the working age shows a high distribution in this cluster, which is reasonable considering lots of its members are labor-export regions (such as Hubei, Hunan, Sichuan and Henan). The characteristic of labor-intensive production is more obvious in this cluster. If compared with other clusters, this cluster can be described as above-average level in public facilities, since the values of medical services and schools per capita are near to cluster 1. But still, the gap between cluster 2 and cluster 1 are not slight in many aspects.

5.3. The third cluster

The third cluster is composed by peripheral and inland provinces and municipalities (Yunnan, Guangxi, Xinjiang, etc.) in the first period, however, the members of this cluster are also not stable as the second cluster. We can see clearly from the maps that it spreads into inland regions in the second period, and then locates into the southeast peripheral region. The MEI report of this cluster is with huge gap to the first cluster. The GRP per capita was only 27.82% of the first cluster and 68.24% of the second cluster in the initial year, and less than half (45.12%) of the first cluster and 75.61% of the second cluster in the last year even though it increased during these years. The FDI level is also with dramatic difference with the first and second clusters. The consumption and compensation levels are more similar with the first and second clusters compared with other indicators. The gap of SII report between this cluster and the first cluster is even larger. It only took 13.54% of the first cluster in the initial year, and then increased to 35.81% in the last year. The difference of SII between this cluster and the second cluster is not as dramatic as with the first cluster in the last year (64.83%), which indicates a tendency that this two clusters may converge to each other in this dimension. We can see a relatively above-average level of governmental expenditure in science and innovation in the last year in this cluster, even slightly higher than the second cluster (0.1%). The labor productivity and income per capita in high science and technology firms develops significantly, more than fivefold and sevenfold when compared the first year and the last year, as the fastest growth rates among all the clusters. The R&D expenditure level also grows faster than other clusters (67%). The ESI report has the smallest gaps with the first and second clusters. Even though the investments of anti-pollution in industry firms are not so high as the first cluster, it has the lower level of living waste discharging level, especially in the first year, which indicates the better environmental condition than the first and second clusters. However, this cluster is undeveloped in new green science and technology, combined with low investments in environmental protection considering its below average level of economic situation, hence it performs not well in the sustainability part. In the HCI dimension, it is clearly showed that this cluster is getting closer with the second one. The differences between the two clusters are only 2 points in the last year. The literature population proportion, labor force distribution, life expectancy and urban population proportion are all very similar yet slightly lower than the second cluster (0.42%, 3.62%, 0.02%, and 7.61% respectively). The employment ratio of the third sector with the first sector grows dramatically, even faster than the first and second cluster (1.77 times). However, the level of this indicator is still lower than the second cluster (18.85%), which is the largest difference in this dimension. In the PFI dimension, the differences of this cluster with the first and second clusters are very large, only 23.43% and 36.49% of the first cluster and 41.79% and 39.22% of the second cluster in 1998 and 2010 respectively. The public medical and educational services per capita are in low level, only less than half (49.20%) and two third (66.03%) of the first cluster. But the indicators referring to transportation such as the paved road and public vehicles per capita are even better than the second cluster. This might be caused by the relatively smaller population of this cluster. To summarize, this cluster has huge differences with the first cluster, and its gap with the second cluster keeps stable from the beginning year to the ending year. It possesses unique advantages in environmental sustainability since the heavy industry occupied a relatively small proportion in the production, which can be confirmed by the relatively good performance in the ESI. A certain amount of the indicators referring to the general conditions are close to but slightly smaller than the second cluster, but still it is in a below-average developing condition.

5.4. The fourth cluster

The fourth cluster contains the least members, all of which belong to peripheral regions, locating northwest and southwest mountain areas. Only one province (Qinghai) composed this cluster in the initial time period, and then more provinces adjacent to Qinghai were added in and reached up to its largest number of members in the end of the second period, several inland regions were added in such as Ningxia and Anhui. In the last period, the numbers of the cluster decreased and gradually formed as the northwest and southwest regions. This cluster is somewhat isolated from other regions, not only because its remote location, but also for its particular socioeconomic history and developing patterns. It shows the lowest level in all dimensions in the first year. The MEI score of this cluster was ranked the last of all clusters. The GRP per capita, consumption level and compensation were similar with the third cluster, however, the FDI level and international trade balance were much lower than the third cluster (the FDI is only 13.92% of the third cluster and 1.64% of the first cluster), which indicated the occlusion of this cluster on some degree. In the last year, the FDI level increased 5% compared with the first year, and its gap with the third cluster decreased (45.84% of the third cluster) whereas the gap with the first cluster enlarged (9.08% of the first cluster). The SII also reported as the last of all the clusters. The governmental expenditure proportion in science and innovation was always the lowest and even decreased slightly which further increasing the gap with other clusters. The gap with third cluster of the labor productivity and income per capita in high science and technology firms was also enlarged dramatically in the last year (the differences were from 2.47% and 31.62% to 78.65% and 69.43% respectively). The R&D expenditure ratio and trading value proportion in science and technology market were increased 52.56% and 5.39%, and both very similar with the third cluster, however still the lower than other clusters. It is reasonable considering the relative backwardness of this cluster. The ESI report of this cluster is the dimension with the smallest gap with other clusters. In the last year, it is even higher than the third cluster (8 points), which, as far as I concerned, may be contributed to its relatively underdeveloped industry and thinner population density, so the waste discharges are not as tense as other clusters. It could be certified somehow from the lowest level of living waste discharge. However, other indicators such as the discharging industry waste water meeting standards and utilization of industry waste are still in a low level, indicating the weak link of sustainability of this cluster. In the HCI dimension, the literature population proportion and working age labor distribution are similar with other clusters, showing an average level of labor force in the cluster and the homogeneous tendency of these two indicators in the whole nation. However, the urban population proportion is significantly lower than other clusters, only half of the first cluster (50.25% in the first year and 51.91% in 2010), indicating the slow process of urbanization in this cluster. The employment ratio of the third sector compared with the first sector of this cluster is higher than the third cluster (0.32% in 1998 and 8.86% in 2010), also certifying that the industry is not so developed in this cluster as I assumed previously. In the PFI dimension, all the indicators are in the similar level with third cluster. The governmental expenditure in public service, public vehicles and public schools per capita are even higher than the first cluster in both years (for governmental expenditure in public service, 1.96% in 1998 and 0.35% in 2010; for public vehicles[38], 8.05% in 1998 and 13.70% in 2010; for public schools, 23.93% in 1998 and 23.52% in 2010).

[38] The relatively higher level of public vehicles and public schools per capita may be caused by the thin population density of this cluster.

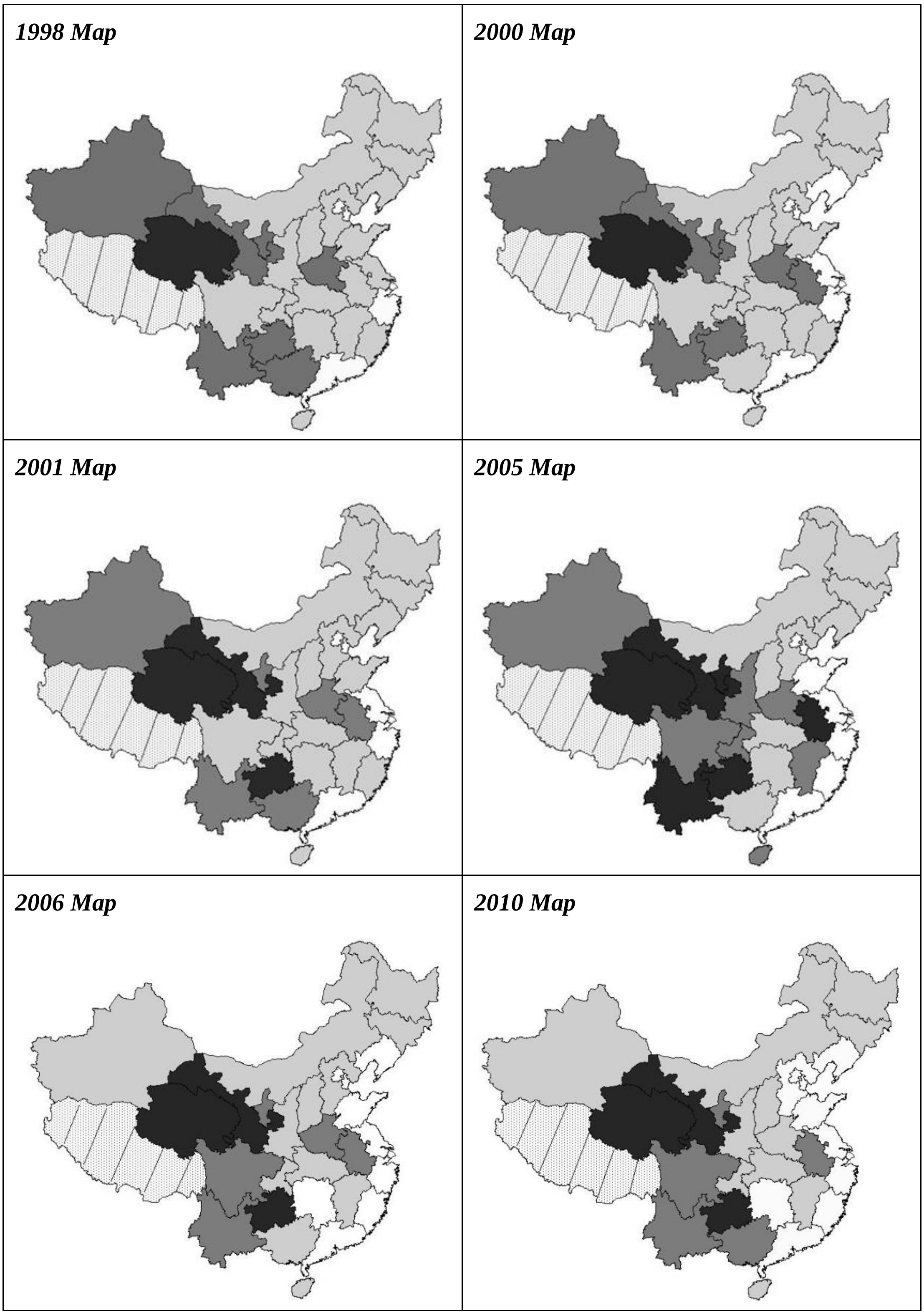


Figure 5 - *Mapping China: Five Dimensions 1998, 2000, 2001, 2005, 2006, 2010*

Source: The maps are drawn according to the cluster analysis result, in which Tibet was excluded for the missing data. The software used here is ArcView GIS.

Note: the colors of light-to-dark indicate the levels from high-to-low; the office map with names of administrative districts is reported in Appendix 7.

Generally speaking, it is obvious that the whole socioeconomic situations China improved. The east coastal regions are spreading their effects from several singles points in 1998 to a long belt in 2010. Even some inland regions such as Hunan, Hubei and Liaoning, are influenced by their spilling-over effects. If we compared maps, there is a trend of economic expansion from east coast to inland. It indicated the impact of the developed regions started to radiate into inland China. From the map 1998, it is clear that only several single points were in the light color, representing the underdevelopment of inland regions. Even though the east coast was in a relatively developed condition compared to inland in this period, most of the regions, if compared with the following years, were just in the level of low economic performances with the min-max sub-scores ranged from 50 to 60, except Shanghai was in the level of average economic performance. The missing parts of good and highest levels in 1998 signified the underdevelopment of the whole country. The fault lines had existed among different regions in 1998's China.

From the map 2010, it is obvious that the group of high-developed economies sharply increased to a long belt locating in the east coast. Nearly half of inland regions contiguous to the east coast improved to the average level, while the other half locating in the west upgraded to the level of low economic performances. The east coast became more advanced compared with 1998. Besides, Shanghai, Beijing and Tianjin ascended to the top group, in the range of 88 to 100; also, the single point (Shanghai) in 1998 was developing to a discontinuous belt in the coastal area. However, the gap inside the east coast still remains, and the fault lines showed in the whole China in 1998 does not disappeared.

6. Application of CIRD: convergence estimation

The previous cluster analysis showed four socioeconomic clusters have been formed up with homogeneity in development paths of their members inside yet heterogeneity against other clusters. It leads me to the assumption that a tendency of club convergence of socioeconomic conditions might exist in China, which sort of debated upon the commonly-accepted conception that the regional disparity in China has been going through an increasing spatial disparity since its 1978 reform. In this section, I will focus on the estimation of the regional developmental trajectory, trying to answer the complicate question that whether there is a tendency of convergence or divergence in Chinese provinces and municipalities. The previous studies on the estimation of convergence trend in China mostly just observed the income dimension. Instead, this study will base on the CIRD to evaluate the comprehensive socioeconomic development in China, which still remains as a gap in the researches of convergence of China. Here I will further apply the dataset of CIRD that obtained from the previous study to estimate the convergence trend in Chinese provinces and municipalities. Five sub dimensions of MEI, SII, ESI, HCI and PFI are analyzed respectively on the basis of their previous PCA results, and then the aggregated CIRD are analyzed with the combined components extracted from the integral PCA. The tests of σ-convergence, β-convergence and Kernel density estimations are presented respectively in this section.

6.1. Convergence of MEI

In this part, I will analyze the convergence tendency of macroeconomic conditions by tests of σ-convergence, β-convergence and Kernel density estimations. As many previous researches on the convergence trend of GDP or income per capita in China, it shows a tendency of convergence. I here applied the estimation pillar beyond GDP, to better describe the development of comprehensive economic conditions in China.

In Table 10, it shows the test results of σ-convergence in macroeconomic conditions across different time intervals from 1998 to 2010. In the second period from

2001 to 2005, it is pretty clear that there is no convergence during these years. The premise of β-convergence cannot be satisfied, indicating that one at relatively higher level increased faster since π>1. T_1 smaller than its critical values shows that the variances of the starting year is smaller than ending year, meaning that the differences among different regions in the last year increased, which is further proved by the T_2 test that falls into the range of critical values, so it is not reasonable to reject the null hypothesis of non-convergence. In the third period from 2006 to 2010, there is a Type II error for the test T_1 as its conclusion of acceptance of null hypothesis has been overturned by the following tests T_2 and T_3. The tricky part concentrates on the interpretation of the three tests in the whole period of 1998-2010 and the first period 1998-2000. In the whole period, π is smaller than 1, indicating the existence of β-convergence, whereas is not a sufficient condition to determine σ-convergence; T_1 is smaller than its critical value, but it is still too soon to determine to accept the null hypothesis only based on it since there might be Type II error, and this requires us to go further to the T_2 and T_3; however, these two tests still blur the real situation of convergence or divergence: T_2 verified the validity of null hypothesis whereas T_3 rejected. In the first period from 1998 to 2000, it was sort of contrary as T_3 fall inside the range of critical value but T_2 rejected the null hypothesis. In my point of view, this blur might also be caused by the short time interval (only 3 years) and the similarity of the variances in two time points.

Table 10 - *σ-convergence estimation of MEI*

	Π	T_1	T_2	T_3	N	
1998-2010	0.98	1.17	0.79	2.41	31	*σ-convergence*
1998-2000	0.99	1.19	7.38	-6.38	31	*σ-convergence*
2001-2005	1.003	0.99	0.01	—	31	*Non-convergence*
2006-2010	0.98	1.49	12.82	6.73	31	*σ-convergence*

Source: this table is calculated by the author on the basis of sigma-convergence theory. T_1 test is one-tail test under P=0.05 with a freedom F (N-2, N-2) = (29, 29), the critical value is 1.86; T_2 test is chi-square distribution with freedom F=1, and the critical value is 3.841; T_3 test is standardized normal distribution with P=0.05, and the critical value is 1.64.

In a word, the macroeconomic situations in the 31 Chinese provinces[39] tend to show σ-convergence and β-convergence during this time interval (1998-2010) with an exception in the sub-period 2001 to 2005. However, the methodology is more relied on the variances in the initial time and ending year, with a leak of the evolvement during the middle years. In the test of macroeconomic situations in different Chinese provinces, the two variances at initial year and ending year are happened to be very similar with each other, and hence the conclusions of the three tests show some confusion.

In order to clearly picture growth rate and initial level of MEI in each individual province and municipality, I specially estimate the β-Convergence of macroeconomic situations in the 31 Chinese provinces and municipalities and benchmark the provinces with their performance in β-convergence of MEI, which is the relation between the initial condition and the long-run growth rate in the time interval 1998-2010. From Table 11 it shows a negative parameter of β, indicating the existence of β-Convergence in MEI, which means poorer regions with lower initial levels in macroeconomic conditions grow faster in the following years. However, the estimation of this method is affected by the small sample number of observations and the high probability of omitting important time-invariant and country- specific effects.

Table 11 - *β-convergence estimation of MEI*

				N. obs = 31
				F(1,29) = 8.24
				Prob>F = 0.0076
				R-squared = 0.3491
Δy (1998-2010)	*coef.*	*st.err*	*P>| t |*	*β*
y1998	*-0.02*	*0.01*	*-2.87*	*-0.39*
Cons	*0.00*	*0.00*	*0.19*	
(i) Δy is the average annual growth rate of MEI during 1998-2010				
(ii) y1998 is the MEI at initial time 1998				
(iii) cons is the constant				

Source: it is calculated by the software STATA 12.0.

Figure 6 identifies the negative correlation between the initial level of macroeconomics and the growth rate during 1998 to 2010. In this dimension, the

[39] Here Tibet is included in this dimension.

distribution was a little bit scattered, the east-coastal regions commonly accepted as developed areas all located in the right part with a high initial macroeconomic level, however with a normal growth rate around national average. These regions have already gone through the first-phase of fast development, so the marginal effects such as capital or investment were decreasing, which might be the reason of the around average growth rates in these regions. On the contrary, the underdeveloped regions with lower initial macroeconomic levels possessed higher marginal profits in their primary developmental phase, so these regions such as Ningxia, Qinghai, Shaanxi, Shanxi, and Inner Mongolia showed a higher growth rate. The one requiring special attentions was the inland area with both relatively lower initial macroeconomic levels and growth rates, these regions, such as Henan, Anhui, Gansu, Sichuan, and Hubei, on one hand, neither possessed the connatural superiority as east-coast, nor obtained as many as financial supports from the central government as the west border areas on the other hand, therefore, these regions developed in a relatively slower rate. It also implied the requirement to further improve the "Go-Central" program initiated in 2004, hence to intensify the role of transition and coordination of this zone.

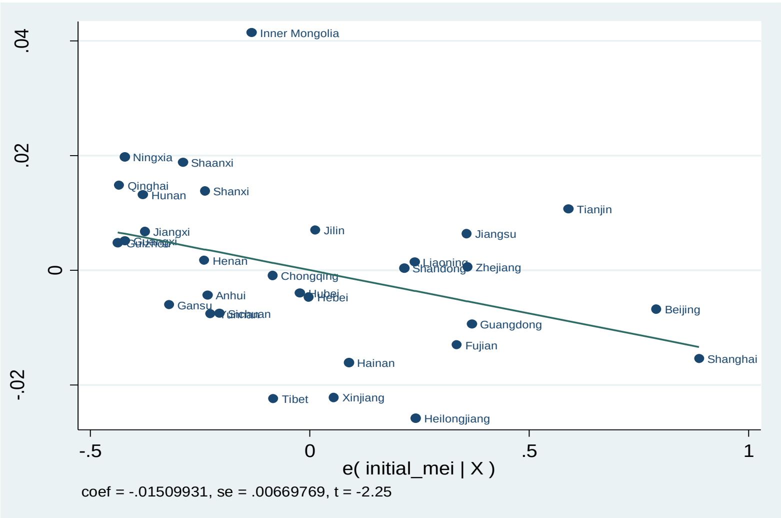


***Figure 6** - The relation between the average growth rate (1998-2010) and the initial level of Macroeconomic Index (1998)*
Source: data from author's own calculation and drawn by STATA 12.0

The previous analyses were just based on the single time points rather than the whole trajectory covering the time interval. Hence, I use kernel density estimation in the following to better picture the convergence trend. The kernel density curves provide a clear picture of the macroeconomic distribution of 31 Chinese provinces in the three time points 1998, 2005 and 2010 (Figure 7). In the beginning year 1998, there were two main modes clustering around the value 0.65 and 1.25, and another two small modes in the right tail around 1.7 and 2.2 times. It indicated the formation of two main clubs and two potential sub clubs, where the two main clubs possessed relatively large sample size. If characterizing them by their relative locations in the coordinator setting, these two main clubs could be classified as a lower-level group and a club with above average level in macroeconomic situations. The lower-level club with the highest peak implied its relatively larger member size compared with the other clubs, whereas its average macroeconomic situations were only 65% of the average as observed in the x-axis. The above-average club was smaller than the low-level club in its membership size, whereas its average macroeconomic situations were 1.25 times of the national average. The other two potential sub clubs were both in a relatively small membership size when compared with the previous two main, since they were just shown up as small bumps in the right tail. These two small sub clubs were in upper class of their macroeconomic conditions, which can be defined as developed club and superior developed club. Their emergence indicated a potential tendency of polarization among different clubs. This assumption was also observed and further confirmed in the middle year 2005. The macroeconomic distribution was more scattered in the right tail, and the previous above-average club in 1998 almost decomposed into two sub modes. On the other hand, the previous lower-level club still remained in the same position in 2005 yet with a larger member size than in 1998, as there might be some new members added into with the decomposition of the above-average club. In the final year 2010, a new arrangement was established on the basis of the former two. Two major clubs was formed and could be defined as club with proxy to average and club with higher level. Another potential sub club could also be observed in this bandwidth as integrated club above average. The club with proxy to average was still shown as the largest club in its membership, but its average level of macroeconomics increased to 0.9, closer to the national average. The integrated club above average, if estimated with larger bandwidth to sharp its distribution then could be divided into two sub clubs: one gathered around 1.15 times of national average and the other gathered around 1.5 times of national average. It could be recognized as a divergence or polarization inside the previous above-average level club in 1998. The club with higher level reached its largest membership if compared with the previous two years. However, its macroeconomic level decreased a little bit, around 2 times of average. It indicated the gaps between the clubs were decreasing in these years, and an overall convergence could be assumed among the clubs.

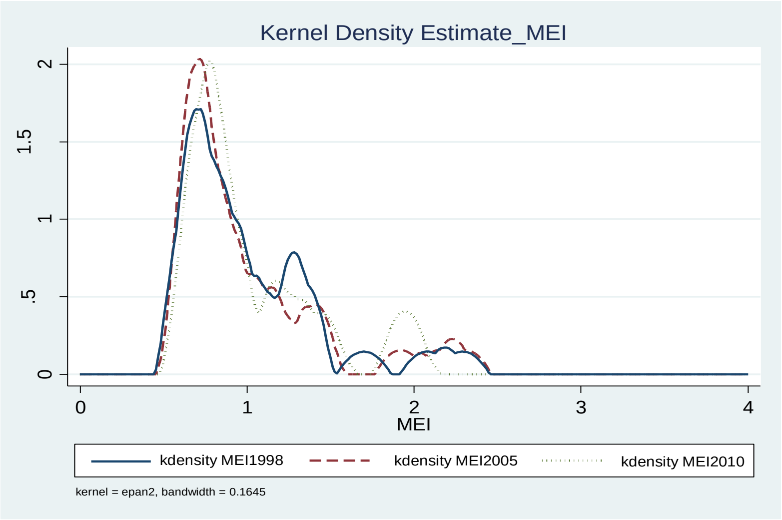


Figure 7 *- Kernel Density Curves of Macro Economic Index (MEI) 1998, 2005, 2010*
Source: this figure is drawn according to the data collected and calculated from the Chinese Statistical Yearbook by the software STATA 12.0.

All in all, the kernel density curves reflected a tendency of multi club convergence in macroeconomic situations, which could also be interpreted as two main clubs could be observed with sub clubs decomposed. There were still a large proportion of Chinese regions staying below national average in their macroeconomic levels as the peak of the left mode was always higher than the other. The gaps between the two clubs decreased as the poorer started to catch up to the richer. In order to better estimate the whole trend, I extend my study to contour plot and kernel surface to estimate the macroeconomic distribution changes in the whole time interval 1998-2010.

The contour plot in the left side of Figure 8 clearly reflected the existence of club convergence on macroeconomic situations of the Chinese provinces. A pattern of dual clubs formed in the trend of development, as the poor regions clustered into one club while the rich regions grouped into another. The club with poor regions clustered around 0.8 times of average value, and the club with rich regions clustered around 2.2 times of average level. Also, the whole curves incline to move in an approach of counterclockwise, indicating inner convergence inside each club. It could also be observed the two clubs are getting closer to each other with some of their members approaching to the national average. Hence, it is safely to conclude the poorer regions started to catch up with the rich ones, whereas no clear sign for the absolute convergence. It was indicated more solid in the following three-dimensional kernel surface at the right side of Figure 8. The sharp turning angle in the ridge could also indicate potential of multi clubs as sub clubs existed beneath the main clubs. The kernel surface depicted two modes moving in a counterclockwise direction along with the diagonal. The higher mode[40] here was actually the lower bump (the group gathering around a higher level far above average) in the kernel curves, and here in the kernel surface it was higher owing to the more active transformations of the right bump (the lower bump) in kernel curves. To be more specific, the regions gathering around the higher level of MEI (above average) transformed more actively compared with the regions gathering around in the lower level of MEI (below or near average). It could be certified in the kernel curves, that the bumps in the right side moved variably in the following years 2005 and 2010, while the left (higher) bump almost kept in a similar shape.

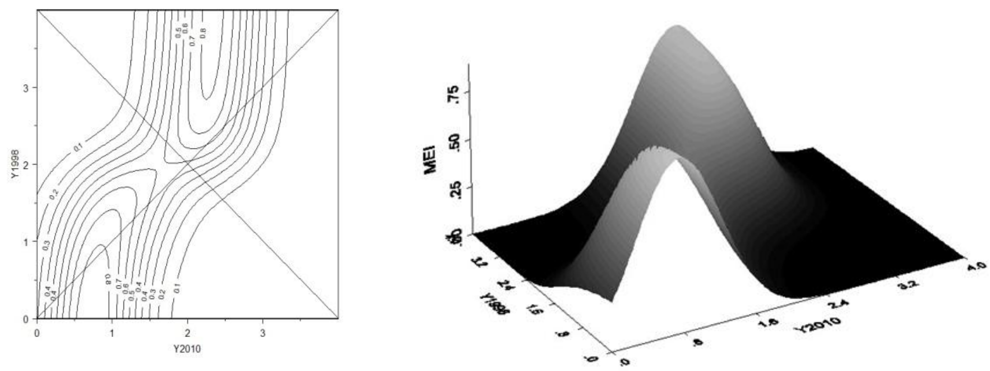


Figure 8 - *Contour Plot and kernel surface of Macro Economic Index (MEI) 1998-201*
Source: this figure is got from the dataset collected and calculated from the China Statistical Yearbooks 1999-2011 by the author, the software used are GAUSS and S-PLUS.

[40] Here I have to emphasize that the height of mode in kernel surface is different with in kernel curves. In kernel surface, the mode reflects the transformation activities, so the higher the mode, the more active the transformation; while in kernel curves, the bump (mode) represents the level compared with average, so the higher the bump (mode), the higher the level over average is.

6.2. Convergence of SII

In this part, I will analyze the convergence tendency of science and innovation development by tests of *σ-convergence, β-convergence* and kernel density estimations. The interaction of scientific progress with economic growth has been widely accepted. Science and technology constitute as the first productivity to promote economic growth, meanwhile the economic growth brings basic supports for scientific technology development. The feature of regional disparity was more evident in this dimension in China. Hence, whether it will present a similar tendency of convergence or an opposite way of divergence with macroeconomics will be an interesting issue to investigate.

Table 12 *- σ-convergence estimation of SII*

	π	T_1	T_2	T_3	N	
1998-2010	1.04	1.35	1.28	__	30	*non-convergence*
1998-2000	1.03	1.201	3.15	__	30	*σ-convergence*
2001-2005	1.04	1.07	0.26	__	30	*non-convergence*
2006-2010	1.01	1.21	1.11	__	30	*non-convergence*

Source: this table is calculated by the author on the basis of sigma-convergence theory.
Note: T_1 test is one-tail test under P=0.05 with a freedom F (N-2, N-2) = (28, 28), the critical value is 1.87; T_2 test is chi-square distribution with freedom F=1, and the critical value is 3.841; T_3 test is standardized normal distribution with P=0.05, and the critical value is 1.65.

Table 12 reported the test results of *σ-convergence* in science and innovation development in different time intervals from 1998 to 2010. It is pretty clear that a non-convergence tendency has been observed in the second and third sub-period and the whole time interval. The premise $\pi>1$ indicated that T_3 could not be computed so that the tendency of *β-convergence* could not be accepted. Moreover, since T_1 and T_2 all fall into the range of their critical values, implying that the variances of the starting years in the three sub periods and the whole time interval are all smaller than ending years, it further certified the assumption of non-convergence. The exception is the first period that the Lichtenburg test has committed a Type II error, since the T_2 rejected the null hypothesis. Therefore, we could observe *σ-convergence* in this period but without accompanying of *β-convergence*. Even though the tests for the whole time interval implied the unacceptance of convergence, it is still not strong enough to conclude the exact developmental trajectory in science and innovation is σ-divergence, since only two time points were estimated in the tests.

In the following I specially estimate the tendency of *β-convergence* of science and innovation development in the 30 Chinese provinces and municipalities. The main aim is to put yardsticks on each region by the relationship between its initial level of science and innovation and long-run growth rate in the time interval 1998-2010. Table 13 reported a negative and significant parameter of β, which implies the existence of β-Convergence in MEI, since one region starting at a relatively lower initial levels in science and innovation will possess a relatively higher growth rate in the following years to catch up with the ones in a higher initial level. This β-convergence tendency in SII shows a similar result of MEI. However, the small sample size also affected the validity of the test result where we can see the R-square is not high.

Table 13 *- β-convergence estimation of SII*

				N. obs = 30
				F(1,28) = 8.23
				Prob>F = 0.0077
				R-squared = 0.3520
Δy (1998-2010)	*coef.*	*st.err*	*P>| t |*	*β*
y1998	*-0.030364*	*0.0105814*	*0.008*	*-0.5933039*
cons	*0.0069728*	*0.0040065*	*0.093*	
(i) Δy is the average annual growth rate of SII during 1998-2010				
(ii) y1998 is the SII at initial time 1998				
(iii) cons is the constant				

Source: it is calculated by the software STATA 12.0.

In the Figure 9, it depicted that the initial level of science and innovation development was negatively correlated with its growth rate during 1998 to 2010. If concentrated on the geographical distribution, the east coastal regions were all stay in the right side of the coordinate setting, implying they are in the higher initial level of science and innovation, as we always considered. However, their growth rates were different. Beijing and Zhejiang were composed as two extreme values of high growth rate and low growth rate respectively, others mostly gather around the curve. Inland regions such as Hebei, Sichuan, Hubei, which began with a relatively low level of science and innovation, locating in the left side of the coordinate setting, showed a higher growth rate. Some peripheral regions, which were commonly accepted as underdeveloped regions in China, such as Qinghai, Guizhou and Guangxi, stayed at the bottom left, indicating that these regions were the ones out of the β-convergence tendency as they had lower levels in both their initial conditions and growth rates of science and innovation. Here we could see even though the general tendency of β-convergence in SII is similar to MEI, the region distribution was totally different. Instead of inland regions locating above of the tendency curve in MEI, here what had to be paid special attentions to were the isolated peripheral regions.

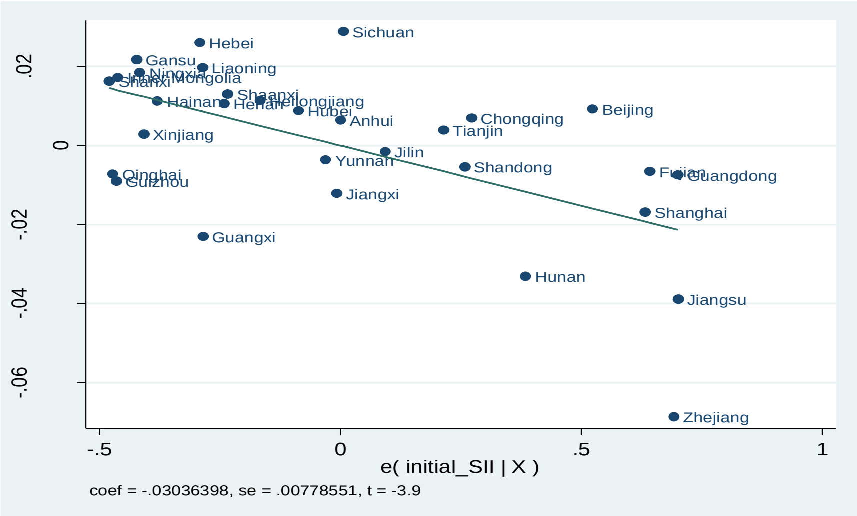


Figure 9 *- The relation between the average growth rate (1998-2010) and the initial level of Science and Innovation Index (1998)*
Source: data from author's own calculation and drawn by STATA 12.0

After the traditional tests of convergence, I apply the kernel density estimation to observe the dynamic trajectory of transformation in science and innovation among the 30 Chinese provinces and municipalities. Figure 10 depicts three single kernel density curves with the distribution of science and innovation Chinese provinces in the years of 1998, 2005 and 2010.

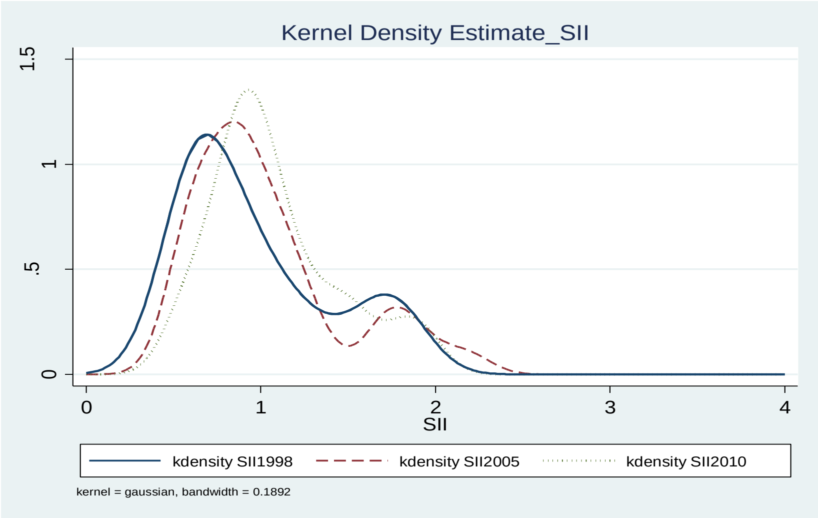


Figure 10 *- Kernel Density of Science and Innovation Index (SII) 1998, 2005, 2010*
Source: this figure is drawn according to the data collected and calculated from the Chinese Statistical Yearbook by the software STATA 12.0.

In the beginning year 1998, two modes gathered around the value 0.65 and 1.75 times of average, which reflects that two main groups showed up as a lower-level than average and a higher-level than average in science and innovation development. In the lower-level group, most members their science and innovation level were only 65% of the average; in the higher-level group, most members their science and innovation levels were 1.75 times over than the average. It is a clear sign to indicate the club-convergence in this year. In the middle year 2005, a club convergence still was shown by the curve, whereas the two main modes were getting closer. The left mode moved from 0.65 in 1998 to 0.8 in 2005 and also heightened up. It implied a larger part of regions in the low-level group grew in a faster rate trying to catch up with the high-level group. The right mode also moved a little bit to the right, gathering around 1.8 of national average. Even though the differences between the two groups got smaller, the two clubs still show no tendency to converge into one. In the final year 2010, the lower-level club moved even right and higher, indicating more regions were gathered in this group, and the average was approximate to 1, the national average level of science and innovation. On the other hand, the high-level group became less obvious in this year, implying a tendency of convergence with the low-level group. Its average of science and innovation also moved to the right, increased to nearly 2 times of national average.

To better picture the dynamic trend, contour plot and a three-dimensional kernel surface figure were provided to estimate the distribution changes of science and innovation in the whole time interval 1998-2010 in the following part. The contour plot in the left side of Figure 11 clearly pictured the convergence tendency of science and innovation among Chinese provinces and municipalities. The whole area was moving counterclockwise along with the diagonal line, and the upper part is almost parallel to the vertical line. It demonstrated an existence of club-convergence but with some complex disturbance. As I concerned, there might be some members in the upper level club, pulling the two clubs closer, but still incapable to change the club-convergence trend. The regions with relatively lower level of science and innovation were grouped a little bit below the national average level, while the regions with relatively higher level of science and innovation gathered around 1.5 times of national average. Hence, it is still too soon to justify that all the regions would converge eventually in science and innovation. In the right-side three-dimensional picture in Figure 11, the whole surface also showed a dual-mode club convergence and general counterclockwise movement along with the diagonal. I. The higher mode here, as discussed before, is actually the lower bump in the kernel curves. Since the transformation activities between 2005 and 2010 in this bump were more dynamic, here the mode in the kernel surface was higher. Or else to say, the regions gathering around the higher level of SII (above average) performed more active in their transformations.

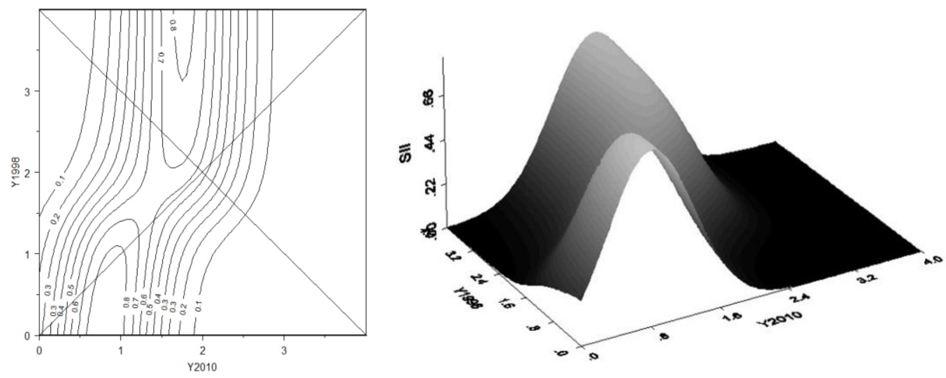


Figure 11 *- Contour plot and kernel surface of Science and Innovation Index (SII) 1998-2010*

Source: this figure is got from the dataset collected and calculated from the China Statistical Yearbooks 1999-2011 by the author, the software used are GAUSS and S-PLUS.

6.3. Convergence of ESI

In this dimension, the main focus was put on the environmental sustainability in the 30 Chinese provinces and municipalities (Tibet was excluded because of missing data). China has been criticized as sacrificing environment to the industry development. Indeed, the discharge of pollutants has been increasing with China's dramatic progress in economic growth. The central government started to put emphasis on the environmental protection in the late 1990s. However, there was a trend of heavy industry transforming from east coast to inland regions. This part investigated how the individual Chinese region performs during the time interval 1998-2010.

Table 14 *- σ-convergence estimation of ESI*

	π	T_1	T_2	T_3	N	
1998-2010	1.01	1.11	0.11	__	30	*non-convergence*
1998-2000	1.08	1.33	1.29	__	30	*non-convergence*
2001-2005	1.002	0.76	0.89	__	30	*non-convergence*
2006-2010	0.97	1.27	0.91	3.01	30	*(σ-convergence)*

Source: this table is calculated by the author on the basis of sigma-convergence theory.
Note: T_1 test is one-tail test under P=0.05 with a freedom F (N-2, N-2) = (28, 28), the critical value is 1.87; T_2 test is chi-square distribution with freedom F=1, and the critical value is 3.841; T_3 test is standardized normal distribution with P=0.05, and the critical value is 1.65.

The Table 14 tested σ-convergence of 30 Chinese provinces and municipalities (expect Tibet)[41] in environmental sustainability across the time intervals from 1998 to 2010. In the last period from 2006 to 2010, even though T_3 was over the critical level, it might still be a non-convergence since T_1 and T_2 fall into the range of critical levels. In the other time periods, it was pretty clear that there was no convergence during these years, β-convergence was first rejected since $\pi>1$, σ-convergence was further rejected as T_1 and T_2 tests concluded the acceptances of null hypothesis. The tricky part concentrated on the interpretation of the convergence tendency in the third period. π was smaller than 1, indicating a possibility of the existence of β-convergence; T_1 was smaller than its critical value, whereas the null hypothesis could not be justified only based on it since there might be Type II error; however, the following tests T_2 and T_3 blurred the real situation of

[41] Tibet has been excluded in this dimension because of the missing data in the selected variables.

convergence or non-convergence: T_2 indicated the validity of null hypothesis whereas T_3 rejected the existence of non-convergence. In a word, the environmental sustainability in the Chinese provinces showed no tendency of σ-convergence. However, the methodology is more relied on the variances in the initial time and ending year, with a leak of the evolvement during the middle years. In the test of environmental sustainability, the π values were all nearly around the value 1, which complicated the further tests, and hence the conclusions of the three tests showed some confusion.

In the following I specially estimate the β-Convergence in environmental sustainability, which is the relation between the initial condition and the long-run growth rate in the time interval 1998-2010. From Table 15, it showed a negative parameter of β, which means a poorer region with lower initial levels of environmental sustainability, grew faster in the following years. However, the estimation of this method was affected by the small sample number of observations and the high probability of omitting important time-invariant and country- specific effects.

Table 15 - *β-convergence estimation of ESI*

				N. obs = 30 *F(1,28) = 10.13* *Prob>F = 0.0036* *R-squared = 0.3488*
Δy (1998-2010)	*coef.*	*st.err*	*P>| t |*	*β*
y1998	*-0.044995*	*0.0141395*	*0.004*	*-0.5906196*
cons	*-0.001783*	*0.0046232*	*0.97*	
(i) Δy is the average annual growth rate of ESI during 1998-2010 *(ii) y1998 is the ESI at initial time 1998* *(iii) cons is the constant*				

Source: it is calculated by the software STATA 12.0.

Figure 12 identifies the negative correlation between the initial level of environmental sustainability and the growth rate during 1998 to 2010. In this dimension, the provinces with a growth rate lower than the average are Chongqing, Gansu, Anhui, Fujian, Guangxi, Henan, Jilin, Sichuan, and Liaoning. These regions are mostly industrial areas and locating in the inner land or northeast. On one hand, the development of industry intensified the pollution; on another hand, their geographical locations are not helpful for the self-discharge of fomites (such as the sulfur dioxide, dust and waste water).

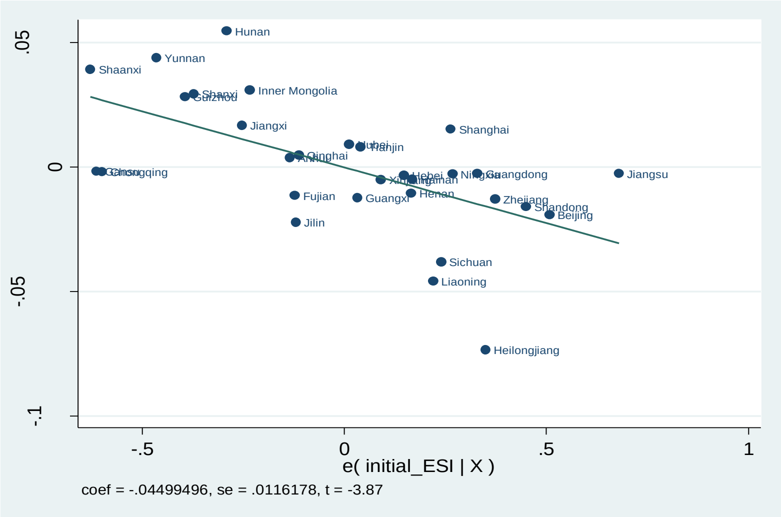


***Figure 12** - The relation between the average growth rate (1998-2010) and the initial level of Environmental Sustainability Index (1998)*

Source: this figure is drawn according to the data collected and calculated from the Chinese Statistical Yearbook by the software STATA 12.0.

In the following part I use kernel density estimation to better picture the dynamic convergence trend of environmental sustainability in a long time interval 1998-2010. In Figure 13, it reported three kernel density curves of environmental sustainability in the 30 Chinese provinces in 1998, 2005 and 2010. All the three curves showed a uni-mode trend, indicating a convergence tendency. In the beginning year 1998, the mode was more dispersed and flat, with a range of 0.6 to 1, implying the majority of regions clustered in this range in their performances of environmental sustainability but with a potential tendency of dispersion. Or else to say, there might be some small sub modes beneath this wide mode if changed the bandwidth, but their values were so consistent to compose as one general mode with this wide range. In the middle year 2005, the kernel density curve of ESI showed a dual-mode club convergence as one gathering around approximately to the national average level 1, and the other gathering around 1.8. The potential dispersion tendency inside the mode in 1998 was verified in this year. In the last year 2010, however, the club-convergence was not obvious as the in 2005, but still two modes could be observed in the curve. The left mode heightened up and moved to the right at the national average level 1, indicating more regions in this year converged around the average level of environmental sustainability. On the other hand, the right mode got lower, so there were fewer regions showing relatively high level in environmental sustainability in this year. To summarize, the kernel density curves showed the existence of convergence in environmental sustainability, but the existence of club convergence showed up later and slackened somewhat again in this ESI pillar.

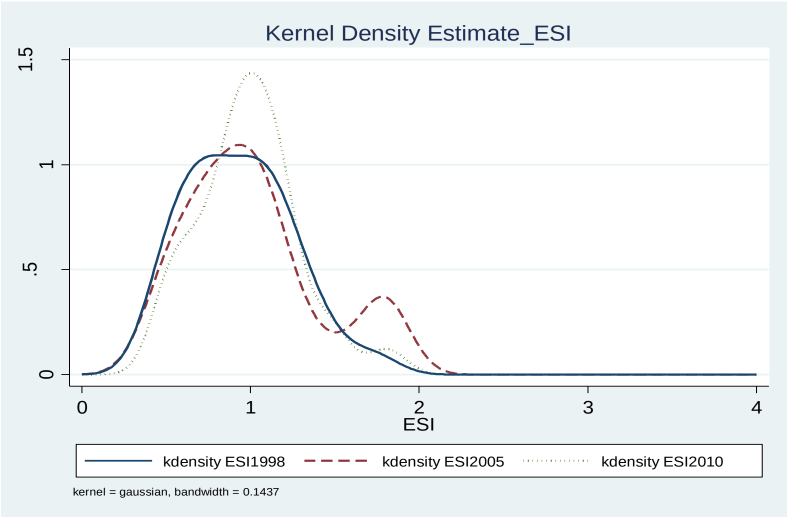


Figure 13 *- Kernel Density of Environmental Sustainability Index (ESI) 1998, 2005, 2010*
Source: this figure is drawn according to the data collected and calculated from the Chinese Statistical Yearbook by the software STATA 12.0.

In order to picture the dynamic trajectory more vividly, I used contour plot and kernel surface to describe the environmental sustainability in the whole time period 1998-2010 as follows. From the contour plot in Figure 14, it was clear that a club convergence trend existed in environmental sustainability dimension. The whole area was moving counterclockwise along with the positive-sloped diagonal line and almost parallel to the vertical line, so there was a possibility of convergence existed between the two clubs. It might be interpreted empirically that the regions with higher initial levels of environmental sustainability were deteriorating and hence the regions with lower initial levels could catch up or even performed better in some aspects of environmental quality than the initially higher-level regions. In the kernel surface in Figure 14, it was clearer that the whole surface was parallel to the vertical squared surface and two modes could be observed but not quite highlighting. It implied that it might be possible that the two clubs would converge together to uni-mode in a long run.

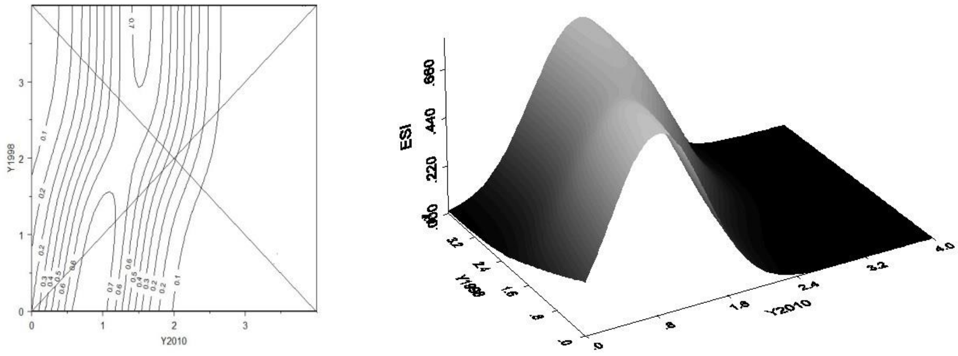


Figure 14 *- Contour Plot and Kernel surface of Environmental Sustainability Index (ESI) 1998-2010*

Source: this figure is got from the dataset collected and calculated from the China Statistical Yearbooks 1999-2011 by the author, the software used are GAUSS and S-PLUS.

6.4. Convergence of HCI

This part focused on the analysis of the convergence tendency in human capital development. It is a developed pillar of the well-know HDI, here I applied additional indicators referring to employment to better estimate the human resources in China, and disregarded the income since it has been involved in the MEI. China has been reported with an increase in HDI after 1980, and this part will give more specific picture of how the different regions in China have been performing in the past years, also with a more comprehensive evaluation. Table 16 shows the test result of HCI in the three sub periods and the whole time interval 1998-2010.

Table 16 *- σ-convergence estimation of HCI*

	π	T_1	T_2	T_3	N	
1998-2010	1.05	0.36	28.27	—	31	*σ-convergence*
1998-2000	1.01	1.09	3.78	—	31	*σ-convergence*
2001-2005	0.99	0.64	26.33	6.87	31	*σ-convergence*
2006-2010	1.03	0.75	5.36	—	31	*σ-convergence*

Source: this table is calculated by the author on the basis of sigma-convergence theory.

Note: T_1 test is one-tail test under P=0.05 with a freedom F (N-2, N-2) = (29, 29), the critical value is 1.86; T_2 test is chi-square distribution with freedom F=1, and the critical value is 3.841; T_3 test is standardized normal distribution with P=0.05, and the critical value is 1.64.

In the whole time period 1998-2010, the first period 1998-2000 and the third period 2006-2010, it was pretty clear that there was no *β-convergence* during these years since the premise of convergence as $\pi<1$ was not satisfied, indicating the existence of a positive relation with initial level of HCI and its growth rate. However, the *σ-convergence* existed in all the time periods since it could show up without *β-convergence* as discussed before. The T_1 were smaller than the critical values, however the T_2 (and T_3 in the second period) rejected the null hypothesis with their outranged values, which implied that T_1 committed Type II error. The exception was the second sub period 2001-2005, which showed a coexistence of *σ-convergence* and *β-convergence* of HCI. The *Lichtenburg test* here also committed a Type II error, as the T_1 showed smaller value than the critical value whereas T_2 and T_3 outranged of their critical values to reject the null hypothesis of no convergence. To summarize, the human capital conditions in the whole time interval from 1998 to 2010 did not show a tendency of *β-convergence*, as the regions with underdeveloped level of human capital also grew slowly (or perhaps even retrogressed) and the regions with developed level of human capital grew faster. This tendency was also shown up in the first and last sub periods in 1998-2000 and 2006-2010. However, in all the time periods, there was a tendency of *σ-convergence*, indicating the regions with lower level of human capital started to catch up with the high developed regions, whereas the high developed regions grew relatively slowly, therefore they inclined to gather to a same level. In the second period, the *σ-convergence and β-convergence* coexisted. As I investigated the original indicators, this period was also the one with the most obvious changes in the employment structure. It had been reported in enormous documents that China had confronted a transformation in the labor force structure during that period, arguing that whether the Lewisian Turning Point. On the other hand, an interesting phenomenon had emerged in the east coast called immigrant labor shortage. Those issues implied the complex situation in that period due to the structure transformation in labor force, so it is empirically reasonable that a convergence tendency was shown up in that period. Nevertheless, the test results might be biased by the approximated level of dispersion in the first and last years, as π proxy to 1.

To better picture the performance of each individual province or municipality in the human capital development, I made a specific estimation of β-Convergence to benchmark the HCI among 31 Chinese provinces and municipalities, according to the correlations between their initial conditions and their long-term growth rates from 1998 to 2010. Table 17 shows a positive parameter of β, indicating the inexistence of β-Convergence in HCI.

Table 17 *- β-convergence estimation of HCI*

				N. obs = 31
				F(1,29) = 4.39
				Prob>F = 0.0450
				R-squared = 0.3912
Δy (1998-2010)	*coef.*	*st.err*	*P>| t |*	*β*
y1998	*0.0246448*	*0.01176*	*0.045*	*0.4373774*
cons	*-0.015758*	*0.0027117*	*0.566*	

(i) Δy is the average annual growth rate of HCI during 1998-2010
(ii) y1998 is the HCI at initial time 1998
(iii) cons is the constant

Source: it is calculated by the software STATA 12.0.

In the Figure 15, the majority distributed in the left-bottom area, indicating that a large part of Chinese provinces and municipalities had been undergoing a relatively low initial level of human capital in 1998 and also with a slow growth rate from 1998 to 2010. Geographically, these regions were all inland regions and most of them locate in the peripheral regions that with less mobility for the labor migration, low level of education and high proportion of agricultural employment and rural population. The regions in the left side but above the linear regression curve possessed a relatively higher growth rate of HCI from 1998 to 2010, but this rate was relative compared with the regions in the left-bottom area. Some of the regions such as Hebei and Anhui developed fast in HCI might be caused by heir adjacency to the east coast, while some peripheral regions such as Tibet and Yunnan also showed a relatively higher growth rate might be attributed to the great supports and investments that the central government had been putting into education and industrial structure transformation. However, this is only two-time-point analysis, so it cannot be determined that these regions commonly accepted as underdeveloped grow relatively faster in HCI among all these years. Additionally, this higher growth rate was relative; if compared with the developed regions, their growth rate was still low. In the right side, all regions belonged to the east coast, which was reasonable since they had better initial levels in HCI. However, there was also dispersion inside these regions, which was mainly shown in their different initial levels of HCI. Beijing and Shanghai, as always, reflected a superior developed level in both their initial conditions and growth rates, whereas other east coastal regions, such as Jiangsu, Zhejiang and Guangdong, stayed a little far away with the superior developed regions but grow even faster (especially Guangdong as the fastest-growing region). Tianjin and Liaoning, on the contrary, showed slower growth rates even though they had a relatively better initial condition in HCI. To summarize, the difference, or else to say, the dispersion among inland regions, stating in the left side, was mainly shown up in their growth rates, while they all had similar initial levels of HCI. On the contrary, the main dispersion among the east coastal regions, locating in the right side, was reflected as their gaps in the initial conditions of HCI.

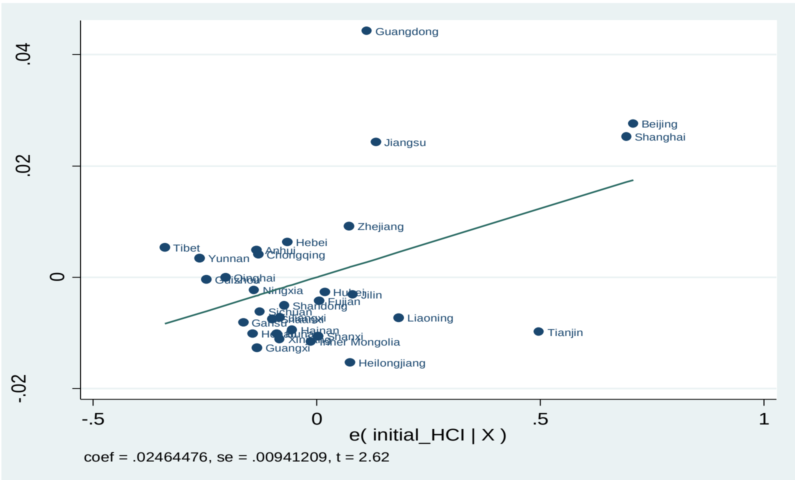


Figure 14 *- The relation between the average growth rate (1998-2010) and the initial level of Human Capital Index (1998)*
Source: data from author's own calculation and drawn by STATA 12.0

Since the unsatisfied depictions of the convergence trajectory in HCI above, the stochastic kernel density estimation was applied to draw the dynamic transformation of human capital in 31 Chinese provinces and municipalities in the three time points 1998, 2005 and 2010. In Figure 15, it showed three kernel density curves in 1998, 2005 and 2010. In the beginning year 1998, there was a multi-mode club convergence showing up, with one main mode gathering around the average level and two small sub modes gathering around the value 1.55 and 1.95, representing that a large proportion of Chinese regions grouped in the average level in human capital, while some proportions grouped in the above average level and superior high level of HCI. In the middle year of 2005, the multi-mode club convergence remained. It showed a tendency of dispersion in human capital between the high-developed group and underdeveloped group, as the right mode moved further while the left mode almost stayed in the same position. On the other hand, there was a tendency of convergence inside right mode as the two sub modes were clustering together. In the last year of 2010, the dispersion of the left mode and right mode even got enlarged. The left mode gathered around 0.8 while the right mode gathered around 2.55, which means that the low developed group showed an average level of human capital below the national average, while the high developed group showed a higher level of 2.55 times of national average in human capital. In the high developed group of HCI, the two sub modes also gathered together into on group in this year, indicating an inner convergence inside this group. However, the left peak was lower than in 2005, indeed implied less regions converging in this group, whereas did not exactly reflect a divergence inside this group. As the number of observed samples is only 31, this lower peak might be just caused by a reduction of one or two provinces.

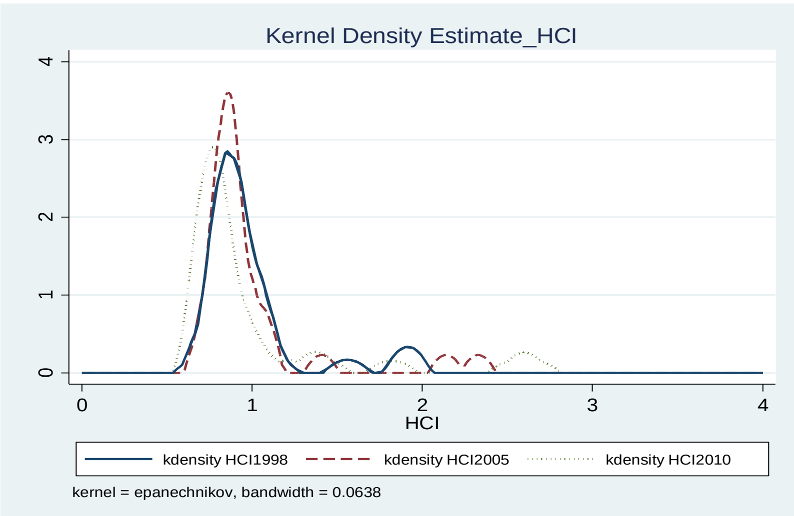


***Figure 15** - Kernel Density Curves of Human Capital Index (HCI) 1998, 2005, 2010*
Source: this figure is drawn according to the data collected and calculated from the Chinese Statistical Yearbook by the software STATA 12.0.

To better interpret the tendency, I applied contour plot and a three dimensional kernel surface in the next part to estimate the distribution changes of human capital in the whole time interval 1998-2010. In the contour plot and kernel surface of Figure 16, it showed a club convergence in HCI. Two clubs are formed and keep persistently away from each other, as the two modes locate in the positively sloped diagonal. Here I assumed that the right small bumps in the kernel density curve in Figure 15 were gathered as one mode. However, inside the two modes, the convergence showed up as they all were a parallel tendency to the vertical edge or surface. To be more specific, the 31 Chinese provinces and municipalities converged into two separate groups according to their performances in human capital development and formed into two clubs over the time interval 1998-2010, but these two clubs had been growing independently with no tendency to converge together. Also, a possibility of multi club convergence might be the reason for the clear ridge turning point between these two clubs. The transformation was also much tenser in the upper level club as the right mode in kernel surface was higher, which could also be verified by the kernel density estimation that the right mode was more active in their changes than the left one.

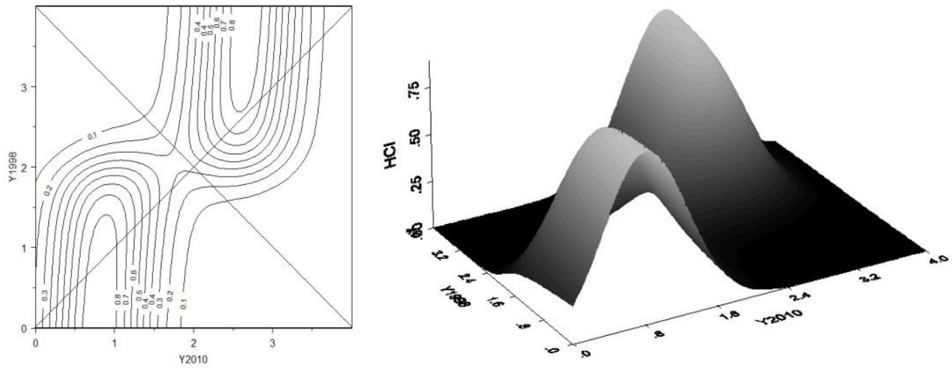


Figure 16 - *Contour Plot and Kernel Surface of Human Capital Index (HCI) 1998-2010*
Source: this figure is got from the dataset collected and calculated from the China Statistical Yearbooks 1999-2011 by the author, the software used are GAUSS and S-PLUS.

6.5. Convergence of PFI

In this part, I mainly focused on measuring the convergence tendency in public facility conditions in the 31 Chinese provinces and municipalities (Tibet was included). China has been generally recognized for its extraordinary accomplishment in infrastructural construction, while the unordinary phenomenon of city slums and binary structure of urban-rural areas which are universally observed in developing economies still cannot be prevented in China’s development. This part investigates how the individual regions in China performed in their pubic facility conditions beneath that dramatic but incongruous dynamics during 1998-2010. Table 18 showed the test results of PFI in the three sub periods and the whole time interval 1998-2010.

Table 18 - *σ-convergence estimation of PFI*

	π	T_1	T_2	T_3	N	
1998-2010	1.199	0.65	8.35	__	31	*σ-convergence*
1998-2000	1.03	1.05	0.44	__	31	*non-convergence*
2001-2005	1.04	0.95	0.25	__	31	*non-convergence*
2006-2010	1.11	0.78	5.27	__	31	*σ-convergence*

Source: this table is calculated by the author on the basis of sigma-convergence theory.
Note: T_1 test is one-tail test under P=0.05 with a freedom F (N-2, N-2) = (29, 29), the critical value is 1.86; T_2 test is chi-square distribution with freedom F=1, and the critical value is 3.841; T_3 test is standardized normal distribution with P=0.05, and the critical value is 1.64.

In the whole time period 1998-2010 and the three sub periods 1998-2000, 2001-2005, and 2006-2010, it is pretty clear that there was no *β-convergence* of public facility conditions in these time intervals since the premise of convergence of π<1 cannot be satisfied. It reflected that there were positive correlations with initial levels of PFI and its growth rate, which could be interpreted as the regions with lower initial levels of public facilities grew in relatively slower rates, while the regions with higher initial levels of public facilities grew faster between 1998 and 2010. Hence, it is solid to conclude that there was no tendency of *β-convergence* in these periods. However, the inexistence of *β-convergence* cannot directly deduce the corollary of inexistence of *σ-convergence* in PFI. In the whole time interval 1998-2010 and the third time period 2006-2010, this conclusion has been further established by larger T_2 over than critical values, indicating that the Lichtenburg test had committed Type II error as the T_1 were smaller than the critical values in these periods. Therefore, it is safely to conclude that *σ-convergence* existed in these periods. However, in the first period 1998-2000 and middle period 2001-2005, T_2 also fall into the range of critical values, which meant acceptance for the null hypothesis of non-convergence. The conflict, in my point of view, might be caused by two reasons: firstly the variances of the first year and the last year in these two time periods were similar as their π are proxy to 1, so it blurred the test results; secondly the time interval was not large enough to observe the convergence tendency clearly.

To better depict the developments of public facilities in each individual province or municipality, *β-convergence* was estimated with a benchmarking picture of the development of public facilities in 31 Chinese provinces and municipalities, based on their correlations between their initial conditions and their long-term growth rates from 1998 to 2010. Table 19 shows an insignificant parameter of β with a slight negative value, therefore the *β-convergence* has been further proved as inexistence in PFI.

***Table 19** - β-convergence estimation of PFI*

				N. obs = 31
				F(1,29) = 1.34
				Prob>F = 0.2563
				R-squared = 0.0215
Δy (1998-2010)	*coef.*	*st.err*	*P>| t |*	*β*
y1998	*-0.049*	*0.043*	*0.256*	*-0.1465*
cons	*0.0135*	*0.0015*	*0*	

(i) Δy is the average annual growth rate of PFI during 1998-2010
(ii) y1998 is the PFI at initial time 1998
(iii) cons is the constant

Source: it is calculated by the software STATA 12.0.

In Figure 17, the provinces and municipalities showed a scattered distribution, indicating the inexistence of *β-convergence*. The east coastal regions such as Beijing, Shanghai, Zhejiang and Jiangsu located around the linear in the right edge, indicating that these regions have relatively high levels of public facilities in the first year 1998, and develop in national average rates of growth. Other east coastal regions such as Guangdong, Fujian and Tianjin, show better performances in the growth of their public facilities even though their initial conditions of public facilities are not as best as the superior developed regions mentioned previously. Among these east coastal regions, Guangdong shows a highest growth rate in its public facility development. In my point of view, this fast growth cannot be separated with governmental supports and investments. The establishment of Special Economic Zones (SEZ) in Guangdong province in 1980s (Shenzhen and Zhuhai cities) dramatically stimulate the development in the general province, especially in the infrastructural construction. This reason can also explain the relatively fast growth in Fujian (as Xiamen and Shantou are two SEZ in this province) and Hainan (as Hainan is SEZ). In the middle part of the settings, the region distribution is more scattered, which might cause the whole spotted. Some central inland regions such as Hubei and Hunan show relatively higher growth rates, while some northern inland regions such as Shanxi and Henan show relatively lower growth rates. Peripheral regions also spread out in their distributions. Some regions such as Qinghai and Guangxi show relatively high growth rates in their public facility development between 1998 and 2010, while some regions such as Inner Mongolia and Yunnan show relatively low growth rates in their public facility development between these two years. Tibet, however, shows a national average level of growth rate in its public facility growth rate.

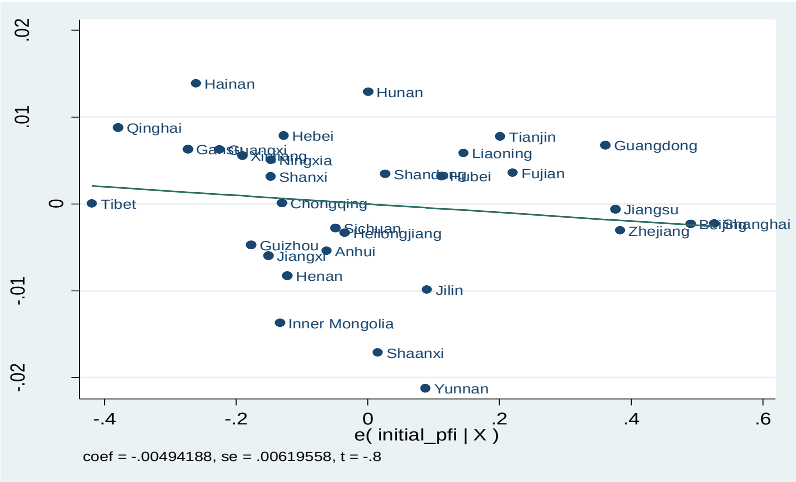


Figure 17 *- The relation between the average growth rate (1998-2010) and the initial level of Public Facility Index (1998)*
Source: data from author's own calculation and drawn by STATA 12.0

This situation was complex to explain the reason, and also smuggled the exact trend of convergence since the invalidity of convergence test in the first year 1998 and last year 2010. Therefore, I used the stochastic kernel density estimation to describe the dynamic evolvement of PFI in the 31 Chinese provinces and municipalities with the consideration of all intermediate years and hence to better define the convergence trajectory.

Figure 18 depicted three kernel density curves in 1998, 2005 and 2010 respectively to provide an initial picture of the dynamic convergence trend in public facility development. In the first year 1998, the majority was clustered around the value of 0.9 of national average, with a slight bump above average at the value of 1.5. In the middle year 2005, the mode moved slightly to the right, and gathered around a value proxy to the national average level. However, the peak was lower than the first year, indicating fewer regions were included in this majority, and the slight bump also moved a little to the right side. In the last year 2010, the mode gathered around the national average level, and the previous slight bump was becoming obvious, and moved to the value proxy to 1.6. The peak in this year kept getting lower, so the number of majority reduced. However, as I discussed before, this diminish cannot significantly imply a trend of divergence since the sample size is only 31, so the movements of one or two single regions may cause a change of the curve. To summarize, the public facilities of 31 Chinese provinces and municipalities show a convergence tendency from the kernel density estimation in 1998, 2005, and 2010. Even though the peaks of the three curves are getting lower, the general shape of the curves still approximate to uni-mode shape. A club convergence tendency is emerging in the public facility development among Chinese provinces and municipalities, as the relatively smaller mode is increasingly obvious. To better interpret the dynamic trajectory, contour plot and kernel surface will be applied in the following.

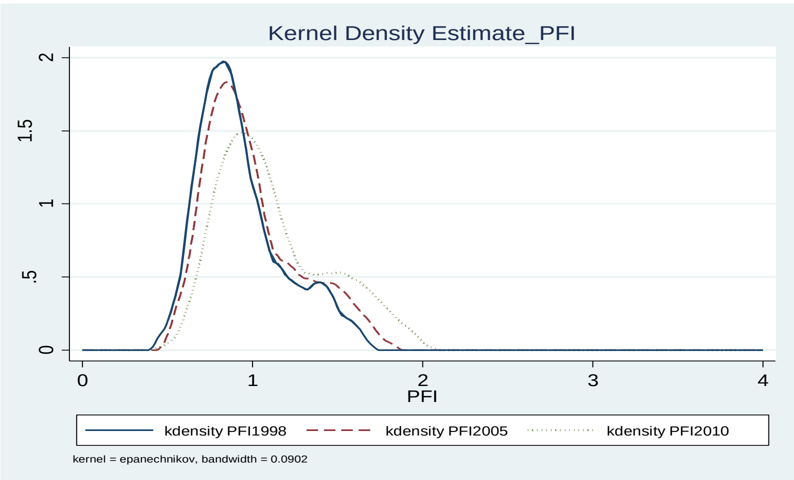


Figure 18 *- Kernel Density Curves of Public Facility Index (PFI) 1998, 2005, 2010*
Source: this figure is drawn according to the data collected and calculated from the Chinese Statistical Yearbook by the software STATA 12.0.

In the contour plot and kernel surface in Figure 19, it showed a club convergence trend in PFI among Chinese provinces and municipalities, since the whole kernel ridge moves counterclockwise with the positive-sloped diagonal and two modes were shown up in both contour plot and kernel surface. In the contour plot, the group with low level of PFI gathered proxy to the national average level 1, and the high level of PFI gathered around 1.5 times of national level. The convergence between two clubs, if estimated from the contour plot, was not strong. This conclusion can also be verified in the kernel surface picture, as the higher peak represents the group with higher level of PFI and verse vice, it showed a more active transformation in the group with higher level of PFI, indicating their relatively larger transformations, and hence the club-convergence of PFI among Chinese regions could be formed if the binary development trend remains.

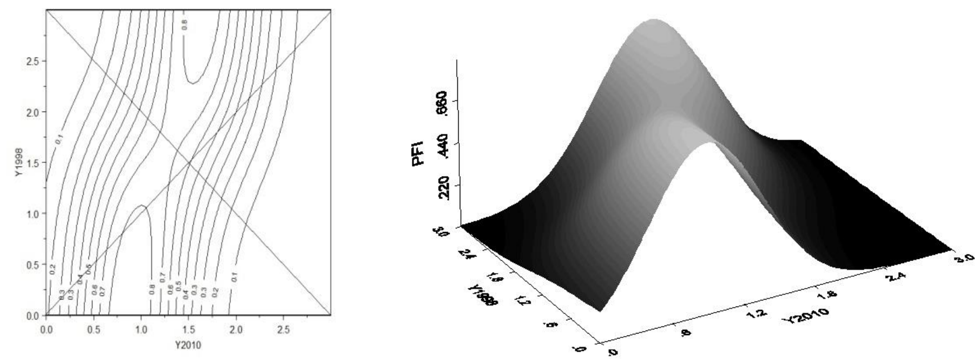


Figure 19 *- Contour Plot and Kernel Surface of Public Facility Index (PFI) 1998-2010*
Source: this figure is got from the dataset collected and calculated from the China Statistical Yearbooks 1999-2011 by the author, the software used are GAUSS and S-PLUS.

6.6. Convergence of CIRD

After the analysis in each sub pillar, I estimate the convergence tendency by the Composite Index of Regional Development (CIRD) of 30 Chinese provinces and municipalities[42] with the aggregation of five dimensions above in this part. This index goes beyond the common measurements such as income per capita, productivity or even HDI, with an endeavor to evaluate Chinese regions in a comprehensive development, involving macroeconomic, science and innovation, environmental sustainability, human capital and public facilities. All of these composed as significant aspects and correlated and interacted with each other to influence the socioeconomic development in different regions in China. Under their interaction, whether there would be different pattern in their convergence trend still remains unclear. In this part, I investigated whether the Chinese regions would converge by their comprehensive socioeconomics in the 13-year development. As always, I firstly started with the tests of *σ-convergence* and *β-convergence*. Table 20 showed the test results in the three sub periods and the whole time interval 1998-2010.

[42] Tibet is excluded because of the missing data in SII and ESI.

Table 20 *- σ-convergence estimation of CIRD*

	π	T_1	T_2	T_3	N	
1998-2010	1.25	0.83	0.74	__	30	*non-convergence*
1998-2000	1.12	1.14	5.53	__	30	*σ-convergence*
2001-2005	1.14	1.25	4.11	__	30	*σ-convergence*
2006-2010	1.29	1.07	0.18	__	30	*non-convergence*

Source: this table is calculated by the author on the basis of sigma-convergence theory.
Note: T_1 test is one-tail test under P=0.05 with a freedom F (N-2, N-2) = (28, 28), the critical value is 1.87; T_2 test is chi-square distribution with freedom F=1, and the critical value is 3.841; T_3 test is standardized normal distribution with P=0.05, and the critical value is 1.65.

In the whole time period 1998-2010 and the third sub period 2006-2010, it is pretty clear that *σ-convergence* of CIRD did not exist since T_1 and T_2 all fall into the range of their critical values, whereas in the first sub period 1998-2000 and the second sub period 2001-2005, there were a tendency of *σ-convergence* since the outranged values of T_2 rejected the null hypothesis of no convergence. On the other hand, *β-convergence* did not occur in the whole time period and all three sub time periods, as the premise of π<1 cannot be satisfied. It also reflects that correlations with initial levels of CIRD and its growth rate is not negative, which can be interpreted as the regions with lower initial socioeconomic levels grow slower than the regions with higher initial socioeconomic levels between 1998 and 2010. Hence, it is solid to conclude that there was no tendency of *β-convergence* in these periods. However, the inexistence of *β-convergence* cannot directly deduce the corollary of inexistence of *σ-convergence* in CIRD. In the first sub period 1998-2000 and the second sub period 2001-2005, the *σ-convergence* occurred without the following of *β-convergence.* This phenomenon was certified Type II error since the T_2 showing larger values than critical values.

To better benchmark the individual provinces or municipalities by their socioeconomic developments in the first year 1998 and the last year 2010, I applied *β-convergence* to picture how the Chinese regions performed with their different starting points in CIRD. Table 21 shows a negative but not significant parameter of β with a small value, which additionally proved that *β-convergence,* did not exist in the comprehensive socioeconomic development among Chinese provinces and municipalities.

Table 21 *- β-convergence estimation of CIRD*

				N. obs = 30
				F(1,28) = 5.94
				Prob>F = 0.0214
				R-squared = 0.2306
Δy (1998-2010)	*coef.*	*st.err*	*P>| t |*	*B*
y1998	*-0.018127*	*0.0074354*	*0.021*	*-0.361433*
cons	*0.0216849*	*0.0030533*	*0*	

(i) Δy is the average annual growth rate of CIRD during 1998-2010

(ii) y1998 is the CIRD at initial time 1998

(iii) cons is the constant

Source: it is calculated by the software STATA 12.0.

In Figure 20, all the provinces and municipalities spread out in their locations and the linear curve fells slightly into the direction of the negative slope. Geographically analyzing, in the east coast, Beijing and Shanghai demonstrated outstanding performances in CIRD, whereas with average growth rates between 1998 and 2010. As elaborated before, this average performance in their growth might be due to that there are not many spaces for them to develop since they are already in a top class of their socioeconomic conditions. Tianjin showed a relatively high quality in its initial socioeconomic conditions, but a low growth rate between 1998 and 2010. Even though it is a municipality locating in the coastal area, it had been overlooked to some degree by the central government. Even though the central government started to put efforts in the integration of Beijing and Tianjin, the effects of policy supporting lagged behind, so here Tianjin showed in a below-average growth rate between 1998 and 2010. Other east coastal regions such as Jiangsu and Zhejiang, demonstrated above-average levels in their initial socioeconomic conditions yet with faster growth rates, but still the difference with the superior developed Beijing and Shanghai cannot be ignored. The inland and peripheral regions all distributed in the left side of the coordinate settings, reflecting their relatively low level of socioeconomic conditions in the initial time; however, their growth rates between 1998 and 2010 are so different that they spread out in the settings. Of course the reasons to form this phenomenon are complicated, but still it is unreasonable to overlook the fact that this is only a two-time-point analysis, so the result might be biased by some extreme values in one single year. This leads me to the stochastic kernel density estimation, in order to evidently describe the dynamic evolvement of CIRD in the 30 Chinese provinces and municipalities over 13 years from 1998 to 2010 and better define the convergence trajectory of CIRD.

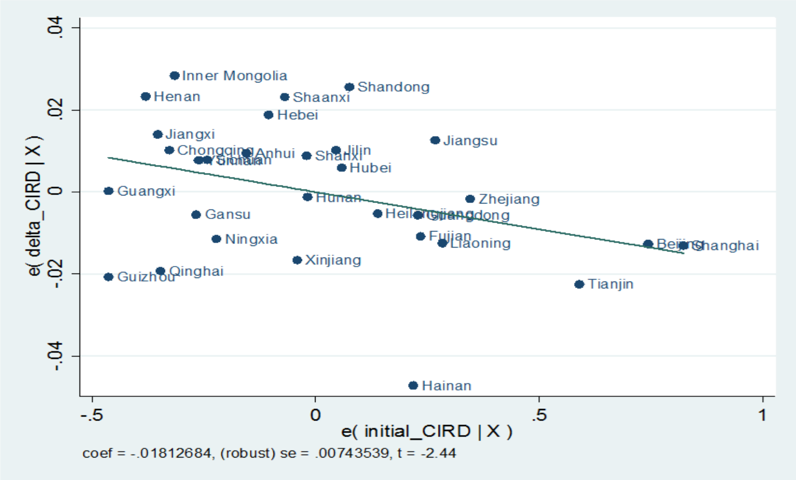


Figure 20 *- The relation between the average growth rate (1998-2010) and the initial level of Composite Index of Regional Development (1998)*

Source: data from author's own calculation and drawn by STATA 12.0

In Figure 21 (a) and (b), three kernel density curves in 1998, 2005 and 2010 were drawn according to the CIRD respectively to provide an initial picture of the dynamic convergence trend among 30 Chinese provinces and municipalities.

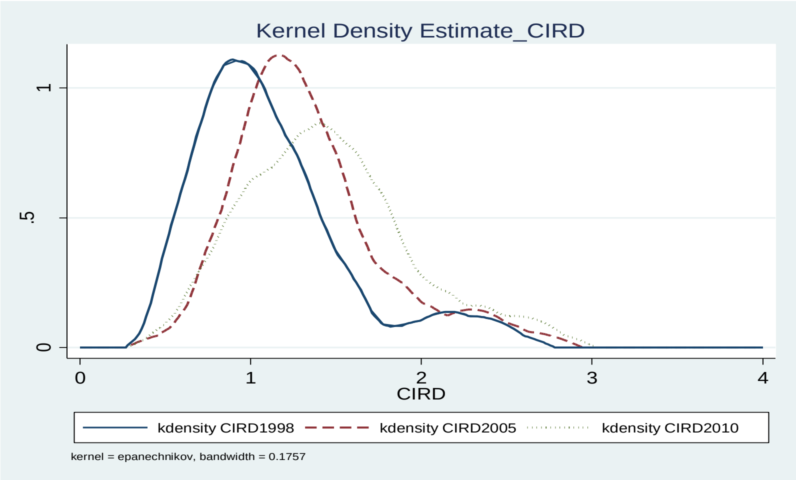


Figure 21a *- Kernel Density Curves of CIRD 1998, 2005, 2010 with bandwidth Epanechnikov*

Source: this figure is drawn according to the data collected and calculated from the Chinese Statistical Yearbook by the software STATA 12.0.

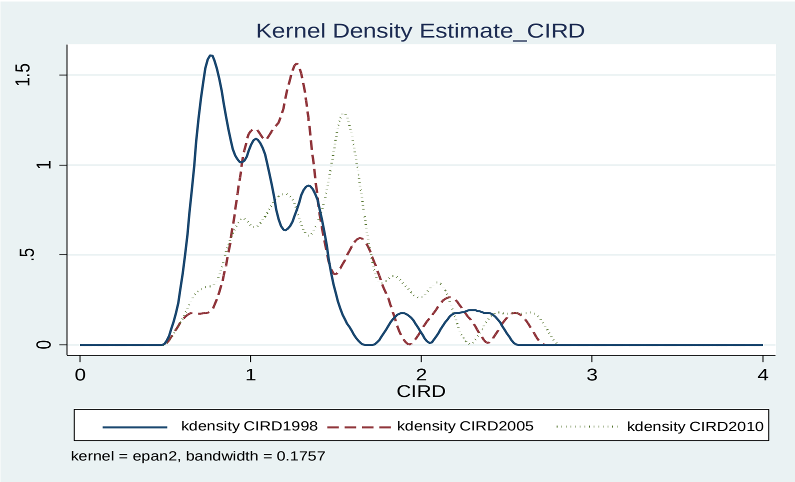


Figure 21b *- Kernel Density Curves of CIRD 1998, 2005, 2010 with bandwidth Epan2*
Source: this figure is drawn according to the data collected and calculated from the Chinese Statistical Yearbook by the software STATA 12.0.

In the first year 1998, two modes were observed in the curve, one is with a higher peak with the value of 0.9 of national average, and another is with a lower peak with the value of 2.1 of national average. It implied a club-convergence existed in this year. In the middle year 2005, the higher mode moved rightly to the value of 1.1, indicating a general improvement of CIRD in the regions inside this group; the lower mode also moved to the right but very slightly, indicating this group in general grew slower than the other group. In the last year 2010, the kernel curve showed multi-modes in case of the application of bandwidth epan2, a trend of club-convergence was observed in this year. The main two peaks were more obvious than the previous curves in Figure 21 (b) and located around 1.1 and 1.6 of the national average level. However, the peaks of 2010 kernel curve in this year got lower, as I discussed before, cannot directly imply a trend of dispersion since the sample size is only 30, since a dramatic change that happened in one or two single regions may reduce the height of the peaks. To summarize, the comprehensive socioeconomic conditions reflected by CIRD in 30 Chinese provinces and municipalities show a club-convergence trend from the kernel density estimation by different bandwidths in 1998, 2005, and 2010. In order to verify this club-convergence pattern in CIRD more precisely, contour plot and kernel surface will be applied over the 13 years.

In the contour plot and kernel surface, a club-convergence trend is observed in the CIRD among 30 Chinese provinces and municipalities. The complete setting of kernel ridges inclines to move counterclockwise along with the positive-sloped diagonal. Meanwhile, two modes are shown up in Figure 22. From the contour plot, one club was composed by the regions with national average level of CIRD, while the other club contains the regions with more than twofold of national average level of CIRD. The club with low level of CIRD shows a tendency to start catching up with the other, as some of its members grow fast, but this tendency was not very obvious since there was still a gap between these two clubs that could be omitted. On the other hand, the club with high level of CIRD showed an inner convergence trend among its members as the mode was proxy parallel to the vertical axis. This inner convergence can also be found out in the club with low level of CIRD. In the kernel surface, it is clear that the club with high level of CIRD transforms more dynamically during these years, as the peak of its mode is higher, which can also be verified by the kernel density curves in Figure 21 (b). To summarize, a club-convergence of comprehensive socioeconomic development occurred during the 13 years (1998-2010), and a tendency that the underdeveloped regions started to catch up with the developed regions was estimated, but not very obviously and clearly. Inner convergences of comprehensive socioeconomic development inside two clubs were also investigated, indicating that the individual regions inside each group tend to develop into a similar level in their comprehensive socioeconomic conditions.

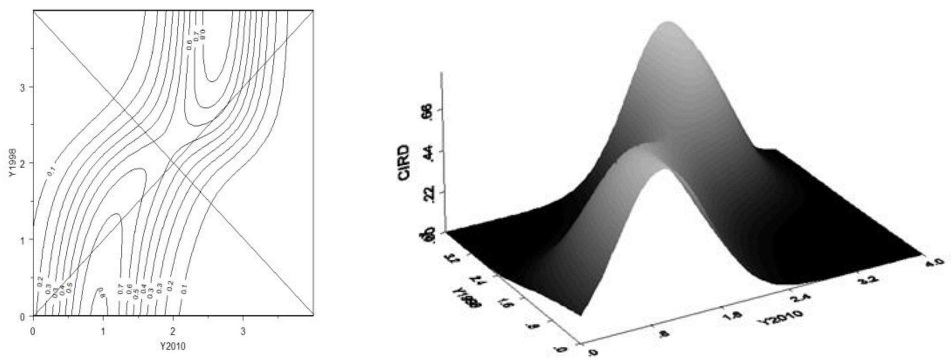


Figure 22 *- Contour Plot of Composite Index of Regional Development (CIRD) 1998-2010*
Source: this figure is got from the dataset collected and calculated from the China Statistical Yearbooks 1999-2011 by the author, the software used are GAUSS and S-PLUS.

7. Conclusion

In this study, a Composite Index of Regional Development (CIRD) is constructed on the basis of theoretical, statistical and empirical methods and techniques, hence to quantify the comprehensive socioeconomic situations in Chinese provinces and municipalities. As the evaluation of the developmental level of a complex economy has to take various aspects into consideration, five sub pillars are set up beneath the CIRD with 25 indicators in general, as Macroeconomic Index (MEI), Science and Innovation Index (SSI), Environmental Sustainability Index (ESI), Human Capital Index (HCI), and Public Facility Index (PFI), in order to describe the sustainable development of regional socioeconomics both respectively and comprehensively. A large dataset is built up with the 25 indicators (as five indicators for each pillar) over a 13-year time interval (1998-2010), which happened to be the periods from the 9th Five-year plan to 11th Five-year plan enacted by Chinese central government. I evaluated the regional sustainable development of Chinese provinces and municipalities indicator by indicator and pillar by pillar. Ranking reports in each dimension and the aggregated CIRD were provided. Regional performances and growth rates based on the aggregated CIRD scores were also used for mapping China. In the dimension of MEI, one principle component is extracted and three groups are classified by the ranking scores from high, middle to low level of MEI. The member size of the middle group is the smallest among the three groups, while the bottom and top groups are in a similarly larger member size, indicating the middle class around national average level in MEI is sort of absent, and hence a tendency of polarization is sort of showing up. In the dimension of SII, there is also only one principle component extracted, however, the three groups are in a balance of their member sizes. The gaps inside the first group are larger than the other two groups, indicating a split-up tendency of SII inside the developed regions. In the dimension of ESI, the first group is almost overlapped with the top quintile, while the second group takes the majority, indicating more regions are gathering together around national average level of ESI. The third group is also in a small member size and totally overlapped with the bottom quintile. Also, there is a vertical split-up in the inland regions from north to central to south in ESI. In the dimension of HCI, the low level group takes the majority of regions while the high and middle groups are in a similarly small size of membership. It implies that a large proportion of Chinese regions are still in a relatively low level in HCI, below the national average level. In the dimension of PFI, I find a dramatic gap inside the high level group, whereas inside the middle and underdeveloped groups, the gaps are not so large. The aggregated scores under two different methods bring two different ranking reports. The one with equal-weighting method shows the majority of Chinese provinces and municipalities concentrated in the low level group, while the high level group is in the smallest size. It means that there is still a great portion of regions in China staying in a low level of comprehensive socioeconomic condition, below the national average level of CIRD, while the developed regions above national average level are a small portion. In the aggregate ranking report according to the parametric method of integral PCA result, it shows a different distribution of the three groups. The majority regions allocated in the middle level group, whereas the smallest member size is in the low level group. It indicates the major part of Chinese regions is in a level around the national average, and the number of underdeveloped regions is reducing, which might be interpreted as a tendency of convergence. As I concerned, the latter is more suitable to describe the current reality in China.

Geographically, the east coastal regions have a good performance compared with other regions in all of these classifications. The only slight exception is observed in the ESI dimension, where some of the east coastal regions located in the middle level group. The inland regions and peripheral regions, however, are complicated to be summarized a routine of their classifications. It seems that a split-up is emerging in the inland and peripheral regions. The central and southern inland and peripheral regions usually performed better than the northern inland and peripheral regions. Also, the inland regions adjacent to the east coast usually show a better performance than the isolated regions because of the spill-over effects from the developed east coast. The ranking of changes of the Chinese provinces and municipalities is more complex to interpret. Some of the central inland regions such as Hunan and Hubei display a distinguished progress in the

CIRD dimensions, implying the rising-up of central area. Nevertheless, some of the east coastal regions were in a relatively slow growth. As I underlined before, it is reasonable to ascribe their relatively low rates of growth to their initial superior levels in these east coastal regions, so there is not enough upside potential for them.

Cluster analysis provides new classification of Chinese provinces and municipalities. Mapping regions with the clustering results represented a geographical-oriented feature. Dispersions are observed horizontally between east coast and inland regions, and vertically among northern, central and southern inland regions; peripheral regions, nevertheless, showed more homogenous even with their scattered geographical locations. Cluster 1 demonstrates as dramatically different with other clusters in its initial conditions of MEI, SII, HCI, PFI and CIRD (1998), whereas the gap of PFI with Cluster 2 diminished in 2010. The gaps among Cluster 2, Cluster 3 and Cluster 4 are also large in MEI, SII, HCI, PFI and CIRD in their initial conditions, however, the difference of HCI between Cluster 3 and Cluster 4 reduced in 2010. Environmental sustainability (ESI) is the only pillar with relatively smaller gaps among all the clusters, whereas does not imply the advanced level in this pillar but contrarily the integral deterioration.

Referring to the dynamic trajectory of socioeconomic development of the Chinese regions, club-convergence is concluded as the most common trend that has been observed in the each separate dimension and the integral CIRD. Besides it did not show up in the initial time in the environmental sustainability, the macroeconomic conditions, science and innovation, human capital development, public facilities and the comprehensive socioeconomic development containing the five sub pillars/dimensions above all demonstrate club convergence. The difference is mainly concentrated on that how the clubs performed during the time interval 1998-2010 in different dimension. In the MEI and HCI, there is a potential tendency of multi club convergence observed in the contour plot. However, in MEI, the clubs are inclined to converge with each other, while HCI shows no such tendency of convergence between the clubs, which means that the clubs almost remained persistently in their distributions in HCI. In the SII, the two clubs behave a tendency to converge, while in the PFI, this convergence inclination between clubs is not very strong, and yet no persistence is shown between the two clubs. In the aggregated

CIRD, there is a convergence tendency between the two clubs; meanwhile, it is also possible to assume the multi club convergence beneath the dual-mode club convergence if the bandwidth is changed and hence more transformation activities can be observed.

In brief, the regional coordinated development cannot rely on sole dimension, multi aspects should be taken into consideration as their integrative impact on the regional socioeconomic development. Due to the different levels and developmental trajectories of regional socioeconomic conditions, it is significant to make targeted policies for different groups to narrow down the regional inequalities and promote the whole economy. The east coast should be fixed as a high-tech belt; in the meantime, it is necessary to transform some of its economic functions and labor-intense industries into the contiguous inland or even central inland regions, hence to further magnify the positive radiation. The development of central and western area can, not only benefit their local booming, but also open new market and provide better resource support for the further development of the east in return. Special supporting projects and financial investments should be established for the underdeveloped peripheral provinces and municipalities since their intrinsic motivations for local prosperity are limited. Also, more efforts should be taken to rural development, following-up works of the tax-free policy and relevant fostering policies should be provided to consolidate the improvements, additionally, particular attentions should be paid to how to properly manage the process of urbanization.

Referring to the further studies and improvements, a sub-dimensional classification of urban and rural disparity can be involved in the CIRD, since great gaps exist between urban and rural area in China in different aspects, for example, in public facilities and services such as education and medical resources, or in human capital such as educational level and employment structure. Another angle can be focused on an analysis of inferior level of administrative district such as prefectural level, and then a large sample size can be observed and further investigated with more potential studies. However, the availability of reliable and consistent data is the main challenge required to be conquered first to pursue the further studies, and hence necessary alterations in each pillar may have to be considered in order to adjust CIRD to adapt for new analysis.

8. References

Alanen, A., Huovari, J. and Kangasharju, A. (2000) “Regional competitiveness In Finland”, ERSA 40th European Congress 29th August - 1st September, 2000.

Alanen, A., Huovari, J. and Kangasharju, A. (2001) Constructing an index for regional Competitiveness, Helsinki: Peelervo Economic Research Institute – Working paper No 44. http://www.ptt.fi/dokumentit/tp44_10080608.pdf

Anand, S. and Sen, A. (1994) “Human development index: methodology and measurement”, Human Development Report Office, Occasional Papers.

Annoni, P. and Kozovska, K. (2010) “EU Regional Competitiveness Index 2010”, European Commission, JRC-IPSC 58169.

Auffhammer, M., Chong, H. and Wolff, H. (2009) “Human development index: are developing countries misclassified?”, selected paper prepared for presentation at the Agricultural & Applied Economics Association 2009 AAEA&ACCI Joint Annual Meeting, Milwaukee, Wisconsin, July 26-26, 2009.

Barone, B. and Benni, F. (2012) Convergence in China: Rural-Urban and Spatial Dynamics, LAP LAMBERT Academic Publishing, Oct. 31, 2012, ISBN: 3659276774

Balogh, V., Bartha, A., Kovács, K., Nábrádi, A., Pető, K. and Szabó, E. (2007) Efficiency indicators in different dimension, APSTRACT: Agroinform Publishing House, Budapest.

Berger, T. and Bristow, G. (2008) “Benchmarking regional performance: a critical reflection on indices of competitiveness”, paper for the Regional Studies Association (RSA) Annual Conference, Regions: The Dilemmas of Integration and Competition, May 27th-29th 2008, University of Economics Prague, Czech Republic.

Betrand Le Gallic. (2002). “Fisheries Sustainability Indicators: The OECD Experience.” OECD. Joint workshop on “Tools for measuring (integrated) Fisheries Policy aiming at sustainable ecosystem”. October 28-29, 2002, Brussels (Belgium)

Benni,F., Brasili,C., Costantini,B., Sandri,E. (2006), Convergence in the European Union Regions: The Impact of Structural Funds, presented at International Workshop on Spatial Econometrics and Statistics, rome, 25-27 May 2006

Bin P., Fanfani R., Brasili C. (2012) “Regional differences and dynamic changes in rural China: the study of 1996 and 2006 National Agricultural Census”. Asian Journal of Agriculture and Rural Development, 2: 260-270.

Bosker, M., Brakman, S., Garretsen, H. and Schramm, M. (2007) “Adding geography to the new economic geography” CESIFO working paper no. 2038.

Bossel, H. (1999) “Indicators for Sustainable Development: Theory, Method, Applications”, a

Report to the Balaton Group, International institute for sustainable development.

Brandt, L. and Holz, C. A. (2006) Spatial price differences in China: estimates and implications, University of Chicago.

Buisman, E., Frost, H.,etc., (2008) "Review on economic indicators", CEVIS.

Cahill, M.B., and N. Sanchez (1998), "Using Principal Components to Produce an Economic and Social Development Index: An Application to Latin America and the US", mimeo, College of the Holy Cross-USA.

Cai F., Wang D. and Du Y. (2002). "Regional disparity and Economic Growth in China: the Impact of Labor Market Distortions." China Economic Review, 13, 197-212.

Cappelen, A. (2004) "Convergence, Divergence and the Kuznets-curve", in Reinert, E. S. (ed.): Globalization, economic development and inequality, Edward Elgar, Cheltenham, UK, 2004, 309-325.

Caree, M. and Klomp, L. (1997) "Testing the convergence hypothesis: a comment", The Review of Economics and Statistics, Vol. 79, No. 4, November 1997.

Carmeci, G. and Mauro, L. (2003) "Imperfect labor market and convergence: theory and evidence for some OECD countries", Journal of Policy Modeling, 2003, vol. 25, issue 8, pages 837-856.

Chen, B. and Woo, Y.P. (2008) A Composite Index of Economic Integration in the Asia-Pacific Region, research report, http://www.asiapacific.ca/sites/default/files/filefield/PECCIntegrationIndex.pdf

Chen, H. (2009) Convergence, productivity and industrial growth in china during the reform era, thesis submitted to the University of Birmingham, Department of Economics February 2009.

China Human Development Report (2005) China Human Development Report

Cowell, F. A. (2009) Measuring Inequality, (third edition), Oxford University Press, Oxford.

Dal Bianco, S. (2007) "Alternative hypotheses of cross-country convergence. A non-parametric analysis of manufacturing sectors", Università degli studi di Pavia, Quaderni di Dipartimento # 196 (04-07).

De la Fuente, A. (2000) "Convergence across countries and regions: theory and empirics", EIB papers, Volume 5, No. 2/2000.

Di Liberto A. (2005) "Convergence and divergence in neoclassical growth models with human capital", Università di Cagliari and CRENoS, working paper.

Dreher, A. (2006), "Does Globalization Affect Growth? Evidence from a New Index of Globalization, Applied Economics 38, 1091-1110.

DEFRA (2004). "Regional quality of life counts-2003. Regional versions of the headline indicators of sustainable development", 4 ed., Department for Environment, Food and

Rural Affairs, London

Egger, P. and Pfaffermayr, M. (2009) “On Testing Conditional Sigma –Convergence”, Oxford Bulletin of Economics and Statistics 71(4), 2009, 453-473.

EUROCHAMBERS (2007). Regional Competitiveness Atlas. Brussels: EUROCHAMBERS.

European Commission (2009a) Progress towards the Lisbon Objectives in Education and Training: Indicators and benchmarks 2009. Brussels: Commission staff Working Document.

Fan C. and Sun M. (2008). “Regional Inequality in China, 1978-2006.” Eurasian Geography and Economics, 49, 1-18.

Fan, S., Kanbur, R. and Zhang, X. (2009) “China’s Regional Disparities: Experience and Policy”, Review of Development Finance 1:47-56, 2011.

Fan, S., Kanbur, R. and Zhang, X. (2009b) Introduction to Fan, S., Kanbur, R., Zhang, X., Regional Inequality in China: Trends, Explanations and Policy Responses, Routledge, 2009.

Fanfani R. and Brasili C. (2003) “Regional Differences in Chinese Agriculture: Results from the 1997 First National Agricultural Census.” Journal of Peasant Studies, 30, 18-44.

Fleisher, B., Li, H. and Zhao M. Q. (2010) “Human capital, economic growth, and regional inequality in China”, Journal of Development Economics 92 (2010) 215–231.

Ghatak, S. and Li, H. (2006) “Economic Growth and Convergence in China’s Provinces: Theory and Evidence”, International journal of economic research, 3(1), pp. 1-15. ISSN (print) 0972-9380.

Gundlach, E. (1997) “Regional convergence of output per worker in China : a neoclassical interpretation”, Asian Economic Journal, ISSN 1351-3958, Vol. 11, Iss. 4, pp. 423-442.

Heshmati A. (2006), “Measurement of a Multidimensional Index of Globalization”, Global Economy Journal 6 (2), Article 1.

Heshmati, A., and J.E. Oh, (2005) Alternative Composite Lisbon Development Strategy Indices, IZA Discussion Paper No.1734.

Hotelling, H. (1933), "Analysis of a complex Statistical Variables into Principal Components", Journal of Educational Psychology, 24, 417-41.

Huggins, R. and Izushi, H. (2008) UK Competitiveness Index 2008. University of Wales Institute, Cardiff – UWIC: Centre for International Competitiveness – Cardiff School of Management. http://www.cforic.org/downloads.php

Huggins, R., Davies, W. (2006) European Competitiveness Index 2006-07. University of Wales Institute, Cardiff – UWIC: Robert Huggins Associates Ltd. http://www.cforic.org/downloads.php

Huggins, R. (2003) “Creating a UK Competitiveness Index: regional and local benchmarking”, Regional Studies, 37(1), pp. 89-96.

Huovari, J., Kangasharju, A., Alanen, A. (2001) "Constructing an index for regional competitiveness." Helsinki: Peelervo Economic Research Institute – Working paper No 44. http://www.ptt.fi/dokumentit/tp44_10080608.pdf

IMD (2008) World Competitiveness Yearbook 2008, Institute for Management Development, Lausanne.

Islam, N. and Sakamoto, H. (2008) "Convergence across Chinese provinces: an analysis using Markov transition matrix", China Economic Review, 19 (2008) 66–79.

Kanbur R.and Zhang X. (2005). "Fifty Years of Regional Inequality in China: a Journey Through Central Planning, Reform, and Openness." Review of Development Economics, 9, 87-106.

Kelley, A. C. (1991) "The Human Development Index: "Handle with Care", Population and Development Review, Vol. 17, No. 2 (Jun., 1991), pp. 315-324.

Kitson, M., Martin, R. and Tyler, P. (2006) Regional Competitiveness. London: Routledge.

Kuusk, A., Paas, T., Schlitte, F. and Võrk, A. (2007) "Econometric analysis of income convergence in selected EU countries and their nuts 3 level regions", The University of Tartu, faculty of Economics and Business Administration, working waper No. 60-2007.

Kuznets, S. (1955) "Economic growth and income inequality", American Economic Review, 49: 1-28.

Lal, S. V. and Yilmaz, S. (2001) "Regional economic convergence: do policy instruments make a difference?" The Annals of Regional Science (2001), 35: 153-166.

Lei, C. K. and Yao, S. (2008) Economic Convergence in greater China. Mainland China, Hong Kong, Macau and Taiwan, Routledge

Lichtenberg, F. R. (1994) "Testing the convergence hypothesis", The Review of Economics and Statistics, Vol. 76, No. 3 (Aug., 1994), pp. 576-579.

Lin Y. J., Cai F., Li Z. (2008): The China Miracle: Development Strategy and Economic Reform. The Chinese University Press, Hong Kong; ISBN 962-201-985-4

Maasoumi, E. and Wang, L. (2007) "Economic Reform, Growth and Convergence in China", Econometrics Journal, Volume 10, Number 1, pp. 1-25.

Mankiw, N. G., Romer, D. and Weil, D. N. (1992) "A contribution to the empirics of economic growth", The Quarterly Journal of Economics, Vol. 107, No. 2. (May, 1992), pp. 407-437.

Mazzocchi, M. (2008). Statistics of Marketing and Consumer Research. London: SAGE.

Meyer-Stamer, J. (2008) "Systematic Competitiveness and Local Economic Development." In Shamin Bodhanya (ed.) Large Scale Systemic Change: Theories, Modelling and Practices. http://www.meyer-stamer.de/2008/Systemic+LED_SouthAfrica.pdf

Michelak, J. and Zanekow, N. (2012) Construction and application of the Rural Development

Index to analysis of rural regions. JRC Scientific and Policy Reports, European Union, printed in Spain.

Mishra, S.K. (2007) A Comparative Study of Various Inclusive Indices and the index Constructed by the Principal Components Analysis, MPRA Paper No.3377

Monasterolo, I. and Pagliacci, F. (2011) Mapping changes on agricultural and rural areas: an ex-post evaluation of the EU membership for Hungary, conference paper, EAAE, http://ageconsearch.umn.edu/bitstream/98988/2/monasterolopagliacci.pdf

NBSC. (1999-2011). China Statistical Yearbook 1999-2011. Beijing: State Statistical Bureau People's Republic of China. http://www.stats.gov.cn/tjsj/ndsj/1997/indexeh.htm

Noll, Heinz-Herbert. (2002) "Social indicators and quality of life research: Background, achievements and current trends," In: Genov, N. B. (eds.) : Advances in sociological knowledge over half a century. Internat. Soc. Sc. Council ISSC: Paris.

OECD. (2009) Rural Policy Review: China 2009. Paris: OECD.

Osberg, L. and Sharpe, A. (2002) "An Index of Economic Well-Being for Selected OECD Countries", Review of Income and Wealth, Blackwell Publishing, 48(3), pp. 291-316.

Peneder, M. (1997) Creating a coherent design for cluster analysis and related policies, conference paper, OECD Workshop on Cluster Analysis and Cluster Based Policies, Amsterdam, 10-11 October 1997

Perloff, J. M. and Wu, X. (2004) "China's income distribution and inequality", Paper provided by Econometric Society in its series Econometric Society 2004 North American Summer Meetings with number 316.

Quah, D. T. (1995) "Empirics for economic growth and convergence", Centre for Economic Performance, discussion paper no. 253.

Quah D. T. (1995b) "Regional convergence clusters across Europe", Centre for Economic Performance, discussion paper no. 274.

Quah, D. and Thomas J. S. (1993) A dynamic index model for large cross sections," In Business Cycles, Indicators, and Forecasting, ed. James Stock and Mark Watson, vol. 28 (Chicago IL: University of Chicago Press and NBER) chapter 7, pp. 285-306

Quah, D. (1992) "The relative importance of permanent and transitory components: Identication and some theoretical bounds," Econometrica, 60(1), 107-118, January

Quah, D. T., (1997) "Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs," Centre for Economic Performance, Discussion Papers 1586.

Sadesky, G. S., (2003) Cluster analysis and its application in standard settings, conference paper, annual meeting of Canadian society for the study of higher education, http://www.asiapacific.ca/sites/default/files/filefield/PECCIntegrationIndex.pdf

Sala-i-Martin, X. (1994) "Cross-sectional regressions and the empirics of economic growth," European Economic Review 38(3/4), 739-747, April

Sala-i-Martin X. (1996) “The Classical Approach to Convergence Analysis”, The Economic Journal, 106, pp. 1019-1036

Wan, G., Lu, M. and Chen, Z. (2007) “Globalization and regional income inequality: empirical evidence from within China”, Review of Income and Wealth, Series 53, Number 1, March 2007.

Weeks, M. and Yao, J. Y. (2002) “Provincial conditional income convergence in China, 1953-1997: a panel data approach”, Faculty of Economics and Politics, University of Cambridge.

World Health Organization (2005) “China: health, poverty and economic development”, office of the WHO representative in China and Social Development of China State Council Development Research Center, December 2005, Beijing.

Xu X. (2002) “Have the Chinese Provinces Become Integrated under Reform?” China Economic Review, 13, 116-133.

Zhang, T. and Zou, H. (1998). Fiscal Decentralization, Public Spending, and Economic Growth in China. Journal of Public Economics 67:221–40.

Zhao Y. (1999) “Labor Migration and Earnings Differences: The Case of Rural China.” Economic Development and Cultural Change, 47, 767-782.

Zhu, N. (2002). “The Impacts of Income Gaps on Migration Decisions in China.” China Economic Review, 13, 213-220.

Zhou, H. and Zou, W. (2007) “Classification of growth clubs and convergence: evidence from panel data in China, 1981–2004”, China &World Economy, 91 – 106, Vol. 15, No. 5, 2007.

ACKNOWLEDGEMENT

First of all, I would like to express my sincere appreciation to my respectable and responsible supervisor, Prof. Roberto Fanfani. His enlightening guidance greatly helps me to conquer the obstacles in pursuing my doctoral degree. His rigorous working style gives me profound and meaningful influences.

High tribute shall also be paid to Prof. Cristina Brasili, for her unselfish helps and amicable supports in my PhD period. I benefit a lot from her professional instruction and substantial expertise.

A special acknowledgement shall be made for my colleague and my best Italian friend Dr. Barbara Barone, who really enriches my life abroad. Not only does she provide massive selfless helps in my research works, her optimistic life attitude and honest open mind also profoundly influence me.

I also would like to extend my thanks to every professor providing me guidance and helps and every colleague working with me in the past years, I have learned a lot from them.

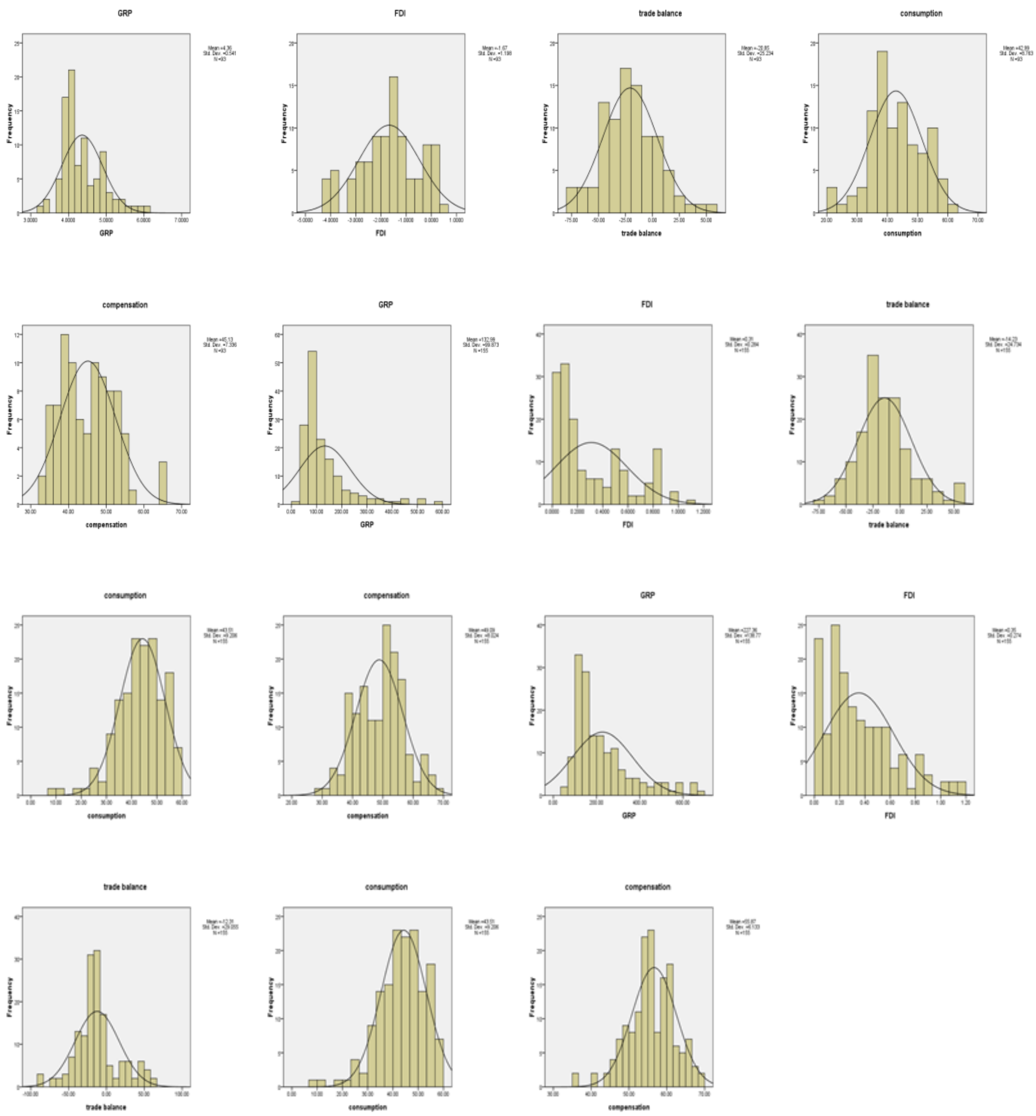
Many thanks to my Chinese friends in Bologna, their friendships warm my heart and color my life.

Special thanks to my boyfriend, Dr. Xinwu Cui, his companionship gives me a sense of home when staying abroad.

I also would like to thank Erasmus Mundus program and DAAD program, which provided me valuable opportunities to study abroad.

Last but not least, I would like to deliver my deeply heartfelt gratitude to my beloved parents, my life mentors. Their graciousness for me is more than I can express. I love them so much, beyond words!

Appendix 1.1 *Data Analysis of Macroeconomic Index (MEI) 1998-2010*



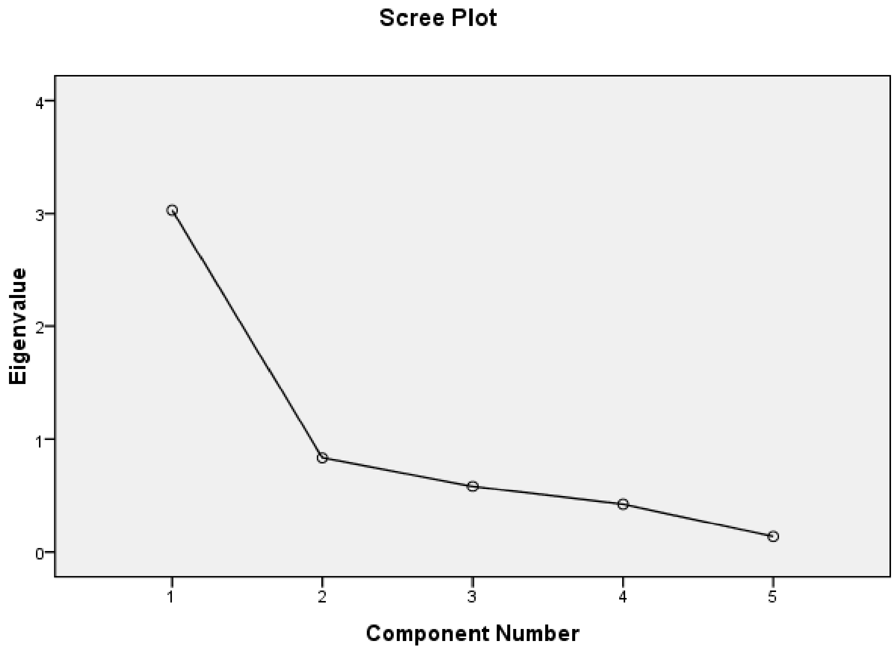
***Appendix 1.2-i** PCA Results of Macroeconomic Index (MEI) 1998-2000*

Component Matrix[a]

	Component
	1
GRP	.950
FDI	.748
trade balance	.515
consumption	.822
compensation	.792

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.029	60.589	60.589	3.029	60.589	60.589
2	.833	16.656	77.245			
3	.580	11.606	88.851			
4	.422	8.433	97.284			
5	.136	2.716	100.000			



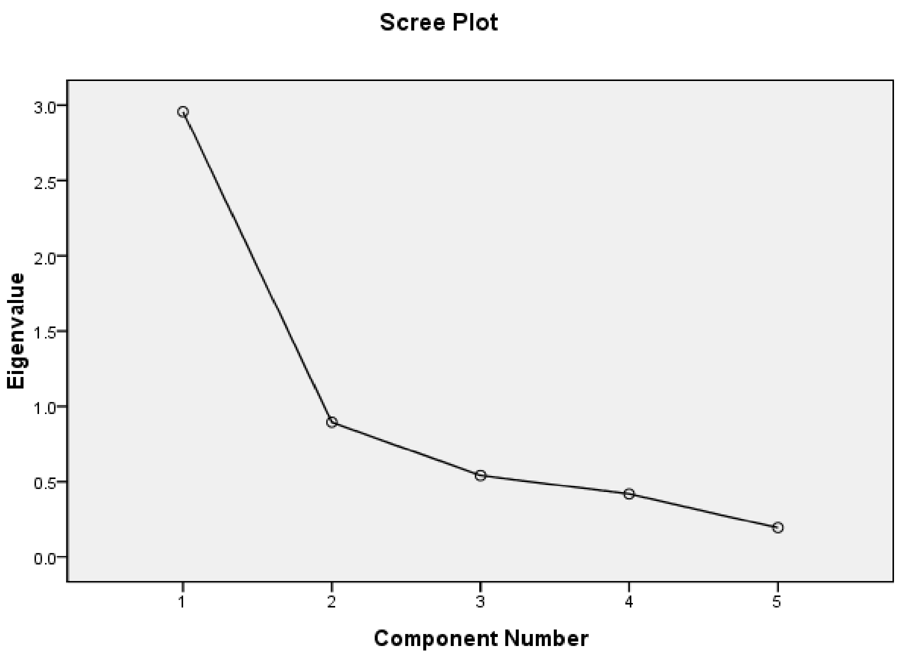
Appendix 1.2-ii *PCA Results of Macroeconomic Index (MEI) 2001-2005*

Component Matrix[a]

	Component
	1
GRP	.926
FDI	.799
trade balance	.470
consumption	.818
compensation	.755

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.957	59.137	59.137	2.957	59.137	59.137
2	.893	17.858	76.995			
3	.541	10.819	87.814			
4	.417	8.337	96.151			
5	.192	3.849	100.000			



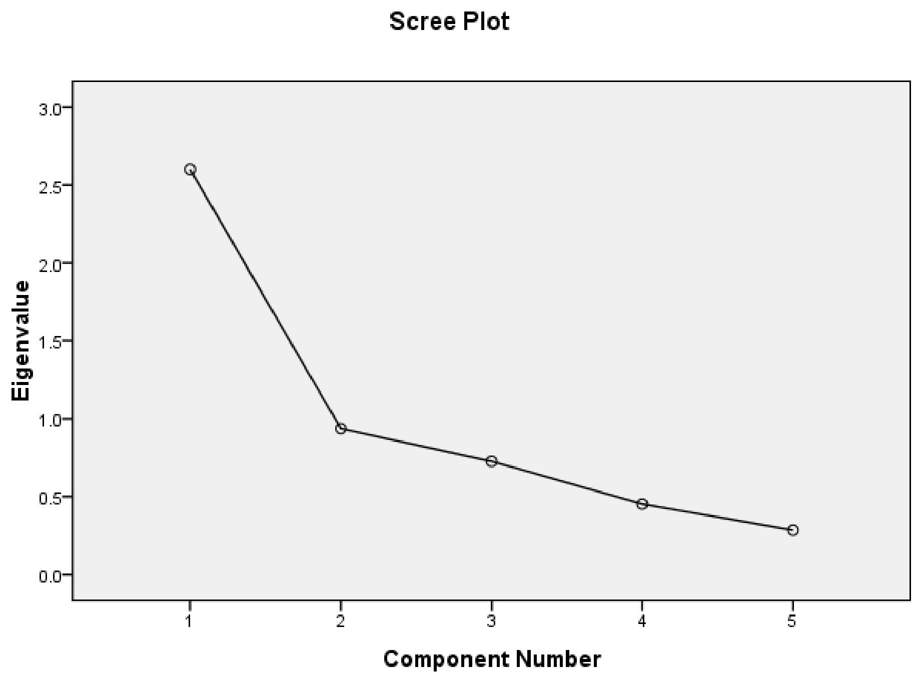
Appendix 1.2-iii *PCA Results of Macroeconomic Index (MEI) 2006-2010*

Component Matrix[a]

	Component
	1
GRP	.843
FDI	.800
trade balance	.363
consumption	.860
compensation	.614

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.601	52.018	52.018	2.601	52.018	52.018
2	.936	18.730	70.747			
3	.726	14.526	85.273			
4	.452	9.041	94.314			
5	.284	5.686	100.000			



Appendix 1.3 *Ranking Report of Macroeconomic Index (MEI) 1998-2010*

Province_1998	*MEI*	*Province_1999*	*MEI*	*Province_2000*	*MEI*
Shanghai	100	Shanghai	100	Shanghai	100
Beijing	87	Beijing	85	Beijing	86
Tianjin	65	Tianjin	63	Tianjin	68
Guangdong	45	Guangdong	42	Guangdong	44
Zhejiang	44	Zhejiang	42	Zhejiang	42
Jiangsu	44	Jiangsu	40	Jiangsu	42
Fujian	42	Fujian	38	Liaoning	37
Heilongjiang	35	Liaoning	33	Fujian	37
Liaoning	35	Heilongjiang	29	Shandong	30
Shandong	33	Shandong	29	Heilongjiang	30
Hainan	25	Hebei	19	Xinjiang	19
Xinjiang	23	Hubei	18	Inner Mongolia	19
Jilin	21	Jilin	18	Hebei	19
Hebei	20	Inner Mongolia	16	Jilin	17
Hubei	19	Chongqing	15	Hubei	17
Tibet	15	Hainan	15	Hainan	15
Chongqing	15	Xinjiang	14	Chongqing	9
Inner Mongolia	13	Sichuan	11	Sichuan	7
Sichuan	9	Shanxi	7	Henan	7
Yunnan	9	Henan	7	Shanxi	5
Anhui	8	Tibet	7	Jiangxi	5
Shanxi	8	Shaanxi	6	Shaanxi	4
Henan	8	Anhui	6	Hunan	4
Shaanxi	6	Yunnan	5	Anhui	4
Gansu	4	Hunan	3	Yunnan	2
Jiangxi	2	Jiangxi	1	Gansu	2
Hunan	2	Ningxia	1	Guangxi	1
Guangxi	1	Guizhou	0	Guizhou	1
Ningxia	1	Guangxi	0	Ningxia	0
Qinghai	0	Qinghai	0	Qinghai	0
Guizhou	0	Gansu	0	Tibet	0

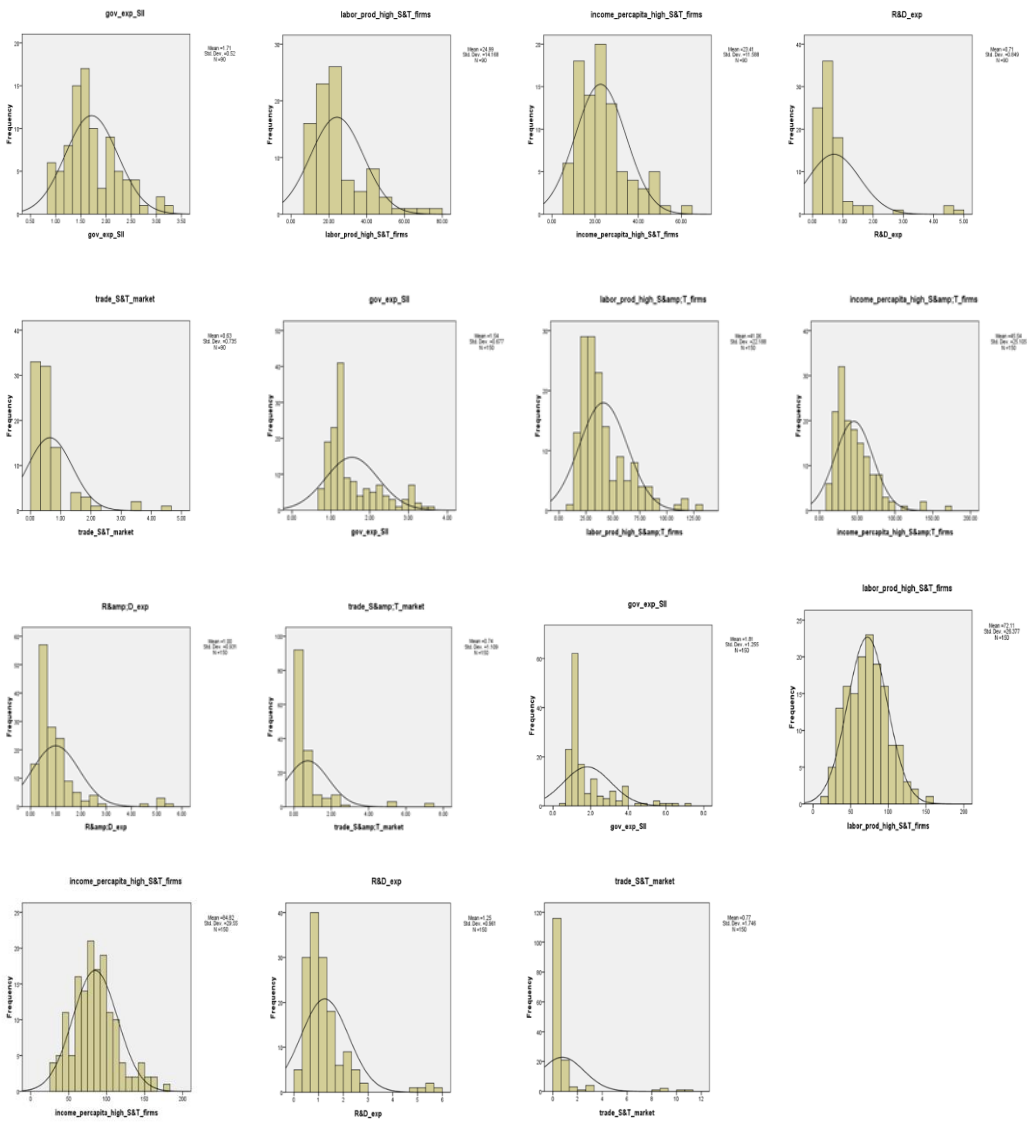
(Continued i)

Province_2001	*MEI*	*Province_2002*	*MEI*	*Province_2003*	*MEI*	*Province_2004*	*MEI*	*Province_2005*	*MEI*
Shanghai	100	Shanghai	100	Shanghai	100	Shanghai	100	Shanghai	100
Beijing	89	Beijing	87	Beijing	87	Beijing	93	Beijing	90
Tianjin	70	Tianjin	70	Tianjin	74	Tianjin	78	Tianjin	77
Guangdong	44	Zhejiang	45	Zhejiang	47	Zhejiang	51	Zhejiang	49
Zhejiang	42	Jiangsu	41	Jiangsu	42	Jiangsu	46	Jiangsu	47
Jiangsu	41	Guangdong	41	Guangdong	41	Guangdong	41	Guangdong	45
Liaoning	37	Fujian	35	Liaoning	32	Shandong	34	Shandong	36
Fujian	36	Liaoning	34	Fujian	32	Liaoning	33	Inner Mongolia	32
Shandong	29	Shandong	29	Shandong	30	Fujian	32	Liaoning	31
Heilongjiang	28	Heilongjiang	27	Heilongjiang	22	Heilongjiang	24	Fujian	28
Xinjiang	25	Xinjiang	19	Hebei	18	Hebei	20	Jilin	22
Hubei	20	Hebei	18	Xinjiang	16	Inner Mongolia	20	Heilongjiang	20
Hebei	19	Hainan	18	Hainan	16	Jilin	20	Hebei	20
Hainan	19	Hubei	17	Jilin	15	Hainan	17	Shanxi	15
Inner Mongolia	18	Inner Mongolia	16	Inner Mongolia	15	Xinjiang	16	Xinjiang	14
Jilin	15	Jilin	15	Hubei	14	Hubei	14	Hainan	14
Sichuan	8	Chongqing	7	Shanxi	6	Chongqing	11	Hubei	13
Henan	7	Anhui	5	Jiangxi	5	Shanxi	7	Henan	10
Chongqing	7	Henan	5	Chongqing	5	Henan	7	Chongqing	8
Shaanxi	7	Guizhou	5	Anhui	5	Jiangxi	7	Shaanxi	8
Hunan	4	Shaanxi	5	Hunan	5	Hunan	6	Jiangxi	6
Ningxia	4	Sichuan	4	Henan	5	Anhui	5	Gansu	6
Anhui	3	Jiangxi	4	Shaanxi	5	Shaanxi	4	Hunan	6
Yunnan	3	Hunan	4	Sichuan	3	Sichuan	4	Anhui	4
Guangxi	3	Ningxia	4	Gansu	3	Ningxia	4	Sichuan	4
Qinghai	2	Shanxi	3	Guizhou	3	Guizhou	2	Guizhou	4
Shanxi	2	Tibet	2	Ningxia	2	Tibet	2	Guangxi	3
Tibet	1	Yunnan	2	Guangxi	1	Qinghai	1	Qinghai	2
Jiangxi	1	Gansu	1	Tibet	1	Yunnan	1	Ningxia	2
Gansu	0	Guangxi	1	Qinghai	1	Gansu	1	Tibet	0
Guizhou	0	Qinghai	0	Yunnan	0	Guangxi	0	Yunnan	0

(Continued ii)

Province_2006	*MEI*	*Province_2007*	*MEI*	*Province_2008*	*MEI*	*Province_2009*	*MEI*	*Province_2010*	*MEI*
Shanghai	100	Shanghai	100	Shanghai	100	Shanghai	100	Beijing	100
Beijing	92	Beijing	85	Beijing	88	Beijing	89	Shanghai	96
Tianjin	75	Tianjin	70	Tianjin	73	Tianjin	77	Tianjin	94
Zhejiang	50	Zhejiang	48	Zhejiang	47	Jiangsu	48	Jiangsu	63
Jiangsu	47	Jiangsu	46	Jiangsu	46	Zhejiang	46	Inner Mongolia	59
Guangdong	44	Guangdong	42	Guangdong	41	Inner Mongolia	44	Zhejiang	55
Shandong	36	Shandong	34	Inner Mongolia	36	Guangdong	40	Liaoning	44
Inner Mongolia	30	Inner Mongolia	31	Shandong	36	Shandong	35	Guangdong	43
Liaoning	29	Liaoning	28	Liaoning	32	Liaoning	33	Shandong	41
Fujian	28	Fujian	27	Fujian	28	Fujian	28	Fujian	36
Hebei	20	Jilin	19	Jilin	21	Jilin	24	Jilin	31
Jilin	19	Hebei	17	Hebei	20	Hebei	18	Shaanxi	20
Heilongjiang	16	Shanxi	12	Shanxi	14	Shanxi	16	Shanxi	19
Hubei	13	Heilongjiang	11	Hainan	13	Shaanxi	15	Hebei	19
Shanxi	13	Hainan	11	Heilongjiang	13	Hainan	13	Hubei	18
Hainan	12	Hubei	11	Hubei	13	Chongqing	13	Heilongjiang	17
Xinjiang	10	Henan	8	Henan	12	Hubei	13	Chongqing	17
Henan	10	Shaanxi	6	Shaanxi	11	Heilongjiang	12	Hainan	15
Shaanxi	8	Xinjiang	5	Chongqing	8	Henan	11	Ningxia	13
Chongqing	7	Chongqing	4	Xinjiang	7	Hunan	9	Hunan	10
Jiangxi	4	Gansu	4	Qinghai	6	Qinghai	8	Henan	10
Hunan	4	Yunnan	3	Gansu	6	Ningxia	8	Xinjiang	9
Gansu	4	Qinghai	3	Hunan	5	Jiangxi	7	Qinghai	8
Sichuan	4	Hunan	2	Jiangxi	5	Xinjiang	5	Jiangxi	6
Anhui	3	Jiangxi	2	Sichuan	4	Gansu	5	Anhui	6
Guizhou	2	Sichuan	2	Ningxia	4	Guizhou	4	Sichuan	5
Guangxi	2	Anhui	1	Anhui	3	Sichuan	4	Yunnan	4
Ningxia	2	Tibet	0	Guangxi	2	Anhui	3	Guangxi	2
Tibet	1	Guizhou	0	Guizhou	2	Yunnan	2	Tibet	1
Qinghai	0	Guangxi	0	Yunnan	0	Tibet	1	Guizhou	1
Yunnan	0	Ningxia	0	Tibet	0	Guangxi	0	Gansu	0

Appendix 2.1 *Data analysis of Science and Innovation Index (SII) 1998-2010*



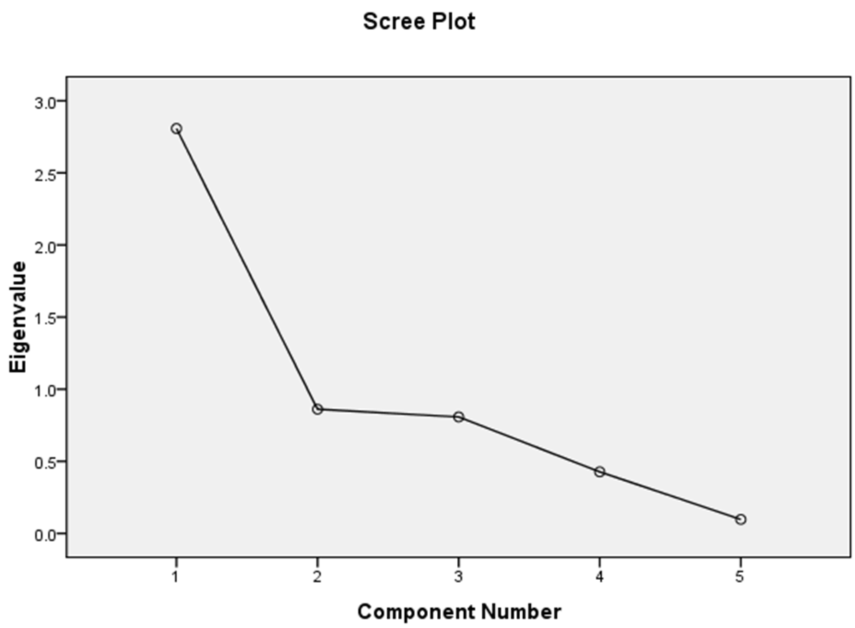
Appendix 2.2-i *PCA result of Science and Innovation Index (SII) 1998-2000*

Component Matrix[a]

	Component
	1
gov_exp_SII	.554
labor_prod_high_S&T_firms	.847
income_percapita_high_S&T_firms	.907
R&D_exp	.666
trade_S&T_market	.720

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.808	56.159	56.159	2.808	56.159	56.159
2	.861	17.217	73.376			
3	.808	16.151	89.527			
4	.427	8.539	98.066			
5	.097	1.934	100.000			



Appendix 2.2-ii *PCA result of Science and Innovation Index (SII) 2001-2005*

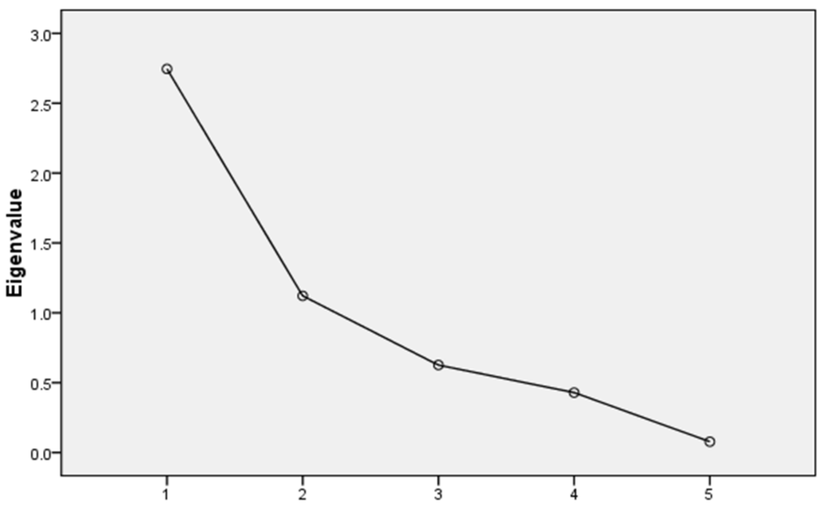
Component Matrix[a]

	Component
	1
gov_exp_SII	.658
labor_prod_high_S&T_firms	.803
income_percapita_high_S&T_firms	.846
R&D_exp	.710
trade_S&T_market	.649

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.746	54.916	54.916	2.746	54.916	54.916
2	1.121	22.427	77.343			
3	.626	12.524	89.867			
4	.429	8.578	98.445			
5	.078	1.555	100.000			

Scree Plot



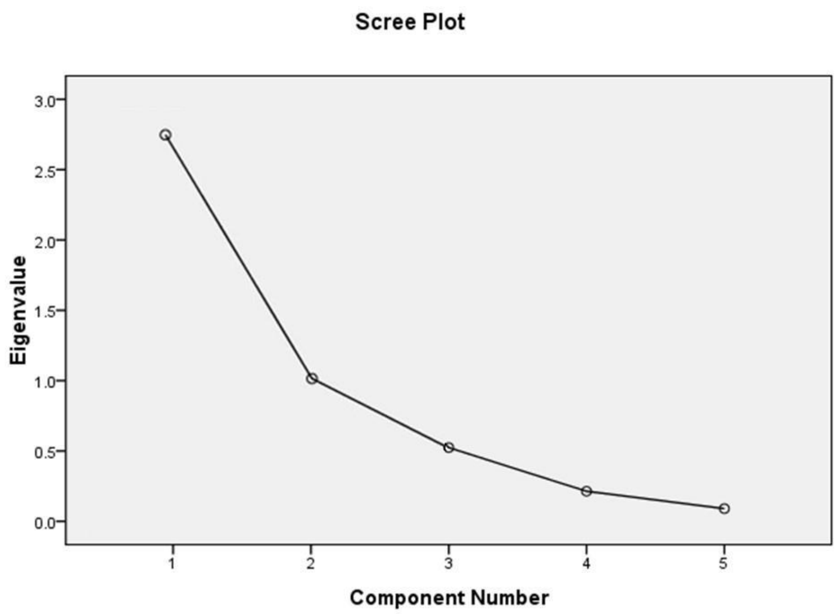
Appendix 2.2-iii *PCA result of Science and Innovation Index (SII) 2006-2010*

Component Matrix[a]

	Component
	1
gov_exp_SII	.741
labor_prod_high_S&T_firms	.659
income_percapita_high_S&T_firms	.704
R&D_exp	.853
trade_S&T_market	.658

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.766	55.324	55.324	2.766	55.324	55.324
2	1.406	28.124	83.448			
3	.524	10.480	93.928			
4	.213	4.269	98.197			
5	.090	1.803	100.000			



Appendix 2.3 *Ranking Report of Science and Innovation (SII) 1998-2010*

Province_1998	*SII*	*Province_1999*	*SII*	*Province_2000*	*SII*
Jiangsu	100	Shanghai	100	Shanghai	100
Guangdong	100	Jiangsu	90	Jiangsu	94
Zhejiang	99	Guangdong	86	Zhejiang	83
Fujian	92	Zhejiang	81	Guangdong	76
Shanghai	90	Beijing	75	Fujian	69
Beijing	77	Fujian	72	Beijing	67
Hunan	61	Tianjin	54	Shandong	46
Chongqing	50	Hunan	42	Tianjin	44
Shandong	48	Shandong	39	Hunan	34
Tianjin	44	Yunnan	35	Chongqing	31
Jilin	34	Chongqing	28	Jilin	21
Sichuan	28	Jiangxi	26	Yunnan	20
Anhui	27	Jilin	25	Shaanxi	19
Jiangxi	27	Hubei	23	Hubei	18
Yunnan	25	Shaanxi	23	Anhui	18
Hubei	21	Anhui	20	Jiangxi	17
Heilongjiang	16	Sichuan	20	Sichuan	16
Shaanxi	12	Henan	13	Hainan	15
Henan	12	Hainan	12	Liaoning	13
Guangxi	10	Heilongjiang	11	Guangxi	13
Liaoning	9	Guangxi	10	Heilongjiang	13
Hebei	9	Xinjiang	9	Henan	10
Hainan	5	Liaoning	8	Xinjiang	7
Xinjiang	3	Hebei	6	Hebei	5
Ningxia	3	Ningxia	3	Ningxia	5
Gansu	3	Shanxi	2	Inner Mongolia	3
Inner Mongolia	1	Guizhou	2	Shanxi	3
Guizhou	1	Inner Mongolia	1	Qinghai	1
Qinghai	0	Gansu	0	Gansu	1
Shanxi	0	Qinghai	0	Guizhou	0

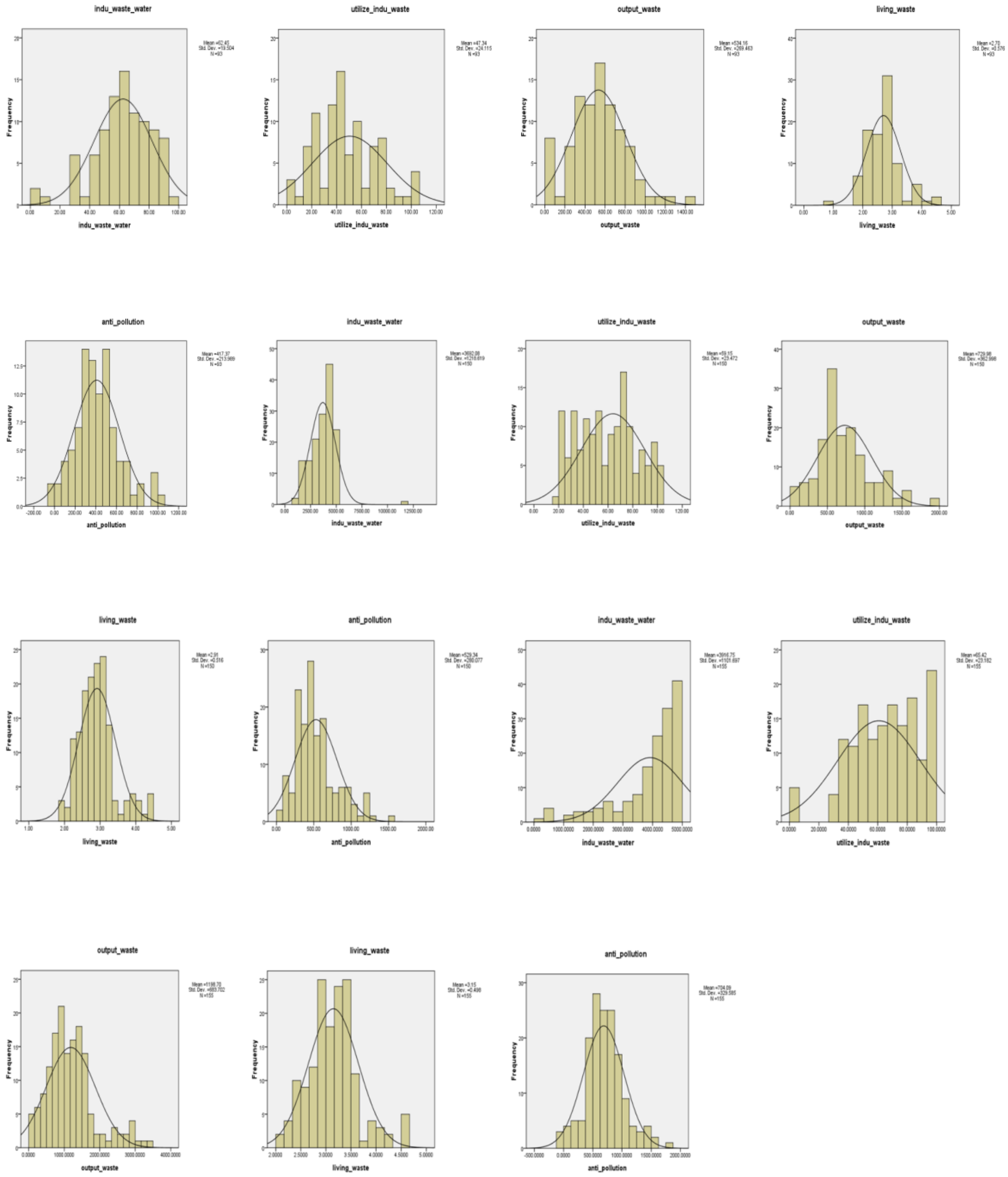
(Continued i)

Province_2001	*SII*	*Province_2002*	*SII*	*Province_2003*	*SII*	*Province_2004*	*SII*	*Province_2005*	*SII*
Shanghai	100	Shanghai	100	Shanghai	100	Shanghai	100	Shanghai	100
Jiangsu	90	Fujian	73	Guangdong	80	Beijing	85	Beijing	85
Zhejiang	89	Beijing	73	Beijing	77	Guangdong	81	Guangdong	77
Beijing	79	Guangdong	72	Jiangsu	75	Fujian	74	Jiangsu	73
Guangdong	74	Jiangsu	69	Zhejiang	74	Jiangsu	71	Fujian	70
Fujian	65	Zhejiang	62	Fujian	73	Tianjin	63	Zhejiang	47
Shandong	49	Tianjin	46	Tianjin	67	Zhejiang	52	Shandong	43
Chongqing	35	Shandong	36	Shandong	40	Shandong	40	Sichuan	37
Tianjin	35	Jilin	30	Chongqing	37	Chongqing	31	Tianjin	36
Jilin	23	Chongqing	29	Anhui	32	Jilin	30	Jilin	34
Hunan	22	Hunan	28	Jilin	30	Anhui	27	Anhui	33
Yunnan	21	Anhui	23	Sichuan	21	Sichuan	26	Chongqing	32
Hubei	18	Yunnan	20	Hunan	20	Liaoning	22	Hunan	24
Shaanxi	17	Heilongjiang	18	Liaoning	19	Hunan	21	Shaanxi	23
Anhui	17	Liaoning	18	Heilongjiang	19	Yunnan	19	Liaoning	23
Heilongjiang	17	Hebei	17	Hainan	19	Heilongjiang	19	Hubei	21
Hainan	17	Hubei	17	Yunnan	19	Shaanxi	17	Hainan	20
Sichuan	15	Shaanxi	15	Hebei	18	Hainan	15	Heilongjiang	20
Liaoning	15	Guangxi	12	Hubei	12	Hebei	14	Yunnan	18
Hebei	9	Sichuan	12	Shaanxi	12	Hubei	13	Henan	18
Guangxi	9	Hainan	11	Xinjiang	8	Henan	10	Hebei	17
Xinjiang	9	Xinjiang	9	Shanxi	8	Xinjiang	6	Ningxia	12
Ningxia	8	Ningxia	6	Guangxi	6	Inner Mongolia	3	Xinjiang	12
Henan	4	Henan	2	Inner Mongolia	5	Shanxi	3	Guangxi	7
Jiangxi	4	Gansu	2	Jiangxi	5	Guangxi	2	Jiangxi	6
Shanxi	2	Inner Mongolia	2	Henan	4	Jiangxi	1	Shanxi	5
Inner Mongolia	1	Qinghai	2	Ningxia	2	Ningxia	1	Inner Mongolia	5
Gansu	1	Jiangxi	1	Gansu	1	Qinghai	1	Gansu	2
Guizhou	0	Shanxi	1	Guizhou	1	Gansu	0	Guizhou	1
Qinghai	0	Guizhou	0	Qinghai	0	Guizhou	0	Qinghai	0

(Continued ii)

Province_2006	SII	Province_2007	SII	Province_2008	SII	Province_2009	SII	Province_2010	SII
Shanghai	100	Beijing	100	Beijing	100	Beijing	100	Beijing	100
Beijing	90	Shanghai	94	Guangdong	89	Guangdong	94	Guangdong	95
Guangdong	73	Guangdong	93	Shanghai	88	Shanghai	82	Fujian	88
Fujian	73	Fujian	79	Fujian	75	Fujian	76	Shanghai	71
Sichuan	63	Sichuan	67	Jiangsu	70	Chongqing	63	Sichuan	67
Jiangsu	61	Jiangsu	66	Shandong	48	Jiangsu	54	Chongqing	65
Zhejiang	49	Shandong	45	Tianjin	46	Sichuan	52	Tianjin	55
Shandong	41	Anhui	40	Sichuan	43	Shandong	50	Jiangsu	49
Tianjin	38	Zhejiang	40	Chongqing	42	Tianjin	47	Shandong	48
Chongqing	37	Tianjin	39	Hubei	39	Hebei	44	Anhui	39
Anhui	37	Jilin	35	Jilin	39	Anhui	41	Jilin	39
Jilin	32	Chongqing	35	Hebei	39	Hubei	39	Hebei	37
Hunan	26	Hubei	34	Anhui	39	Jilin	38	Hubei	35
Hainan	26	Hebei	33	Heilongjiang	37	Hunan	33	Heilongjiang	32
Hubei	26	Shaanxi	31	Shaanxi	35	Heilongjiang	32	Liaoning	31
Liaoning	24	Henan	29	Zhejiang	33	Yunnan	31	Hunan	30
Hebei	24	Hainan	29	Liaoning	31	Liaoning	29	Shaanxi	28
Shaanxi	23	Hunan	28	Hunan	30	Shaanxi	29	Yunnan	27
Heilongjiang	22	Jiangxi	27	Henan	29	Jiangxi	29	Henan	26
Henan	19	Liaoning	26	Jiangxi	26	Ningxia	27	Gansu	23
Xinjiang	9	Heilongjiang	24	Hainan	24	Henan	26	Jiangxi	22
Jiangxi	9	Ningxia	22	Gansu	22	Gansu	24	Ningxia	21
Yunnan	9	Yunnan	19	Inner Mongolia	19	Zhejiang	24	Zhejiang	19
Inner Mongolia	8	Inner Mongolia	14	Yunnan	18	Inner Mongolia	21	Hainan	17
Shanxi	8	Gansu	14	Ningxia	17	Hainan	20	Inner Mongolia	17
Ningxia	7	Xinjiang	13	Shanxi	14	Shanxi	20	Shanxi	15
Gansu	6	Shanxi	11	Xinjiang	7	Xinjiang	7	Xinjiang	10
Guangxi	4	Guangxi	5	Guangxi	4	Guangxi	1	Qinghai	1
Guizhou	0	Qinghai	3	Qinghai	0	Qinghai	0	Guizhou	0
Qinghai	0	Guizhou	0	Guizhou	0	Guizhou	0	Guangxi	0

Appendix 3.1 *Data Analysis of Environmental Sustainability Index(ESI) 1998-2010*



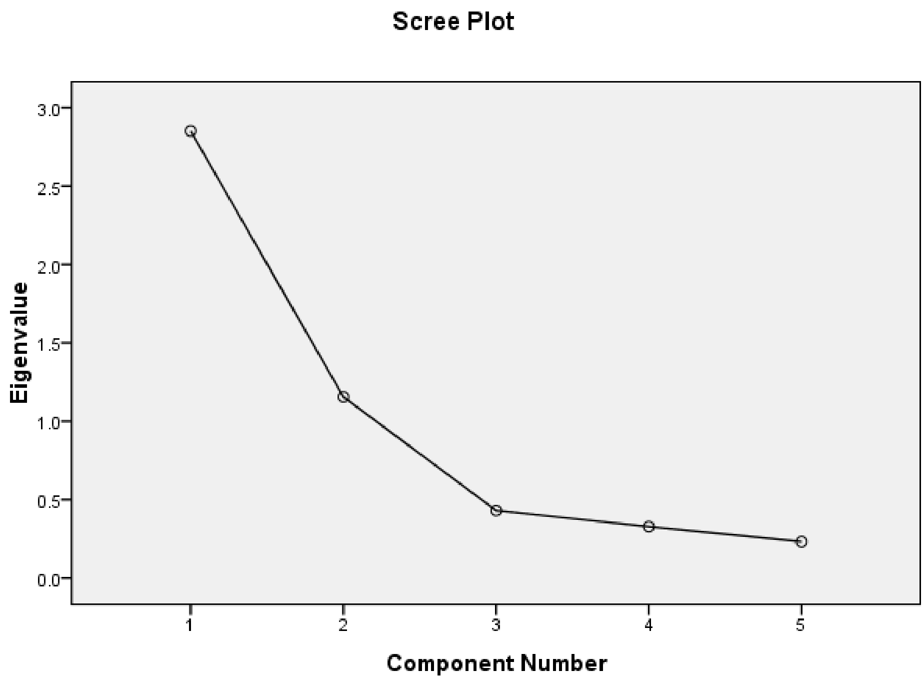
Appendix 3.2-i *PCA Results of Environmental Sustainability Index (ESI) 1998-2000*

Component Matrix[a]

	Component	
	1	2
indu_waste_water	.831	-.243
utilize_indu_waste	.838	-.269
output_waste	.598	.710
living_waste	.688	-.543
anti_pollution	.793	.475

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.852	57.039	57.039	2.852	57.039	57.039
2	1.156	23.113	80.152	1.156	23.113	80.152
3	.431	8.620	88.772			
4	.329	6.576	95.347			
5	.233	4.653	100.000			



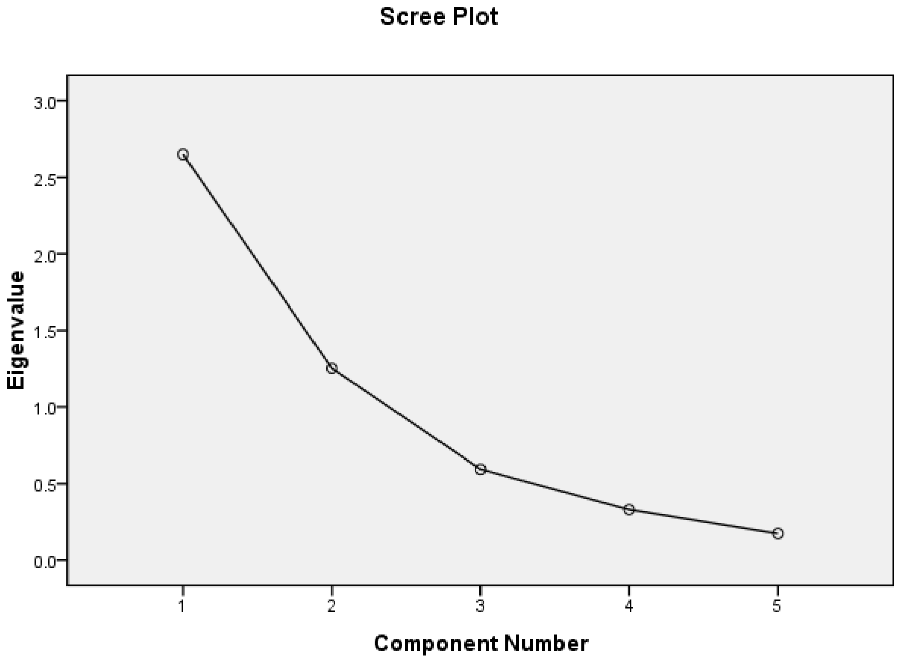
Appendix 3.2-ii *PCA Results of Environmental Sustainability Index (ESI) 2001-2005*

Component Matrix[a]

	Component	
	1	2
indu_waste_water	.718	.208
utilize_indu_waste	.830	.344
output_waste	.744	-.576
living_waste	.547	.707
anti_pollution	.770	-.509

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.649	52.986	52.986	2.649	52.986	52.986
2	1.253	25.056	78.042	1.253	25.056	78.042
3	.594	11.872	89.913			
4	.331	6.612	96.526			
5	.174	3.474	100.000			



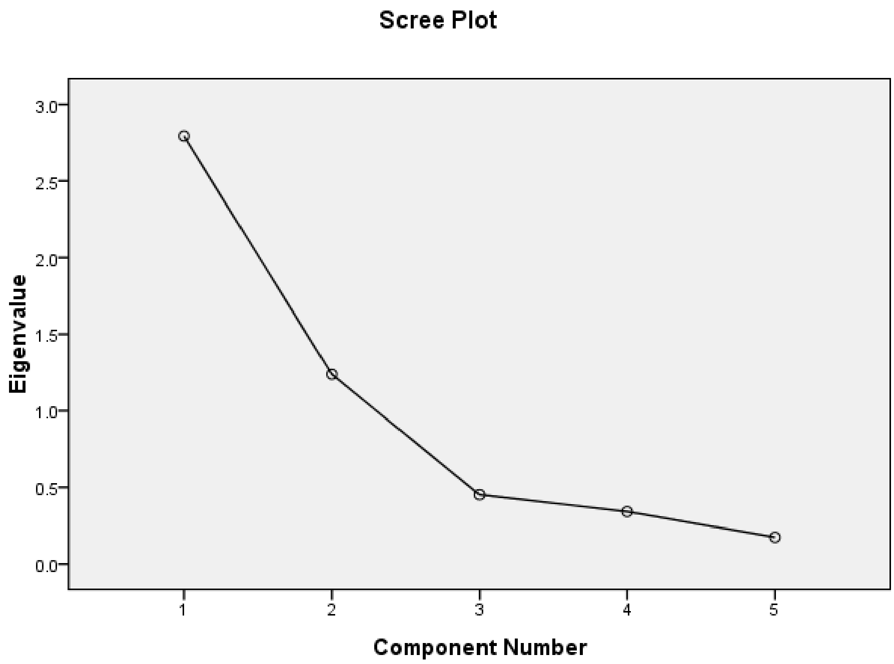
***Appendix 3.2-iii** PCA Results of Environmental Sustainability Index (ESI) 2006-2010*

Component Matrix[a]

	Component	
	1	2
indu_waste_water	.832	-.159
utilize_indu_waste	.899	-.263
output_waste	.715	.558
living_waste	.596	-.697
anti_pollution	.653	.589

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.794	55.872	55.872	2.794	55.872	55.872
2	1.238	24.768	80.640	1.238	24.768	80.640
3	.452	9.038	89.678			
4	.342	6.833	96.512			
5	.174	3.488	100.000			



Appendix 3.3 *Ranking Report of Environmental Sustainability Index (ESI) 1998-2010*

Province_1998	*ESI*	*Province_1999*	*ESI*	*Province_2000*	*ESI*	*Province_2001*	*ESI*
Jiangsu	100	Jiangsu	100	Jiangsu	100	Guangdong	100
Beijing	78	Guangdong	93	Guangdong	100	Zhejiang	100
Shandong	72	Beijing	86	Zhejiang	94	Sichuan	80
Zhejiang	64	Zhejiang	83	Hainan	89	Jiangsu	78
Heilongjiang	61	Hainan	76	Shandong	80	Heilongjiang	54
Guangdong	59	Hebei	73	Heilongjiang	55	Ningxia	53
Ningxia	54	Ningxia	61	Ningxia	55	Hubei	51
Shanghai	53	Shandong	52	Hubei	51	Hebei	50
Sichuan	51	Shanghai	50	Shanghai	49	Shandong	42
Liaoning	49	Henan	49	Hebei	46	Liaoning	41
Hainan	45	Sichuan	46	Jilin	45	Hunan	41
Henan	45	Guangxi	42	Liaoning	43	Hainan	38
Hebei	43	Heilongjiang	41	Sichuan	43	Tianjin	36
Xinjiang	39	Hunan	40	Inner Mongolia	38	Guangxi	36
Tianjin	35	Liaoning	38	Jiangxi	35	Beijing	30
Guangxi	35	Xinjiang	37	Shanxi	33	Shanghai	26
Hubei	33	Qinghai	34	Hunan	32	Inner Mongolia	24
Qinghai	25	Anhui	33	Xinjiang	31	Jiangxi	22
Jilin	25	Jilin	32	Beijing	30	Qinghai	20
Fujian	24	Hubei	31	Guangxi	30	Xinjiang	20
Anhui	24	Tianjin	28	Tianjin	29	Shanxi	19
Inner Mongolia	18	Fujian	26	Qinghai	26	Yunnan	13
Jiangxi	17	Jiangxi	24	Henan	24	Anhui	13
Hunan	15	Yunnan	22	Fujian	24	Fujian	12
Shanxi	11	Shanxi	14	Yunnan	22	Jilin	9
Guizhou	10	Gansu	10	Anhui	18	Henan	6
Yunnan	7	Chongqing	8	Guizhou	11	Guizhou	6
Chongqing	1	Inner Mongolia	7	Gansu	10	Chongqing	3
Gansu	1	Guizhou	1	Chongqing	1	Gansu	2
Shaanxi	0	Shaanxi	0	Shaanxi	0	Shaanxi	0
Tibet	--	Tibet	--	Tibet	--	Tibet	--

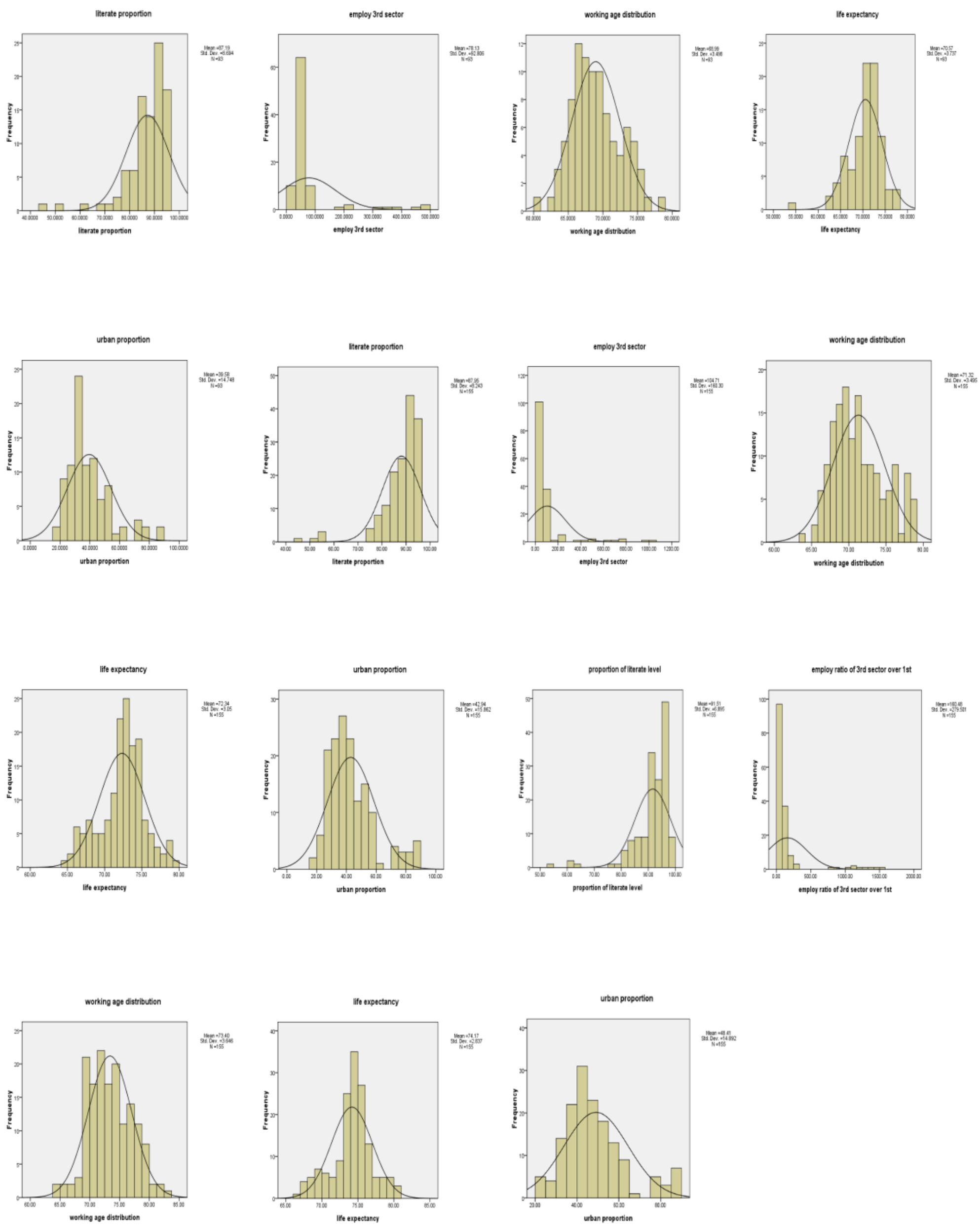
(continued i)

Province_2002	*ESI*	*Province_2003*	*ESI*	*Province_2004*	*ESI*	*Province_2005*	*ESI*
Jiangsu	100	Guangdong	100	Guangdong	100	Guangdong	100
Zhejiang	88	Jiangsu	79	Jiangsu	90	Jiangsu	100
Hebei	72	Zhejiang	68	Hainan	84	Zhejiang	100
Heilongjiang	66	Hainan	64	Shandong	82	Hainan	95
Ningxia	61	Hebei	62	Hubei	72	Hubei	77
Guangdong	60	Shanxi	52	Shanxi	63	Inner Mongolia	61
Hainan	53	Ningxia	51	Zhejiang	62	Fujian	58
Shanghai	53	Hubei	48	Ningxia	61	Yunnan	50
Shandong	50	Tianjin	47	Fujian	52	Shanxi	48
Hubei	47	Shandong	47	Beijing	44	Ningxia	48
Tianjin	43	Beijing	46	Tianjin	43	Shandong	47
Sichuan	41	Sichuan	43	Liaoning	43	Sichuan	44
Jiangxi	36	Fujian	39	Sichuan	42	Xinjiang	43
Xinjiang	35	Jiangxi	35	Qinghai	38	Hunan	40
Hunan	35	Liaoning	34	Shanghai	38	Shanghai	38
Shanxi	33	Inner Mongolia	33	Henan	37	Beijing	35
Qinghai	30	Xinjiang	33	Hunan	33	Tianjin	33
Inner Mongolia	28	Shanghai	32	Xinjiang	28	Jiangxi	32
Guangxi	28	Yunnan	31	Yunnan	26	Guangxi	29
Henan	26	Qinghai	31	Hebei	25	Henan	27
Beijing	24	Chongqing	30	Guizhou	23	Chongqing	25
Liaoning	23	Henan	29	Inner Mongolia	21	Liaoning	22
Anhui	20	Hunan	28	Anhui	21	Qinghai	18
Fujian	18	Anhui	27	Guangxi	20	Guizhou	17
Yunnan	16	Guangxi	23	Heilongjiang	20	Hebei	17
Chongqing	10	Heilongjiang	21	Jiangxi	18	Jilin	10
Jilin	8	Jilin	16	Chongqing	17	Anhui	9
Guizhou	7	Guizhou	8	Jilin	11	Gansu	2
Gansu	3	Gansu	0	Shaanxi	1	Shaanxi	1
Shaanxi	0	Shaanxi	0	Gansu	0	Heilongjiang	0
Tibet	-	Tibet	-	Tibet	-	Tibet	-

(continued ii)

Province_2006	*ESI*	*Province_2007*	*ESI*	*Province_2008*	*ESI*	*Province_2009*	*ESI*	*Province_2010*	*ESI*
Jiangsu	100	Zhejiang	100	Jiangsu	100	Jiangsu	100	Jiangsu	100
Zhejiang	81	Jiangsu	100	Zhejiang	85	Zhejiang	72	Shanghai	77
Guangdong	75	Shandong	69	Shandong	72	Shanghai	64	Hunan	72
Hainan	74	Shanghai	65	Shanghai	68	Shandong	59	Guangdong	59
Shanghai	63	Guangdong	63	Guangdong	68	Tianjin	58	Beijing	55
Shandong	63	Hubei	59	Xinjiang	61	Hunan	55	Shandong	54
Xinjiang	56	Shanxi	57	Shanxi	60	Ningxia	50	Ningxia	53
Ningxia	51	Hainan	50	Ningxia	55	Hebei	44	Zhejiang	51
Shanxi	51	Ningxia	49	Hebei	53	Hubei	44	Inner Mongolia	47
Hunan	46	Henan	49	Tianjin	51	Guangdong	44	Tianjin	45
Henan	45	Beijing	46	Hubei	50	Inner Mongolia	43	Hubei	44
Sichuan	45	Hunan	46	Inner Mongolia	48	Sichuan	42	Yunnan	42
Liaoning	44	Liaoning	46	Hainan	45	Hainan	41	Hebei	42
Hubei	43	Sichuan	45	Yunnan	45	Shanxi	41	Hainan	42
Inner Mongolia	38	Yunnan	43	Hunan	43	Yunnan	40	Henan	36
Tianjin	38	Inner Mongolia	38	Henan	43	Henan	36	Xinjiang	36
Yunnan	36	Guizhou	36	Sichuan	39	Xinjiang	36	Shanxi	35
Qinghai	32	Xinjiang	32	Anhui	38	Guizhou	34	Guizhou	32
Beijing	31	Hebei	32	Beijing	35	Liaoning	33	Jiangxi	32
Hebei	27	Qinghai	31	Liaoning	30	Fujian	31	Qinghai	31
Guangxi	25	Tianjin	30	Guizhou	29	Anhui	31	Anhui	28
Guizhou	24	Jiangxi	27	Guangxi	29	Beijing	26	Shaanxi	26
Jiangxi	22	Guangxi	27	Qinghai	26	Guangxi	23	Guangxi	25
Fujian	20	Anhui	16	Jiangxi	25	Qinghai	22	Sichuan	18
Chongqing	20	Chongqing	15	Heilongjiang	25	Jiangxi	19	Fujian	17
Anhui	8	Fujian	13	Shaanxi	9	Shaanxi	8	Liaoning	11
Jilin	2	Heilongjiang	2	Fujian	8	Jilin	5	Jilin	10
Shaanxi	2	Gansu	1	Jilin	7	Gansu	4	Heilongjiang	1
Gansu	0	Jilin	0	Chongqing	1	Chongqing	0	Chongqing	0
Heilongjiang	0	Shaanxi	0	Gansu	0	Heilongjiang	0	Gansu	0
Tibet	-	Tibet	-	Tibet	-	Tibet	-	Tibet	-

Appendix 4.1 *Data analysis of Human Capital Index (HCI) 1998-2010*



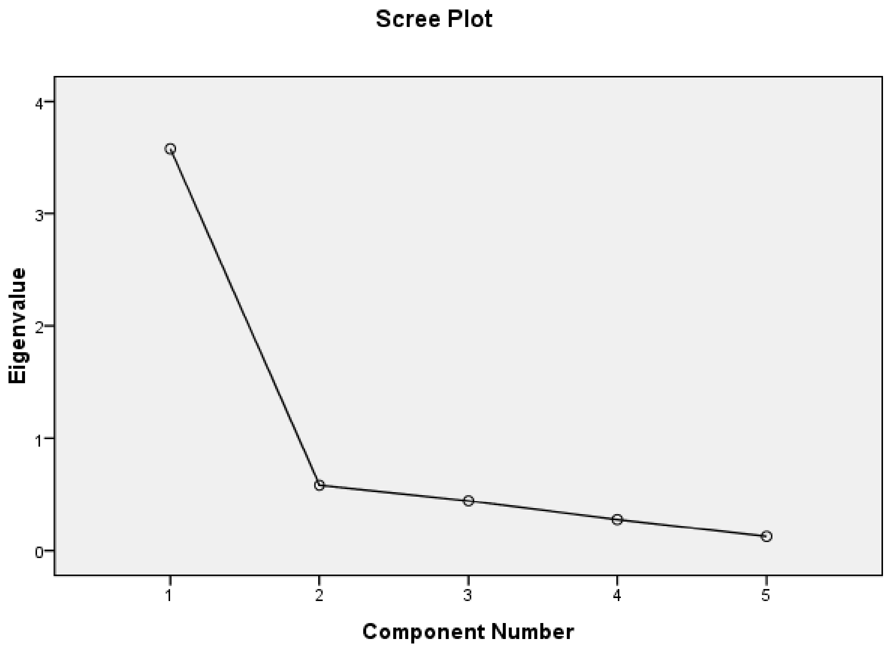
Appendix 4.2-i *PCA result of Human Capital Index (HCI) 1998-2000*

Component Matrix[a]

	Component
	1
literate proportion	.750
employ 3rd sector	.918
working age distribution	.833
life expectancy	.835
urban proportion	.884

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.578	71.564	71.564	3.578	71.564	71.564
2	.580	11.607	83.171			
3	.441	8.826	91.997			
4	.275	5.507	97.503			
5	.125	2.497	100.000			



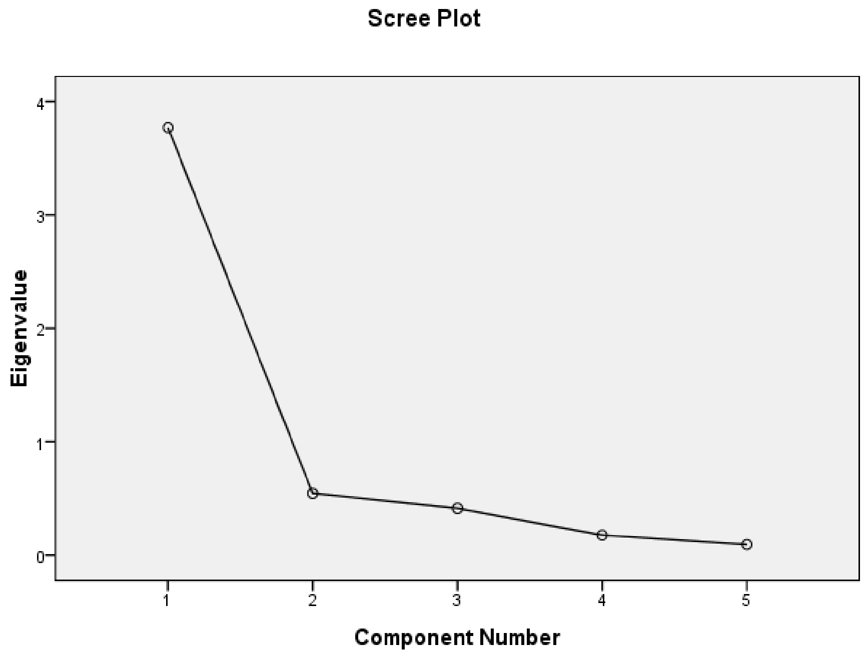
***Appendix 4.2-ii** PCA result of Human Capital Index (HCI) 2001-2005*

Component Matrix[a]

	Component
	1
literate proportion	.777
employ 3rd sector	.880
working age distribution	.816
life expectancy	.914
urban proportion	.944

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.772	75.445	75.445	3.772	75.445	75.445
2	.544	10.885	86.330			
3	.414	8.275	94.605			
4	.177	3.532	98.137			
5	.093	1.863	100.000			



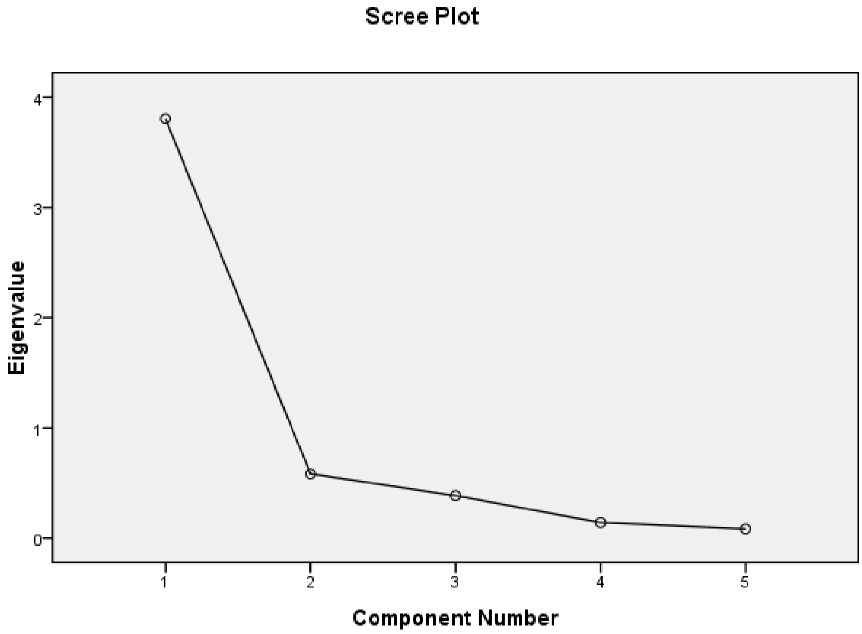
Appendix 4.2-iii *PCA result of Human Capital Index (HCI) 2006-2010*

Component Matrix[a]

	Component
	1
literate proportion	.786
employ 3rd sector	.841
working age distribution	.832
life expectancy	.929
urban proportion	.963

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.806	76.111	76.111	3.806	76.111	76.111
2	.583	11.664	87.775			
3	.386	7.710	95.485			
4	.142	2.838	98.323			
5	.084	1.677	100.000			



Appendix 4.3 *Ranking Report of Human Capital Index (HCI) 1998-2010*

Province_1998	*HCI*	*Province_1999*	*HCI*	*Province_2000*	*HCI*
Beijing	100	Beijing	100	Beijing	100
Shanghai	98	Shanghai	89	Shanghai	98
Tianjin	71	Tianjin	61	Tianjin	60
Liaoning	37	Liaoning	33	Liaoning	33
Jiangsu	33	Jiangsu	29	Guangdong	28
Guangdong	31	Guangdong	27	Zhejiang	28
Jilin	28	Zhejiang	26	Jiangsu	27
Heilongjiang	28	Jilin	25	Jilin	24
Zhejiang	28	Heilongjiang	24	Heilongjiang	24
Hubei	23	Hubei	22	Hubei	22
Fujian	22	Shanxi	21	Fujian	21
Shanxi	22	Inner Mongolia	19	Shanxi	19
Inner Mongolia	21	Fujian	19	Inner Mongolia	19
Hainan	18	Hebei	17	Jiangxi	17
Hebei	17	Hainan	17	Hainan	16
Shandong	17	Xinjiang	17	Shandong	16
Jiangxi	16	Jiangxi	16	Chongqing	16
Xinjiang	16	Shandong	16	Hebei	15
Hunan	15	Hunan	15	Shaanxi	15
Shaanxi	15	Shaanxi	15	Xinjiang	15
Sichuan	13	Chongqing	13	Guangxi	13
Chongqing	13	Sichuan	13	Hunan	13
Guangxi	12	Guangxi	13	Sichuan	13
Anhui	12	Anhui	12	Ningxia	12
Ningxia	12	Ningxia	12	Anhui	11
Henan	12	Gansu	12	Qinghai	11
Gansu	10	Qinghai	11	Gansu	10
Qinghai	8	Henan	10	Henan	8
Guizhou	5	Guizhou	7	Guizhou	6
Yunnan	4	Yunnan	6	Yunnan	5
Tibet	0	Tibet	0	Tibet	0

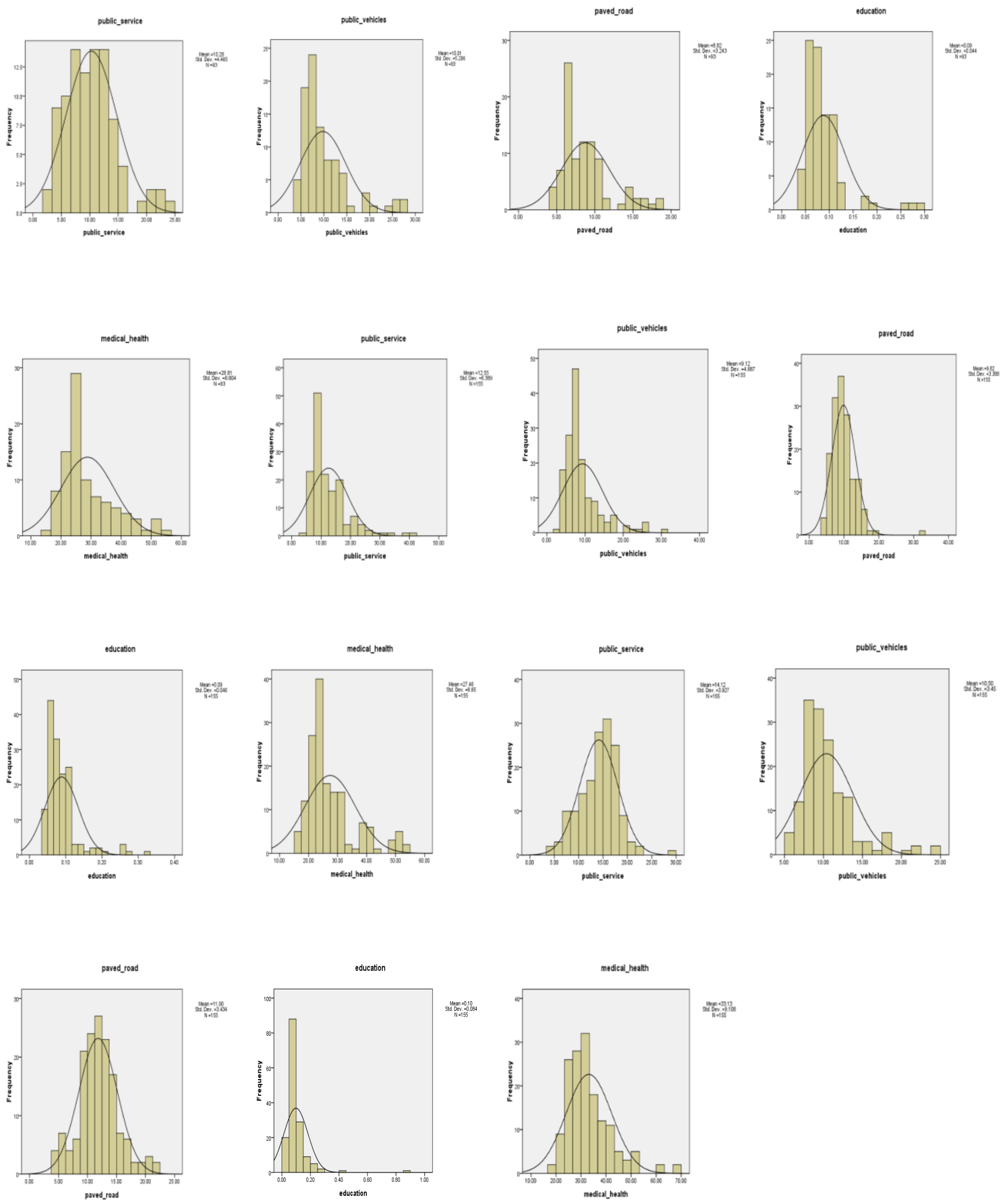
(Continued i)

Province_2001	*HCI*	*Province_2002*	*HCI*	*Province_2003*	*HCI*	*Province_2004*	*HCI*	*Province_2005*	*HCI*
Beijing	100	Beijing	100	Beijing	100	Beijing	100	Beijing	100
Shanghai	92	Shanghai	90	Shanghai	88	Shanghai	80	Shanghai	77
Tianjin	55	Tianjin	49	Tianjin	44	Tianjin	37	Tianjin	35
Liaoning	31	Liaoning	27	Jiangsu	25	Jiangsu	28	Jiangsu	27
Jiangsu	28	Jiangsu	26	Liaoning	23	Zhejiang	20	Zhejiang	20
Zhejiang	27	Zhejiang	25	Zhejiang	23	Liaoning	19	Guangdong	18
Guangdong	26	Guangdong	23	Guangdong	19	Guangdong	17	Liaoning	18
Jilin	22	Jilin	22	Jilin	16	Jilin	15	Jilin	12
Heilongjiang	22	Heilongjiang	19	Heilongjiang	15	Heilongjiang	13	Fujian	12
Hubei	20	Fujian	17	Hubei	15	Hubei	12	Hubei	12
Fujian	19	Hubei	16	Fujian	14	Fujian	12	Heilongjiang	11
Shanxi	18	Shanxi	15	Shanxi	13	Shanxi	11	Shandong	10
Inner Mongolia	17	Chongqing	15	Chongqing	13	Shandong	10	Shanxi	10
Jiangxi	16	Inner Mongolia	14	Shandong	11	Chongqing	10	Chongqing	10
Chongqing	16	Shandong	14	Inner Mongolia	11	Inner Mongolia	9	Inner Mongolia	8
Hainan	15	Hainan	14	Jiangxi	11	Hainan	9	Hainan	8
Shandong	15	Jiangxi	13	Hainan	10	Jiangxi	9	Jiangxi	8
Xinjiang	14	Xinjiang	13	Shaanxi	10	Shaanxi	8	Hebei	7
Hebei	14	Hebei	12	Xinjiang	10	Xinjiang	8	Shaanxi	7
Shaanxi	14	Shaanxi	12	Hebei	9	Hebei	8	Xinjiang	7
Guangxi	13	Hunan	11	Hunan	8	Hunan	7	Anhui	7
Hunan	13	Guangxi	11	Sichuan	8	Ningxia	7	Hunan	6
Sichuan	12	Sichuan	10	Guangxi	8	Guangxi	7	Ningxia	6
Ningxia	12	Ningxia	10	Ningxia	8	Sichuan	7	Guangxi	6
Anhui	11	Qinghai	10	Anhui	7	Qinghai	7	Qinghai	6
Qinghai	10	Anhui	9	Qinghai	7	Anhui	6	Sichuan	6
Gansu	9	Gansu	7	Gansu	5	Henan	4	Henan	4
Henan	8	Henan	7	Henan	5	Gansu	4	Gansu	3
Guizhou	6	Guizhou	5	Guizhou	3	Guizhou	3	Guizhou	2
Yunnan	5	Yunnan	3	Yunnan	1	Yunnan	1	Yunnan	1
Tibet	0	Tibet	0	Tibet	0	Tibet	0	Tibet	0

(Continued ii)

Province_2006	HCI	Province_2007	HCI	Province_2008	HCI	Province_2009	HCI	Province_2010	HCI
Beijing	100	Beijing	100	Beijing	100	Beijing	100	Beijing	100
Shanghai	86	Shanghai	86	Shanghai	93	Shanghai	94	Shanghai	94
Tianjin	32	Jiangsu	34	Jiangsu	38	Guangdong	57	Guangdong	58
Jiangsu	28	Tianjin	31	Tianjin	36	Jiangsu	37	Jiangsu	38
Zhejiang	21	Guangdong	25	Guangdong	29	Tianjin	34	Tianjin	33
Guangdong	17	Zhejiang	23	Zhejiang	22	Zhejiang	22	Zhejiang	22
Liaoning	16	Liaoning	16	Liaoning	16	Liaoning	17	Liaoning	16
Fujian	11	Jilin	15	Jilin	14	Jilin	14	Jilin	14
Hubei	11	Fujian	12	Hubei	11	Fujian	12	Hebei	13
Jilin	11	Hubei	11	Fujian	11	Hubei	11	Hubei	11
Heilongjiang	10	Chongqing	10	Chongqing	10	Chongqing	10	Fujian	10
Shandong	10	Heilongjiang	10	Anhui	9	Anhui	8	Anhui	9
Chongqing	9	Shandong	10	Heilongjiang	9	Heilongjiang	8	Chongqing	8
Shanxi	9	Shanxi	9	Shandong	8	Shandong	8	Heilongjiang	6
Inner Mongolia	7	Hainan	7	Shanxi	8	Shanxi	8	Shanxi	6
Hainan	7	Jiangxi	7	Hebei	7	Hebei	7	Shandong	6
Jiangxi	7	Inner Mongolia	7	Inner Mongolia	7	Inner Mongolia	7	Inner Mongolia	5
Hebei	6	Hebei	7	Jiangxi	6	Ningxia	7	Ningxia	5
Shaanxi	6	Anhui	7	Shaanxi	6	Jiangxi	6	Jiangxi	4
Xinjiang	6	Shaanxi	6	Hainan	6	Hainan	6	Hainan	4
Ningxia	6	Ningxia	6	Ningxia	5	Shaanxi	5	Shaanxi	4
Anhui	6	Xinjiang	6	Sichuan	5	Sichuan	5	Qinghai	3
Hunan	6	Qinghai	6	Xinjiang	5	Qinghai	5	Sichuan	3
Qinghai	6	Hunan	6	Hunan	5	Hunan	5	Yunnan	3
Sichuan	5	Sichuan	6	Qinghai	5	Xinjiang	4	Hunan	3
Guangxi	4	Henan	3	Henan	3	Henan	3	Xinjiang	3
Henan	3	Guangxi	3	Guangxi	2	Yunnan	3	Guizhou	2
Gansu	2	Gansu	3	Gansu	2	Gansu	2	Tibet	1
Guizhou	2	Guizhou	3	Guizhou	2	Guangxi	2	Gansu	1
Yunnan	0	Tibet	0	Yunnan	1	Guizhou	2	Henan	1
Tibet	0	Yunnan	0	Tibet	0	Tibet	0	Guangxi	0

Appendix 5.1 *Data analysis of Public Facility Index (PFI) 1998-2010*



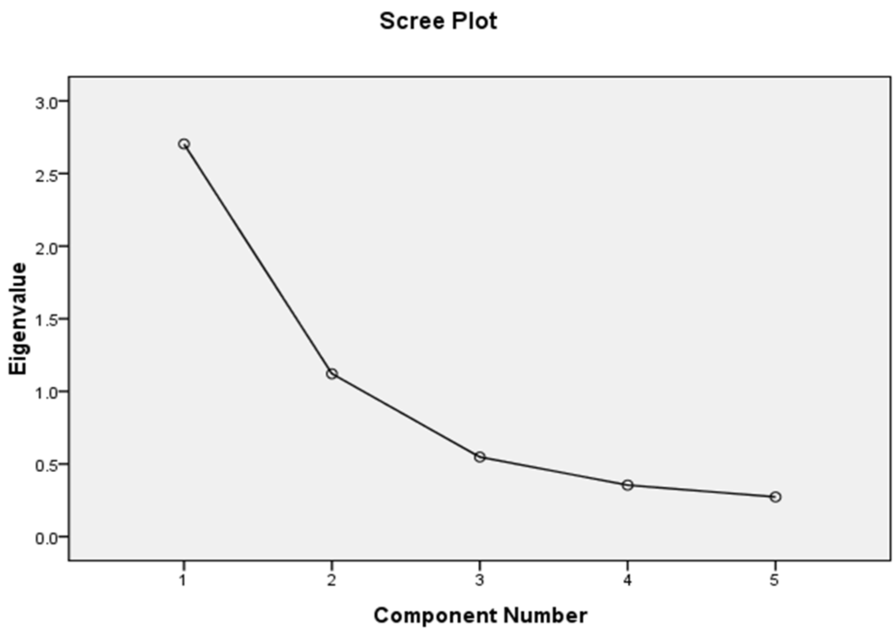
Appendix 5.2-i *PCA result of Public Facility Index (PFI) 1998-2000*

Component Matrix[a]

	Component	
	1	2
public_service	.684	-.626
public_vehicles	.810	-.119
paved_road	.599	.693
education	.794	.402
medical_health	.768	-.272

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.704	54.075	54.075	2.704	54.075	54.075
2	1.121	22.424	76.499	1.121	22.424	76.499
3	.548	10.955	87.454			
4	.355	7.091	94.545			
5	.273	5.455	100.000			



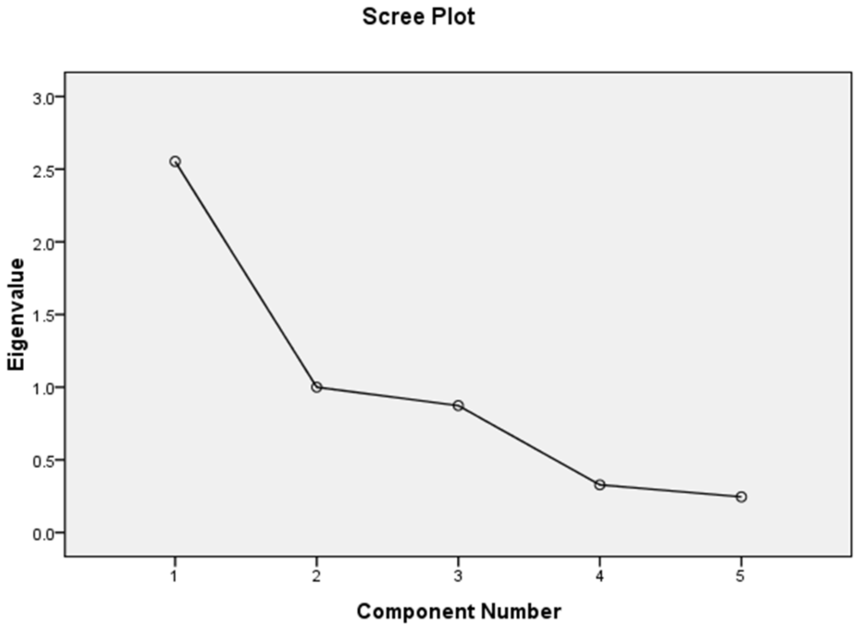
Appendix 5.2-ii *PCA result of Public Facility Index (PFI) 2001-2005*

Component Matrix[a]

	Component	
	1	2
public_service	.789	-.313
public_vehicles	.836	.253
paved_road	.544	-.234
education	.845	-.276
medical_health	.471	.841

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.554	51.072	51.072	2.554	51.072	51.072
2	1.000	20.001	71.073	1.000	20.001	71.073
3	.873	17.462	88.535			
4	.328	6.562	95.097			
5	.245	4.903	100.000			



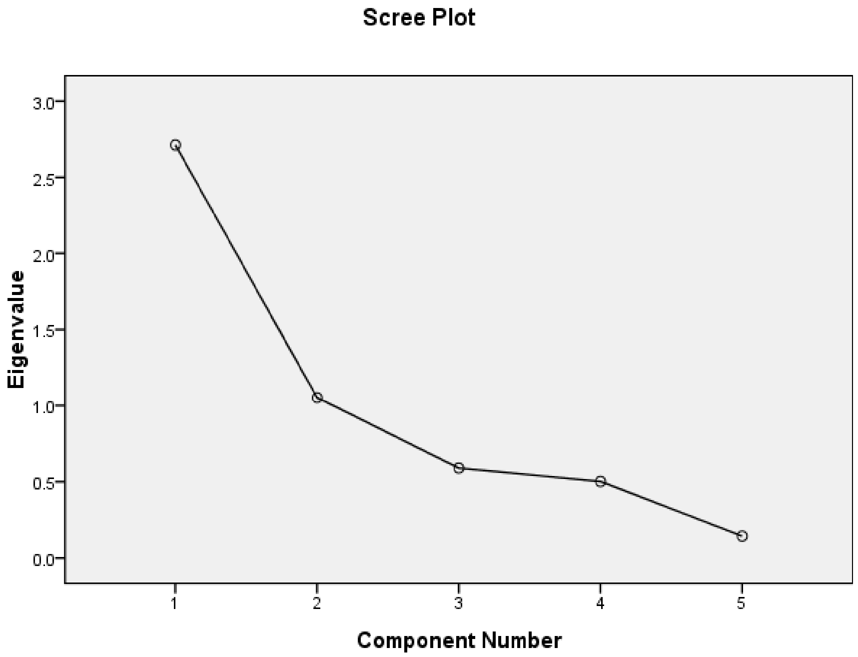
Appendix 5.2-iii *PCA result of Public Facility Index (PFI) 2001-2005*

Component Matrix[a]

	Component	
	1	2
public_service	.857	.073
public_vehicles	.921	.172
paved_road	.655	-.532
education	.467	.785
medical_health	.695	-.344

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.713	54.265	54.265	2.713	54.265	54.265
2	1.052	21.049	75.314	1.052	21.049	75.314
3	.589	11.786	87.100			
4	.502	10.033	97.133			
5	.143	2.867	100.000			



Appendix 5.3 *Ranking Report of Public Facility Index (PFI) 1998-2010*

province_1998	*PFI*	*province_1999*	*PFI*	*province_2000*	*PFI*
Shanghai	100	Shanghai	100	Shanghai	100
Beijing	94	Beijing	91	Beijing	91
Zhejiang	78	Zhejiang	73	Zhejiang	81
Jiangsu	77	Jiangsu	72	Jiangsu	80
Guangdong	75	Guangdong	66	Guangdong	65
Fujian	57	Tianjin	61	Fujian	60
Tianjin	55	Fujian	52	Tianjin	59
Liaoning	48	Liaoning	48	Liaoning	47
Hubei	44	Hubei	43	Hunan	46
Jilin	42	Yunnan	41	Hubei	43
Yunnan	42	Hunan	38	Yunnan	39
Shandong	36	Jilin	37	Shaanxi	36
Shaanxi	35	Shandong	34	Jilin	35
Hunan	33	Shaanxi	34	Heilongjiang	33
Heilongjiang	30	Heilongjiang	30	Shandong	32
Sichuan	28	Sichuan	26	Sichuan	27
Anhui	27	Anhui	25	Xinjiang	27
Hainan	22	Xinjiang	24	Hebei	26
Hebei	21	Hebei	24	Anhui	25
Chongqing	21	Guizhou	22	Shanxi	23
Inner Mongolia	21	Hainan	21	Guizhou	22
Shanxi	20	Ningxia	21	Hainan	21
Ningxia	20	Jiangxi	19	Jiangxi	21
Jiangxi	20	Chongqing	19	Ningxia	19
Guizhou	17	Shanxi	19	Chongqing	19
Xinjiang	16	Inner Mongolia	17	Inner Mongolia	16
Guangxi	14	Guangxi	13	Guangxi	15
Henan	11	Henan	11	Henan	12
Gansu	10	Gansu	8	Gansu	11
Qinghai	3	Qinghai	8	Qinghai	8
Tibet	0	Tibet	0	Tibet	0

(Continued i)

province_2001	*PFI*	*province_2002*	*PFI*	*province_2003*	*PFI*	*province_2004*	*PFI*	*province_2005*	*PFI*
Shanghai	100	Shanghai	100	Shanghai	100	Shanghai	100	Shanghai	100
Beijing	97	Beijing	97	Beijing	92	Beijing	96	Beijing	89
Jiangsu	83	Guangdong	82	Jiangsu	77	Jiangsu	81	Jiangsu	85
Zhejiang	78	Jiangsu	80	Zhejiang	75	Zhejiang	74	Zhejiang	75
Guangdong	61	Zhejiang	78	Guangdong	70	Guangdong	71	Guangdong	71
Tianjin	58	Tianjin	61	Tianjin	61	Tianjin	65	Tianjin	63
Fujian	55	Fujian	48	Fujian	54	Fujian	63	Fujian	54
Liaoning	47	Hunan	46	Hunan	44	Hunan	44	Hunan	48
Hunan	46	Liaoning	44	Liaoning	41	Liaoning	41	Liaoning	45
Hubei	43	Shandong	42	Hubei	38	Hubei	37	Hubei	40
Yunnan	41	Hubei	41	Shandong	38	Shandong	33	Shandong	37
Shandong	41	Yunnan	36	Jilin	31	Jilin	29	Jilin	32
Jilin	35	Jilin	35	Chongqing	28	Hainan	24	Chongqing	27
Chongqing	29	Shaanxi	33	Shaanxi	24	Shaanxi	20	Hainan	25
Shaanxi	28	Chongqing	28	Yunnan	22	Heilongjiang	17	Shaanxi	22
Guizhou	27	Heilongjiang	26	Sichuan	20	Chongqing	17	Yunnan	21
Heilongjiang	24	Xinjiang	25	Hainan	19	Sichuan	17	Sichuan	21
Shanxi	23	Hebei	24	Heilongjiang	19	Yunnan	16	Xinjiang	21
Xinjiang	23	Hainan	22	Xinjiang	18	Ningxia	15	Ningxia	19
Hebei	22	Ningxia	20	Ningxia	17	Hebei	15	Heilongjiang	19
Sichuan	22	Shanxi	20	Hebei	15	Xinjiang	14	Henan	17
Hainan	21	Anhui	19	Shanxi	13	Shanxi	11	Shanxi	17
Ningxia	21	Jiangxi	18	Inner Mongolia	12	Anhui	11	Guangxi	16
Anhui	19	Inner Mongolia	18	Anhui	11	Henan	8	Hebei	15
Guangxi	17	Guangxi	17	Jiangxi	10	Guangxi	6	Anhui	14
Inner Mongolia	16	Sichuan	13	Henan	7	Jiangxi	5	Inner Mongolia	8
Jiangxi	12	Guizhou	12	Qinghai	6	Inner Mongolia	5	Jiangxi	7
Gansu	12	Gansu	9	Gansu	6	Tibet	3	Gansu	4
Qinghai	8	Qinghai	7	Guizhou	5	Gansu	2	Qinghai	4
Henan	7	Henan	3	Guangxi	1	Qinghai	0	Guizhou	0
Tibet	0	Tibet	0	Tibet	0	Guizhou	0	Tibet	0

(Continued ii)

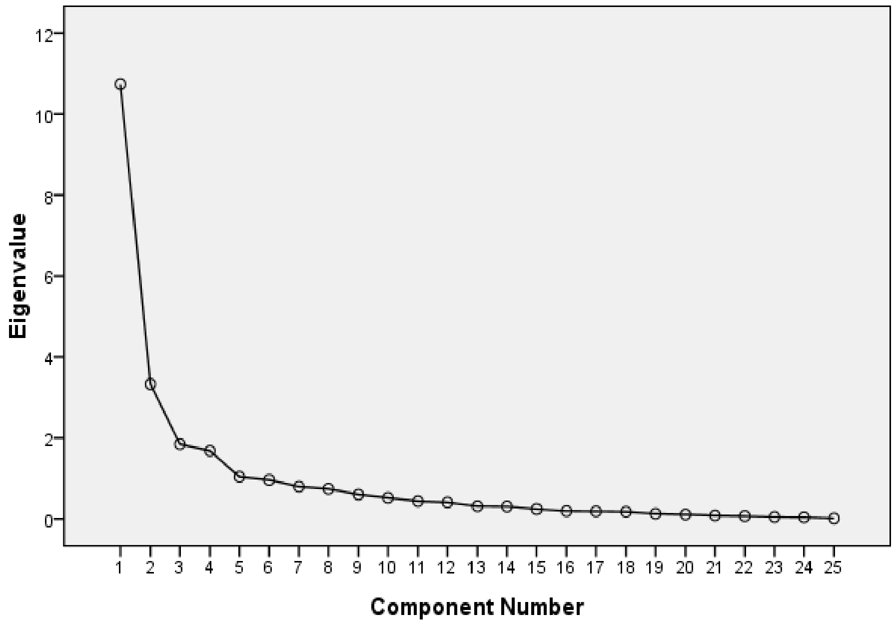
province_2006	*PFI*	*province_2007*	*PFI*	*province_2008*	*PFI*	*province_2009*	*PFI*	*province_2010*	*PFI*
Shanghai	100	Shanghai	100	Shanghai	100	Shanghai	100	Shanghai	100
Beijing	96	Beijing	89	Beijing	97	Beijing	94	Beijing	94
Jiangsu	86	Jiangsu	81	Jiangsu	89	Guangdong	85	Guangdong	92
Zhejiang	79	Guangdong	79	Guangdong	87	Jiangsu	84	Jiangsu	80
Guangdong	75	Zhejiang	78	Zhejiang	86	Zhejiang	75	Zhejiang	76
Tianjin	65	Tianjin	61	Tianjin	75	Tianjin	70	Tianjin	71
Hunan	52	Fujian	52	Liaoning	62	Fujian	62	Fujian	66
Fujian	48	Hunan	48	Fujian	59	Hunan	56	Liaoning	60
Liaoning	47	Hubei	46	Hunan	54	Hubei	55	Hunan	53
Shandong	44	Liaoning	44	Hubei	52	Liaoning	49	Hubei	52
Hubei	44	Shandong	43	Shandong	50	Shandong	43	Shandong	42
Jilin	34	Jilin	29	Hebei	34	Jilin	30	Hebei	32
Shaanxi	29	Shaanxi	28	Jilin	34	Hebei	29	Jilin	31
Hainan	28	Hebei	28	Shaanxi	33	Shaanxi	29	Heilongjiang	27
Chongqing	26	Xinjiang	24	Heilongjiang	33	Heilongjiang	27	Henan	27
Heilongjiang	25	Heilongjiang	24	Ningxia	32	Ningxia	26	Ningxia	27
Xinjiang	24	Sichuan	23	Yunnan	29	Xinjiang	25	Sichuan	26
Ningxia	23	Ningxia	23	Xinjiang	28	Shanxi	23	Shanxi	24
Anhui	21	Yunnan	22	Shanxi	26	Sichuan	23	Xinjiang	23
Henan	21	Henan	20	Sichuan	22	Henan	22	Chongqing	22
Yunnan	21	Hainan	20	Henan	22	Guangxi	22	Anhui	22
Shanxi	20	Guangxi	19	Anhui	21	Anhui	19	Guangxi	21
Sichuan	19	Shanxi	18	Guangxi	21	Yunnan	17	Yunnan	17
Hebei	18	Anhui	18	Jiangxi	19	Gansu	16	Gansu	17
Gansu	17	Chongqing	15	Gansu	19	Chongqing	16	Shaanxi	16
Guangxi	13	Jiangxi	15	Hainan	18	Jiangxi	13	Jiangxi	14
Inner Mongolia	10	Gansu	11	Chongqing	17	Hainan	12	Hainan	14
Qinghai	10	Qinghai	8	Inner Mongolia	14	Qinghai	8	Guizhou	13
Jiangxi	9	Guizhou	8	Guizhou	11	Guizhou	7	Qinghai	11
Tibet	0	Inner Mongolia	8	Qinghai	11	Inner Mongolia	6	Inner Mongolia	8
Guizhou	0	Tibet	0	Tibet	0	Tibet	0	Tibet	0

Appendix 6.1 *Principal Component Analysis of Five Dimensions (1998-2000)*

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10.737	42.946	42.946	10.737	42.946	42.946
2	3.329	13.317	56.263	3.329	13.317	56.263
3	1.847	7.387	63.650	1.847	7.387	63.650
4	1.681	6.726	70.376	1.681	6.726	70.376
5	1.041	4.163	74.539	1.041	4.163	74.539
6	.963	3.852	78.391			
7	.795	3.179	81.570			
…	…	…	…	…	…	…
25	.012	.049	100.000			

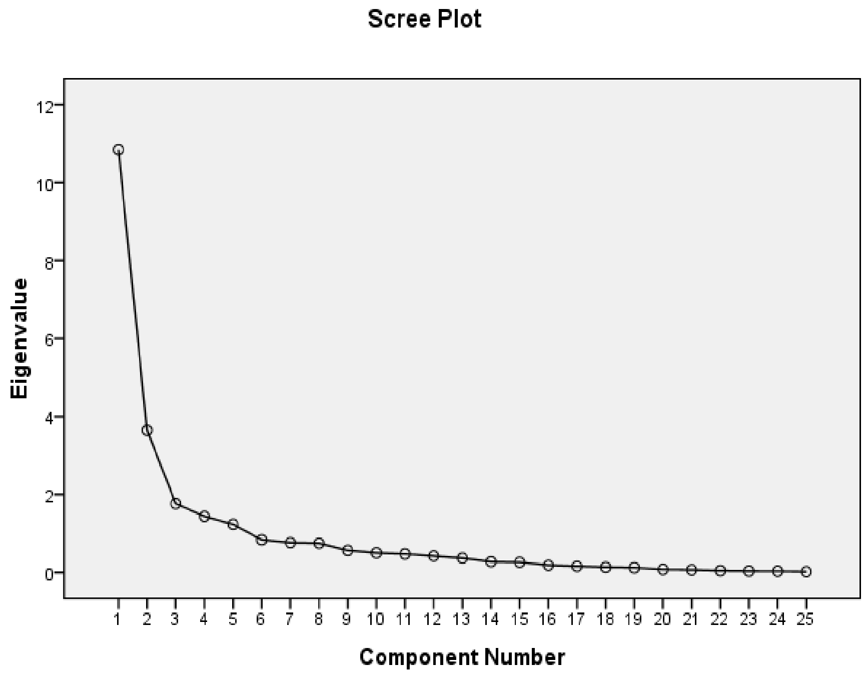
Scree Plot



Appendix 6.2 *Principal Component Analysis of Five Dimensions (2001-2005)*

Total Variance Explained

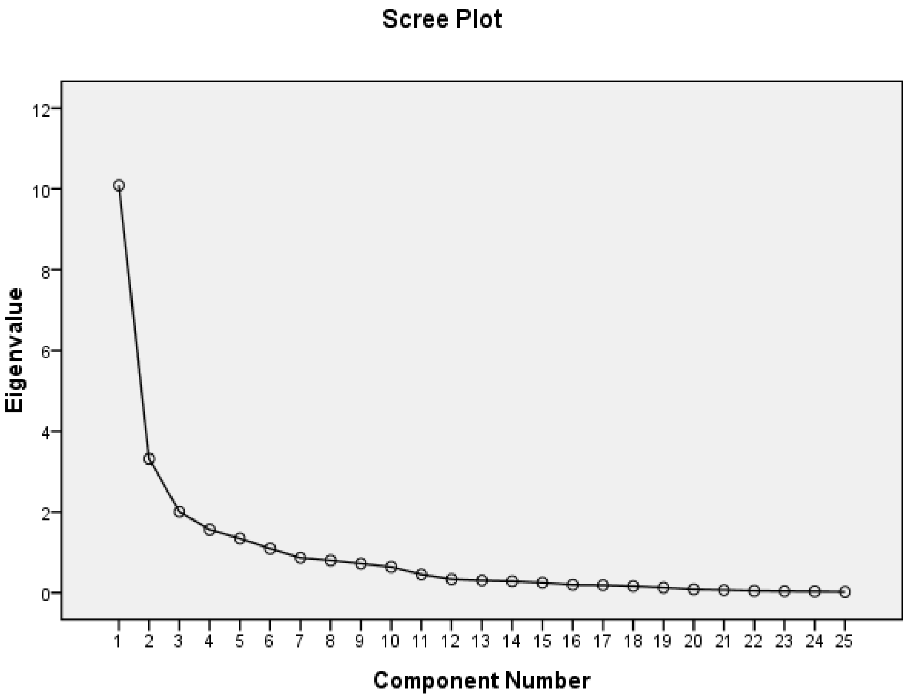
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10.851	43.402	43.402	10.851	43.402	43.402
2	3.647	14.587	57.989	3.647	14.587	57.989
3	1.764	7.055	65.044	1.764	7.055	65.044
4	1.438	5.751	70.795	1.438	5.751	70.795
5	1.232	4.928	75.723	1.232	4.928	75.723
6	.835	3.338	79.061			
7	.762	3.047	82.109			
…	…	…	…	…	…	…
25	.018	.072	100.000			



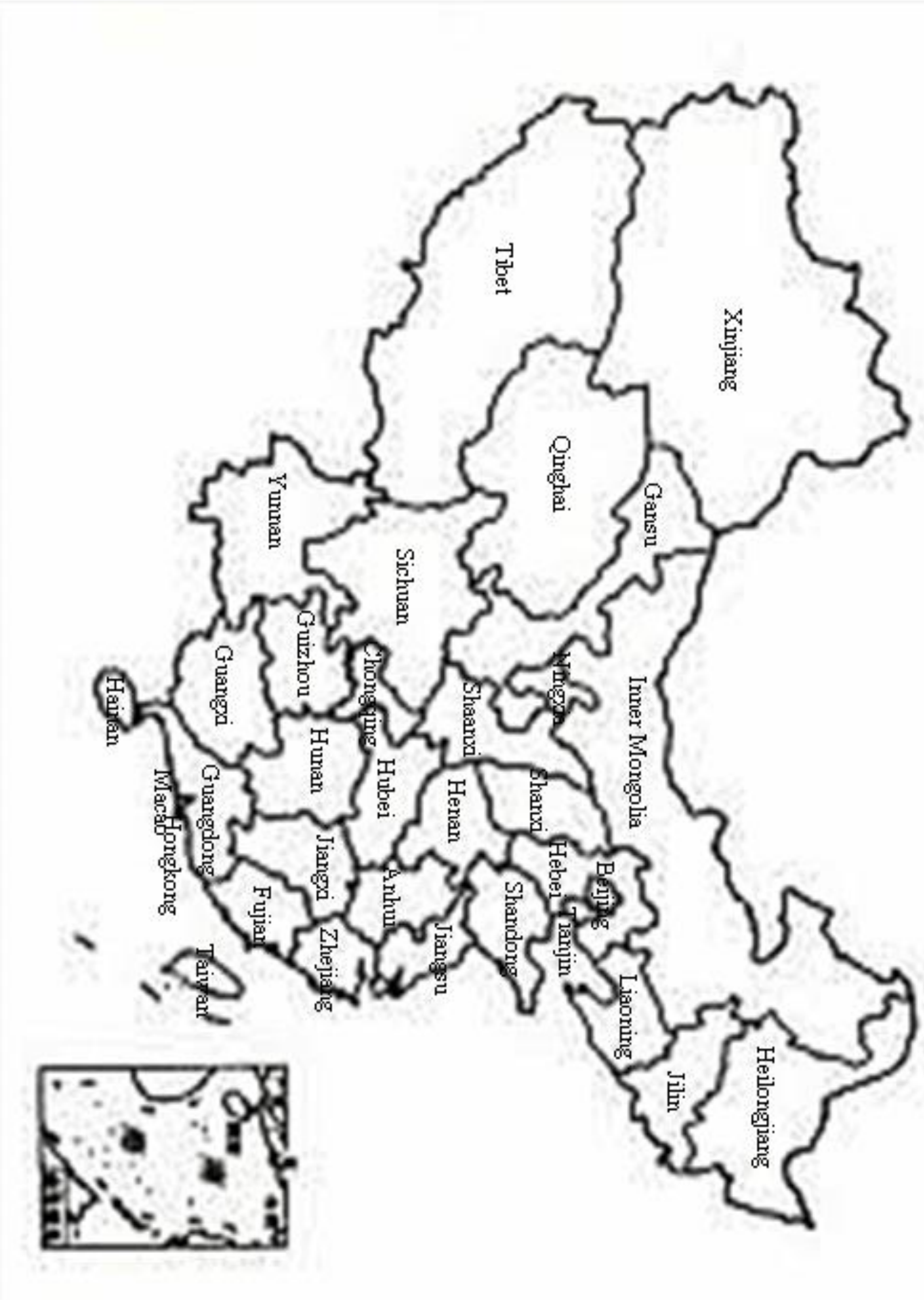
Appendix 6.3 *Principal Component Analysis of Five Dimensions (2001-2005)*

Total Variance Explained

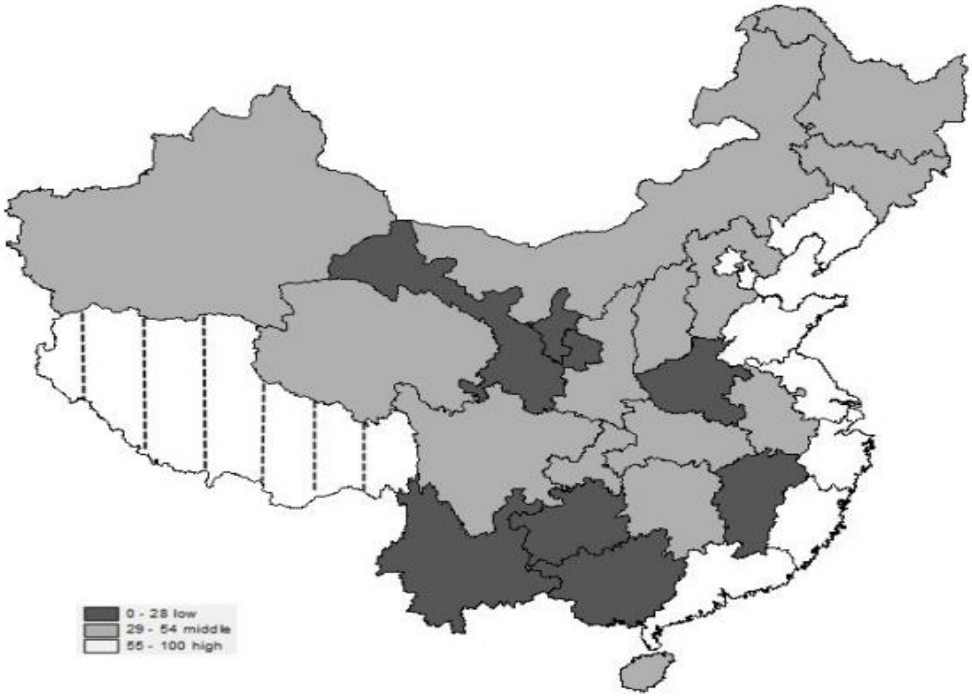
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10.087	40.347	40.347	10.087	40.347	40.347
2	3.317	13.266	53.613	3.317	13.266	53.613
3	2.004	8.018	61.631	2.004	8.018	61.631
4	1.560	6.239	67.870	1.560	6.239	67.870
5	1.343	5.374	73.244	1.343	5.374	73.244
6	1.092	4.368	77.612	1.092	4.368	77.612
7	.863	3.450	81.062			
…	…	…	…	…	…	…
25	.021	.085	100.000			

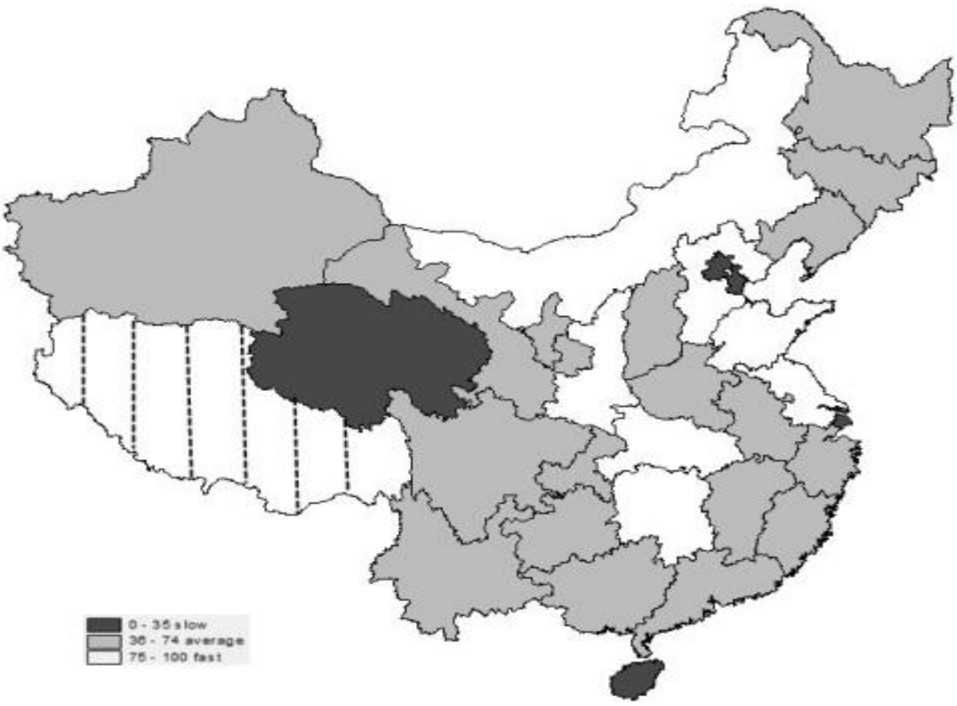


Appendix 7 *Chinese Map of Administrative Districts*



Appendix 8 *Maps of CIRD total average score and total change calculated by integral PCA (1998-2010)*





Curriculum Vitae

Personal Information

Family Name: Bin **First Name:** Peng

Gender: Female **Date of Birth:** 19/11/1984 **Nationality:** China

Academic Experience

09/2013—Present **Visiting PhD student (DAAD program)**
Food Net Center, University of Bonn
Supervisor: Professor Gerhard Schiefer, schiefer@uni-bonn.de

08/2010—Present **PhD candidate (Erasmus Mundus-CONNEC program)**
Agri-Food Economics and Statistics, University of Bologna
Supervisor: Professor Roberto Fanfani, roberto.fanfani@unibo.it

09/2007—07/2010 **Master Degree of Management**
Forestry Economics and Management, Beijing Forestry University
Supervisor: Professor Wu Tiexiong, youyou@bjfu.edu.cn

09/2003—07/2007 **Bachelor's Degree of Management**
Dept. of Economics and Management, Beijing Forestry University

09/2003—07/2007 **Minor Certificate of Law**
Dept. of Humanistic and Social Science, Beijing Forestry University

Conferences and Presentations

06/2013 Attended in the 133rd EAAE seminar and presented paper "A Latent Growth Curve Modeling: Dynamic Trajectory of Agricultural Labor Productivity on a Prefectural Level and New Agricultural-oriented Policy Reforms in China", Chania, Greece

10/2012 Attended in the congress "Italy-China: An ancient cultural heritage and the challenge for future development", and presented paper "Regional Rankings and Dynamic Changes in China: Composite Indicator of Economic Performances", Bologna, Italy.

04/2012 Attended in the 86th AES Annual Conference and presented working paper "Regional Differences and Dynamic Changes in Rural China: the Comparison Study of 1996 and 2006 National Agricultural Census, Warwick University, UK.

07/2011 Attended in the 16th IEA congress (International Economic Association Sixteenth World Congress) and co-presented working paper "The Rural Areas in China: Dynamic Changes and New Classification", Tsinghua University, Beijing, China.

05/2011 Attended in 2011 ETM International Conference on Environmental Technology and Management and presented working paper "Regional Differences of Public Sanitation in China's Rural Areas", Leuven, Belgium.

Awards

09/2013—12/2013 DAAD Short-term Research Visiting Scholarship

09/2010—07/2013 Erasmus Mundus Scholarship

07/2010	Excellent Master Graduate Thesis
07/2007	Excellent College Graduate Thesis
09/2006	Beijing Forestry University Top Grade Scholarship
09/2005	Beijing Forestry University Top Grade Scholarship
09/2004	Beijing Forestry University Top Grade Scholarship

Publications

First author

P. Bin, M. Vassallo, etc. (2013). A Latent Growth Modeling: Dynamic Change of Agricultural Productivity on a Prefectural Level and New Agricultural-Oriented Policy in China. Book of Abstracts: Developing Integrated and Reliable Modeling Tools for Agricultural and Environmental Policy Analysis, 133rd EAAE Seminar, MAICh, Chania, Greece, 15-16 June, 2013.

P. Bin, B. Barone, R. Fanfani (2013). Regional Rankings and Dynamic Changes in China: Composite Indicator of Economic Performances. Proceedings: "Atti del Convegno Italy-Cina: An ancient cultural heritage and the challenge for future development", ISBN 978-88-6155-515-0, 2013.

P. Bin, R. Fanfani, C. Brasili (2012). Regional Differences and Dynamic Changes in Rural China: the Study of 1996 and 2006 National Agricultural Census. Asian Journal of Agriculture and Rural Development, Vol.2, No.2, June, 2012: 260-270.

P. Bin, J. Wu (2011). Regional Differences of Public Sanitation in China's Rural Areas. Proceedings of 2011 International Conference on Environmental Technology and Management (ETM 2011), 2011.

P. Bin, T. Wu (2010). Analysis of Pork Price Fluctuation in China from Spider Web Theory. (in Chinese) Rural Economy and Science Technology, CN42-1374/S, January, 2010.

Second author

B. Barone, P. Bin, etc. (2013). Le dinamiche delle disuguaglianze spaziali in Cina
a livello regionale e tra zone rurali e urbane. L'industria / n.s., a. XXXIV, n. 3, luglio-settembre 2013.

N. Calo, P. Bin, etc. (2013). A New Map of Rural China: An Analysis of Living Conditions and Dynamic Changes. Proceedings: "Atti del Convegno Italy-Cina: An ancient cultural heritage and the challenge for future development", ISBN 978-88-6155-515-0, 2013.

H.D. Gao, P. Bin, etc. (2008). Some Considerations on Ecological Compensation Implementation in Beijing Mountainous. (in Chinese) Forest Resources Journal, DOI: CNKI: SUN: LYZY.0.2008-05-006, May 2008.