THE EFFECT OF HUMAN CAPITAL ON INCOME IN THAILAND: A SPATIAL EFFECT ANALYSIS

BY

MR. JETSON CHITDECHA

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER DEGREE OF ECONOMICS
(INTERNATIONAL PROGRAM)
FACULTY OF ECONOMICS
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2016

THE EFFECT OF HUMAN CAPITAL ON INCOME IN THAILAND: A SPATIAL EFFECT ANALYSIS

BY

MR. JETSON CHITDECHA

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER DEGREE OF ECONOMICS
(INTERNATIONAL PROGRAM)
FACULTY OF ECONOMICS
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2016

THAMMASAT UNIVERSITY
FACULTY OF ECONOMICS

THESIS

BY

MR. JETSON CHITDECHA

ENTITLED

THE EFFECT OF HUMAN CAPITAL ON INCOME IN THAILAND:
A SPATIAL EFFECT ANALYSIS

was approved as partial fulfillment of the requirements for
the degree of Master Degree of Economics (International Program)

on August 10, 2017

Chairman ______________________________
(Dr. Monthien Satimanon)

Member and Advisor ______________________________
(Asst.Prof. Dr. Nattapong Puttanapong)

Member ______________________________
(Dr. Pan Ananapibut)

Dean ______________________________
(Assoc.Prof. Dr. Chayun Tantivasadakarn)

Thesis Title	The Effect of Human Capital on Income in Thailand: A Spatial Effect Analysis
Author	Mr. Jetson Chitdecha
Degree	Master of Economics (International Program)
Major Field/Faculty/University	Economics Faculty of Economics Thammasat University
Thesis Advisor	Assistant Professor Dr. Nattapong Puttanapong
Academic Years	2016

ABSTRACT

This paper aims to analyze the effect of human capital on income in Thailand with the inclusion of spatial influence by applying three spatial econometric models based on a Mincer-type equation to Thailand's Labor Force Survey (LFS) for the years 2001 to 2015. Specifically, this study applies three spatial econometric models, namely, the spatial lag model, the spatial error model, and the spatial durbin model, to correct the problem of omitted variables representing the influences caused by spatial factors. The statistically significant results indicate that the wage determination mechanism in Thailand is compatible with Mincer's theory. In addition, all three spatial econometric models reveal an inter-province influence on wage determination, thereby indicating that the level of labor wage in a particular province is induced by factors within the province and its neighbors that are not considered in the examined model. These findings highlight the importance of developing new policies that focus on mitigating the spatial inequality of wage determination in Thailand, such as promoting multicenter development by taking advantage of the spatial effect from super cluster development. This study also provides directions for future research on the region-specific influences on the wage premium in Bangkok and other main provinces in Thailand.

Keywords: Return on human capital, Labor income, Education, Spatial analysis, Spatial econometrics.

ACKNOWLEDGEMENTS

This thesis has been done with the support of many various precious persons and without them, this important achievement will never happen. I would like to express my sincere gratitude to my advisor Asst. Prof. Dr. Nattapong Puttanapong for the wholehearted support, kind advices and suggestions to improve this thesis and make it more concrete. I would like to thank my committees Dr. Monthien Satimanon and Dr. Pan Ananapibut for their generous suggestions as well as deep economics knowledge which they always share during doing the thesis.

I also would like to thank my warm and happy family, especially my parents, for understanding and encouragement everything I do, and supporting everything I need throughout the course. In addition, I would like to thank all of professors at Faculty of Economics, Prince of Songkla University, and Thammasat University for their kind advice and suggestion both in the past and present.

I would like to thank faculty members of the graduate study program, The Faculty of Economics Thammasat University for information schedule, and all supports which they always give me.

I also thank all precious friends who have been through thick and thin together for providing a contribution to this thesis in a variety of ways and they always share the good idea and enjoyable experiences. I would also thank my best friend, Arinchaya Ritsorakrai, who always supports me everything, and shares nice ideas and enjoyable time.

I always enjoy every day in the university. Memorable days in this place have passed quickly. Thank you to the study room for the workplace, thanks to the Chao Phraya River for the great views and thanks to delicious foods around Tha Phra Chan and Tha Wang Lang which made me happy every day.

I would like to dedicate this thesis to all of them, but all errors are mine.

Mr. Jetson Chitdecha

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

This thesis examines the effect of human capital on the geographical labor income in Thailand while considering the influence of geographical location and the correlated neighboring provinces within a specified space. People need to invest in human capital to receive high returns in the future. However, the return is not only determined by the economic value of a labor's skill set but also depends on geographical economic conditions. Thus, a model that considers the interaction among different provinces within a specified space must be applied to capture a spatial autocorrelation or neighboring effects because the data observations are not truly independent.

1.1 Motivation

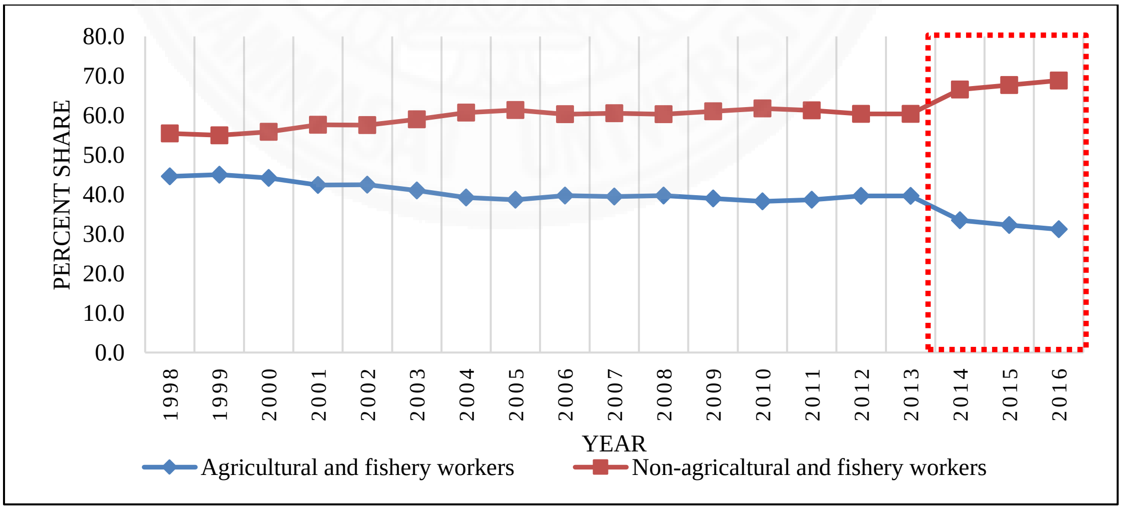
Thailand's economic structure has shifted from agriculture to manufacturing and services as indicated in Krongkaew, Chamnivickorn, and Nitithanprapas (2006), who studied the economic growth, employment, and poverty in the country. Krongkaew and Kakwani (2003) show that the sectoral share of employment in agriculture has decreased from 82.3% in 1960 to 48.8% in 2000. These findings are consistent with the results of the Thai Labor Force Survey in Figure 1.1, which shows the proportion of agricultural (and fishery) workers and non-agricultural workers in Thailand from 1998 to 2000. Figure 1.1 clearly illustrates that the number of workers in the non-agricultural sector has increased by more than 50% since 1998[1]. The manufacturing and service sector demands high-skill and educated workers more than low-educated workers. Therefore, those labors with limited knowledge or working skill are deterred from securing jobs in the higher-income sector. However, the informal sector plays a dominant role in the Thai labor market, with every region in Thailand,

[1] Since 2014, the survey framework was modified based on the population and housing census and the household socio-economic survey.

except Central Thailand and Bangkok, considering the informal sector as a primary source of employment. Figure 1.2 shows that the informal sector accounts for more than 50% of the labor force in these regions. Approximately 80% of the labor force in the northeastern region is working in the informal sector, while the proportion of informal workers in the northern and southern regions is just the same as that in the whole kingdom. The informal sector is generally characterized by small-scale activities that cannot create high-value added output in manufacturing because of the use of very simple production technologies. However, the National Statistical Office of Thailand defines informal workers as those who are neither protected nor have social security in their work[2]. Most informal sectors in Thailand are in the agricultural sector and mostly comprise low-educated workers. The variances in the labor market structure across each region highlights an important geographical influence on the labor market (Lathapipat & Chucherd, 2013). The geographical dimension may also affect the average labor income in each region and province. Therefore, the rate of returns to human capital must be investigated in light of the geographical interactions in Thailand by using the spatial analysis technique.

Figure 1.1

Sectoral share of employment

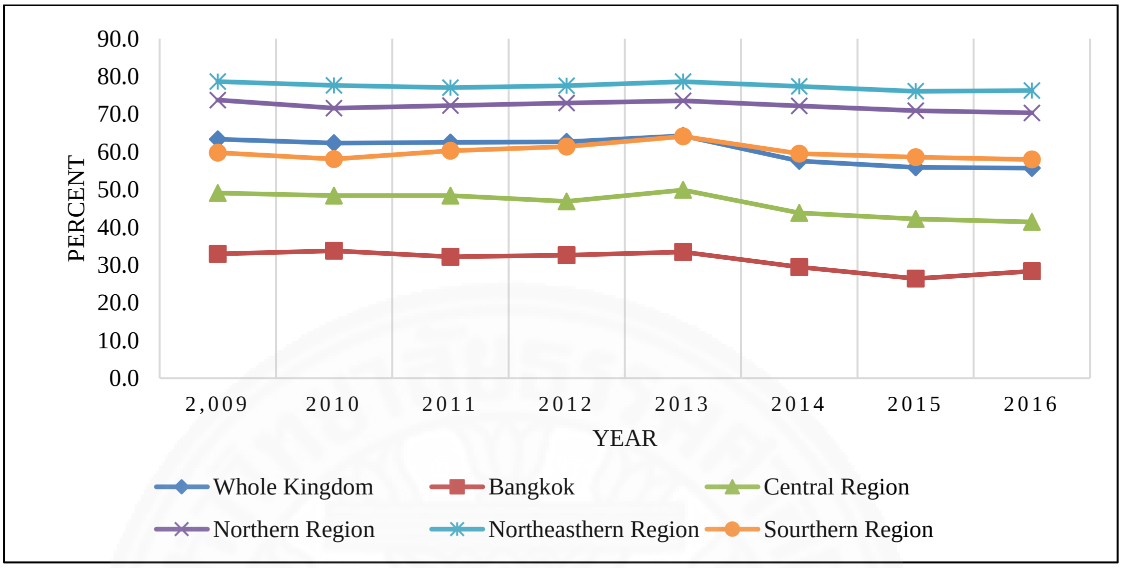


Source: Author's calculation from Labor Force Survey based on Bank of Thailand

[2] Workers in the formal sector are protected and have social security in their work, while workers in the informal sector do not have such protection.

Figure 1.2

Percent share of informal sector in each region



Source: Authors calculation from Informal Employment Survey

A labor with more schooling years is expected to generate a higher income in the future. Although some people have the same level of education, experience and other characteristics, they may not receive the same rate of return. At the global scale, labor income and return on education vary across each landmass, country, region, and state where the labors are working. This pattern can be observed in Thailand as well. The estimated return on education model shows that the labor income distribution by level of education depends on the region (Tangtipongkul, 2015). Income distribution also reflects the equality of a socio-economic situation and the success of policy implementation.

The Office of the National Economics and Social Development Board and the National Statistical Office of Thailand both show some geographical differences in income as the labors from various regions have different average labor income and average schooling years. Therefore, income and return on education must not be explained by only the traditional reason because they may be affected by the geographical dimension as well. For instance, high-income families reside in neighborhoods that do not have any middle- and low-income families because living together with other high-income families will encourage the development of better social services, social capital, school quality, and green space, all of which affect the quality of their lives (Reardon & Bischoff, 2011).

Moreover, if the geographical location and neighbor set are studied in detail, then the influence of these factors on the labor income distribution in Thailand can be determined. However, the spatial effect on the return on human capital at the provincial level has never been investigated in Thailand.

1.2 Significance of the study

At a disaggregated level, the relationship between spatial segregation and income segregation may be linked via the housing market. Compared with middle- and low-income families, high-income families have better opportunities and a higher purchasing power to reside in preferable locations and neighborhoods that can provide them with a positive externality and quality of life, which in turn can increase their income further in the future (Watson, 2009).

The relationship among close proximity neighbors may have been omitted in the traditional model, and ignoring such relationship can lead to biased estimations because this omitted spatial effect is correlated with the error term in the model. For example, the average labor income in each province is affected by the differences in climate, soil quality, attractions, and industrial estate across neighborhoods. Therefore, the model must consider the effect of geographical decomposition to generate accurate results. The educational return in Thailand has been investigated in several studies, including Hawley (2003), Hawley (2004), Mehta, Felipe, Quising, and Camingue (2013), Srinang (2014), Tangtipongkul (2015), and Warunsiri and McNown (2010), but these studies have never considered the geographical influence of proximity in their models.

Using a spatial econometric method can help researchers accurately calculate the estimated coefficient of return on human capital and investigate the spatial effect from the neighbors. Therefore, geographical effects must be considered when examining the actual rate of financial return on human capital in the provinces across Thailand.

1.3 Statement of the problem and research question

In general, an increase in education level generally increases the labor income. However, the return on human capital differs across each region in Thailand. At the disaggregate scale, the geographical location of a workplace may affect labor income because of the spatial effect or spatial externality from neighbors. Thus, the traditional ordinary least squares (OLS) model cannot estimate the actual return on education because of the omitted spatial dimension.

The previous studies and the traditional model are still limited to represent the influence of the spatial dimension. Therefore, it is interesting to analyze that how the labor income is influenced by the spatial dimension in light of the changes in human capital and how the income determinants in the spatial model change from the traditional model.

This thesis measures the effect of human capital, such as education, by using a spatial econometric model, a qualitative econometric technique, and by combining data from the Labor Force Survey and a geographic information system (GIS). To understand further the distribution and concentration of labor income in each province of Thailand, the results are plotted on a map of Thailand to show the geographical effects on wage differentiation.

1.4 Objectives of the study

1. To investigate whether the geographical dimension does matter in measuring the effect of human capital change on labor income.

2. To analyze the spatial correlation of provincial labor income.

3. To determine the effects of spatial dimension on labor income and to compare the spatial econometric model with the traditional model.

1.5 Scope of the study

To measure the geographical decomposition effect in the spatial estimated model, this thesis focuses only on the effect of education on labor income in Thailand by using Thailand Labor Force Surveys data from the 2001 to 2005 published by the National Statistical Office of Thailand

The provincial boundary in Thailand is changed by adding Bueng-Kan Province, which is established by consolidating some districts from Nong Khai Province into a new province since 2011. However, the data for 2012 are only collected from 77 provinces, thereby preventing the researchers from comparing inequitable data across years. In addition, a polygon map for these 77 provinces is unavailable. In this case, only 76 provinces are used as a benchmark. The data from Bueng-Kan Province after 2011 are combined with those from Nong Khai Province.

1.6 Organization of the study

This thesis is organized as follows.

Chapter 1 presents the introduction, motivation, significance, objectives, and scope of the study.

Chapter 2 reviews the literature on return on human capital and spatial income effect as well as presents the relevant geographical theories and a theoretical framework for education in Thailand.

Chapter 3 describes the data description employed in this paper. The research methodology and the empirical econometric model are also discussed.

Chapter 4 presents the empirical econometric results of the regressions to answer the questions and achieve the objectives of this thesis.

Chapter 5 concludes and summarizes the thesis as well as presents policy recommendations, research limitations, and directions for future study.

CHAPTER 2
REVIEW OF LITERATURE

The effect of human capital accumulation can be explained by using human capital theory proposed by Gary S. Becker. Meanwhile, the return on education and the work experience of labors can be evaluated by using the Mincer equation proposed by Jacob Mincer. A spatial model is constructed to explain the geographical economic influence and spatial influence from neighboring provinces by synchronizing the needed data with those from GIS. The results can also be explained further by plotting them on the map of Thailand. This chapter reviews the literature on return on human capital and spatial influence as well as presents the related theoretical framework for Thai education and geographical theory.

2.1 Return on human capital

People invest in human capital, such as education, on the job training, and healthcare, to improve their quality of life. As another form of human capital, working experience can improve labor productivity, which in turn increases their income. The return on investment in human capital depends on the number of years spent in school and years of working experience. Apart from the number of years in school, the return on education also depends on the level and type of education (Becker, 1962).

Education is a type of human capital development, and most people hope that their investment in education will boost their future earnings. Many studies have evaluated the effect of education on income distribution at the global, regional, and country scales by using the Mincer-type wage equation as a standard model for measuring the impact of education, work experience, and other characteristics on the economic returns of the labor force (Mincer, 1974).

The Mincer-type model has been used in previous studies as a benchmark for analyzing the microdata on the return on investments in education. A study on return on schooling around the world which covers 131 countries including Thailand from 1970 to 2011 reveals the positive correlation of return on education in the stable Mincer basic model and working experience provides a positive relationship with the diminishing rate. In addition, the private returns to education increase along with level of education. Nevertheless, the rate of return on education in every level depends on the income group of a region or country. The estimated coefficient indicates that the rate of return in Thailand in 2009 and 2011 are 13.6 and 9.4 respectively (Montenegro, Montenegro, & Patrinos, 2014). Such estimated rates of return do not greatly differ from those obtained by Psacharopoulos and Patrinos (2004), who found that the rates of return on an additional year of education is 10.7%, 9.9%, and 11.5% in a middle-income country, an Asian country, and Thailand, respectively. In addition, these results are consistent with those of Hanushek, Schwerdt, Wiederhold, and Woessmann (2015), who studied the return on skill in 24 countries around the world, excluding Thailand, between August 2011 and March 2012 by using data from the Program for the International Assessment of Adult Competencies. They found that the return on human capital depends on a country's specification of its labor market and institutions as well.

Education and work experience in the European Union (EU), which comprise 14 members countries as of 2000, affect labor income in the same way as the worldwide scale. Data from the EU show that the average income differs across each country and region, with the countries in the Nordic region having the lowest average income. This result indicates that the private return on schooling increases along with education level (De la Fuente & Jimeno, 2005). By using data from the International Social Survey Program in 1995, Harmon, Oosterbeek, and Walker (2002) examined almost western and eastern European countries and found that a higher level of education can increase future income and has a greater impact on high-income households than on low-income ones. This phenomenon represents a complementarity between education and initial ability in the EU. However, both of the aforementioned studies observe a geographical influence in each country.

The pattern of return on schooling in Thailand follows that in EU and other countries. The result from the Mincer model, which are obtained by using quarterly data from the Thai Labor Force Survey from 2007 to 2011, indicate a positive rate of return on years of education and a positive with diminishing rate of return on working experience. Meanwhile, the labor income distribution in Thailand depends on regional influences. For instance, labors in Bangkok Metropolitan Area get the highest income, while labors in the northern region get the lowest income. In addition, the private return on university education for women is higher than that for men, and vocational education provides a higher return than general education at the secondary level. An additional year of schooling without controlling any variable generated 13.37% and 12.63% rates of return in 2008 and 2010, respectively (Tangtipongkul, 2015). Employing a pseudo-panel approach to examine data from 1986 to 2005 can prevent unobserved biases and reveal that urban labors receive a higher return than rural labors because the former is given more opportunities to secure a better job than the latter. The overall rate of return is in the rank between 14% and 16% depending on the pseudo estimating detail, and this rate is higher than that obtained by standard Mincer model (11.5%) (Warunsiri & McNown, 2010). All the above findings are consistent with those of Wannakrairoj (2013) and Srinang (2014), who use dummy of the level of education in their modeling instead of years of education. However, they still discovered some differences in the labor earnings across regions, thereby indicating a geographical influence in Thailand. Table 2.1 presents the rates of return on an additional year of schooling in Thailand as indicated in previous studies.

Table 2.1

Rate of return on education in Thailand

Study	Level of study	Rate of return (percent)[1]
Montenegro et al. (2014)	Around the world including Thailand	9.4 % (Thailand, 2011) 13.6% (Thailand, 2009)
Psacharopoulos and Patrinos (2004)	Around the world including Thailand	9.9% (Asian country) 11.5% (Thailand)
Tangtipongkul (2015)	Country level, only Thailand	13.37% (2008) 12.63% (2010)
Warunsiri and McNown (2010)	Country level, only Thailand	11.5% (Standard approach) 14%–16% (Pseudo penal approach)

Source: Author's compilation

[1] Percent of income increasing from additional 1 year of schooling.

The above literature review represents similar results. Interestingly, these studies show that income distribution depends on geographical segregation (i.e., country or region). Therefore, the geographical location of a workplace can influence the income and return on education of the labor force even if all other characteristics are fixed. In this case, geography and adjacent area spillover effects should be considered in the adopted modeling approach.

2.2 Spatial influence

Spatial analysis has been initially developed to understand the pattern of spatial expression in respect of geometry, statistics, and mathematics. The spatial analysis results can be represented by plotting them on a map and by analyzing other geographical properties, such as nearest neighbor analysis etc. This technique can be applied to many areas of formal knowledge, including economics and econometrics, which examine financial returns on investment. In this case, the spatial econometric analysis in this paper focuses on spatial dependence, which is a key issue that can reduce the explanatory power of the traditional model because the data observations are not truly independent. Normally, cross-sectional data analysis is known for several challenges, such as the heteroscedasticity across different locations. As another important, the autocorrelation among the data observations indicates that the neighbor variable may also be treated as a dependent exogenous variable in the model. The spatial correlation implies a relationship between the locations of the economy, market, and socio-economic characteristics that depend on the distance between two point on locations or geographical contiguity (Miller & Wentz, 2003; Yano, 2000).

In the US, income inequality, which is measured by using the Gini index from the 100 largest metropolitan areas in 1970 to 2000, has been identified as a driver of wide income and spatial segregation. However, the segregation of affluence is greater than the segregation of poverty. The poor do not live together with other low-income families, while high-income households live with equally wealthy families to gain access to high-quality public goods and services, large green spaces, and high-paying jobs in the area (Reardon & Bischoff, 2011). These phenomena can be attributed to the ability and purchasing power of high-income households that allow them to live in areas with many available public goods and services as well as excellent healthcare, environment, and education, such as metropolitan areas. All of these elements may further boost their income in the future. Naturally, the house prices in these areas are very high, thereby indicating that spatial segregation is linked to the mean income distribution in these areas via the housing market. Segregation will affect education and the labor market, and then subsequently influence the distribution and segregation of income (Watson, 2009).

In general, the income differences across each area are affected by many factors that differ across each country, but education remains a key common factor. Previous studies on this sub-topic use the spatial econometric model in their estimations, but the dependent variable relies on the specified model.

The spatial econometric model has been employed to estimate the effect of average income on the income inequality in Austria by using the tax dataset of all Austrian wage earners from 1996 to 2010; apart from revealing spatial dependence at different geographical levels, the spatial econometric model shows that an additional year of schooling, number of years working on a part-time job, and labor earnings all increase the earning inequality in Austria These results are consistent with that of Rodríguez-Pose and Tselios (2009), who used European Community Household Panel data from 1995 to 2000 to examine the effects of per capita income and educational inequality on income inequality in EU. Furthermore, By using data from 1949 to 1998, a study in China found that spatial decomposition influences the growth of income inequality across cities and that the geography of the market is correlated with the per capita income in each location (Hering & Poncet, 2010).

The effect of regional income on dynamic convergence can be estimated by using a spatial model to consider the effect of geographical dependency. In Turkey, the spatial error and spatial lag models estimated by using data from the Turkish Statistical Institute between 1987 and 2001 can be used to improve the accuracy of the estimated parameters. The variation in convergence speed across different provinces as estimated by the spatial model is more effective and suitable than that estimated by the traditional model. These results also indicate a strong correlation between regional inequality and geographical clustering (Yildirim, Öcal, & Özyildirim, 2009). In the same way, Up and Donghyun (2015) found the geographical influence on regional income convergence from their study by using data from the U.S. Bureau of Economic Analysis from 1969 to 2009.

In addition, the empirical results from many studies on the spatial effect on different income levels empirically find that the sizes of cities and the characteristics of adjacent cities, including their economic conditions and economic concentrations, also affect labor income. In this case, the geographical location of a workplace can induce an income disparity in each area and encourage labor mobility as well as migration.

The evolution of urban development in the US has been examined by using a geographical model that focuses on income and facilities and uses the microdata of urban workers from 1980 to 2000. The model shows that workers earn a higher income if they work in larger cities where facilities are abundant. These findings are mostly attributed to the ease of stimulating productivity in these areas. Moreover, the income gap between cities is caused by the different composition of workforce in each city. Therefore, different labor income levels also depend on the city where the laborers are working. Furthermore, the migration of labors can be attributed to housing quality, transportation cost, and other facility-related factors (Kemeny & Storper, 2012). In addition, the spatial proximity impact of neighborhoods must also be considered to describe the economic condition in the host area. A study on Appalachia, a poor area in the US, from 1990 to 2000 reveals that the employment level, migration, and household income in a country are also determined by the performance of economic conditions in its neighboring counties. The empirical spatial models in this study has been estimated simultaneously on the basis of the Feasible Generalized Three-Stage Least Squares technique (Gebremariam, Gebremedhin, & Schaeffer, 2011). Similar to previous results, return on education and migration are influenced by the geographical dimension. A spatial durbin model based on the Mincer-type equation represents that the labor market in EU does not discriminate the return on schooling between migrant and non-migrant workers. However, the factors that influence return on education include the externality from the neighbor household or the neighbor region. The neighboring income produces a positive effect, but the neighboring schooling year produces a negative effect (Rodríguez-Pose & Tselios, 2010).

In conclusion, the geographical influence cannot be ignored because the correlation of the error term, which represents the spatial correlation variation across neighbors, will lead to a biased estimation model. In other words, if the geographical process is responsible for the geographical effect, then the spatial model is more efficient and suitable than the traditional model. Spatial modeling can also provide appropriate policy recommendations for the taxation and provision of public goods (Case, 1991).

In this chapter, the first topic suggests that labor income studies are based on the Mincer-type equation and shows that the return on human capital differs across each country and region. The second topic discusses the spatial influence or spatial component that accurately changes the value of the estimated parameters. Both of these topics reveal some gaps in the literature, particularly the failure of previous studies to consider the influence of location and spatial proximity when evaluating the returns on investment in human capital in the model. Therefore, the income determinant of human capital must be studied along with the spatial component effect.

2.3 Theoretical framework

This section briefly discusses the theory of return on investment in human capital and the geographical theory in terms of how this theoretical framework reveals empirical phenomena by using compatible Labor Force Survey data. The theoretical framework begins by investigating the theory of human capital, return on education, and education system in Thailand before discussing the properties of spatial composition.

2.3.1 Background of human capital and Thai educational system

The human capital theory is proposed by Gary S. Becker and Jacob Mincer, human capital theory points out that the accumulation of human capital can increase a labor’s skill and consequently affect his/her labor productivity and income. However, human capital is inseparable from its owner, and the expected return differs across each person. People need to spend much time for investment, which may involve an opportunity cost, such as losing income by working full time, as well as a direct cost, such as paying tuition when investing in education. A huge part of human capital theory focuses on educational investment. Many studies show that being educated in schools and universities around the world can significantly increase one's personal income. Years of work experience also affect labor income but with a diminishing rate. However, people do not have the same socio-economic status and background that affect their income and access to educational opportunities. Those people with higher education levels often have a higher ability, a richer parent, and a higher educated parent, all of which can increase their income in the future. Gender, marital status, educational program, occupation, number of children in the family, and other household resources greatly affect labor income as well. Therefore, the socio-economic status and characteristics of labor must be controlled in modeling (Becker, 1962, 2008). Based on a report published by the World Bank, Brixi (2012) has represented that health science and engineering programs generate the highest returns on higher education in Thailand, while social science, business, and law programs generate very low returns. However, many students continue to graduate from these fields.

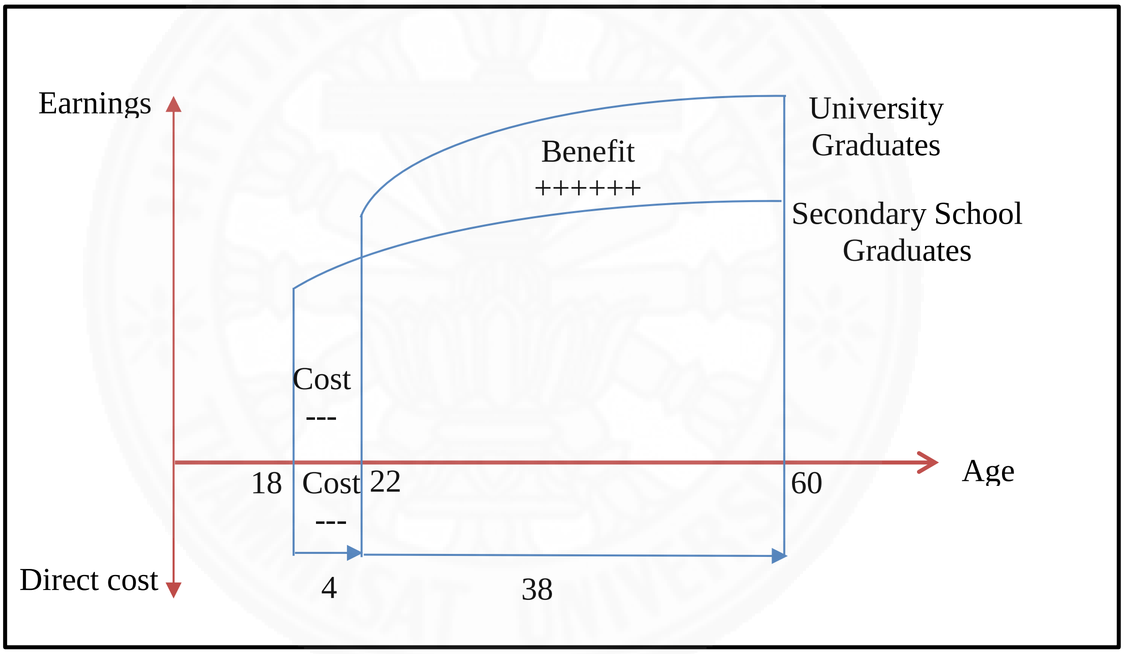
Education in Thailand has been divided into several levels by the National Education Act of 1999 (revised in 2002) and the Compulsory Education Act of 2002. Students begin in pre-primary or early childhood education, which is purely voluntary. They spend two to three years in pre-primary education to develop their cognitive and non-cognitive abilities, such as their mental, intellectual, and socio-emotional abilities, and to establish a foundation for their lifelong learning. After early childhood education, the National Education Act of 1999 requires children to enroll in primary and lower secondary education for six and three years, respectively, until they complete Grade 9 or turn 16 years old. Those students who study at this level develop and improve their general cognitive and necessary skills for working, such as basic literacy skills, numerical skills, communication skills, IT skills, and desirable behavior. After completing lower secondary school, students can take upper secondary education for three years during which they must choose between vocational or general education before proceeding to higher education. These students are also allowed to choose their fields[2] of interest and learn a specific skillset to prepare themselves for the labor market.

[2] The major is separated into the science and mathematics program, mathematics and arts program, and arts program.

Higher level education includes diploma, post-secondary, undergraduate, and graduate programs. Students at the diploma and post-secondary level aim to develop and improve their semi-cognitive skills and entrepreneurial knowledge within three years. The undergraduate and graduate programs, which require four years and two to five years[3], respectively, involve intensive academic studies that encourage students to improve their skills by applying theory and other knowledge in practice that can contribute to the country's development.

Figure 2.1

Stylized age-earning profiles



Source: Author's modification from Psacharopoulos (2006), Figure 2 (p.116)

Generally, investment in education positively and significantly impacts either the private or public sector. However, despite explaining the direction of the return on investment, the theory cannot clearly illustrate the magnitude of returns at each education level. As mentioned earlier, the cost and benefit of investment in human capital can be represented by the age–earning profiles of education. For example, when

[3] A master's student must spend approximately two years, while a doctoral student must spend approximately five years when taking the course.

secondary school and university graduates do not work during their study at the university, their indirect cost of education is equal to that of secondary school graduates with the same characteristics gain from working. These students also face some direct cost in terms of tuition and incidental expenses for studying. After graduation, university graduates receive a higher return than secondary school graduates, and such income advantage will continue throughout their lifetime. Therefore, the return of a higher level of education must not be less than the previous level as shown in Figure 2.1, which is adapted from Psacharopoulos (2006). This phenomenon is known as the screening hypothesis, in which the labor tries to signal something and the employer tries to screen that signal. A higher education level may act as an efficient filter for identifying and screening effective workers.

2.3.2 Geographical theory

Before going further, an issue about the spatial analysis used in the estimation process should be interpreted by using the statistical and mathematics theory of spatial expression. A spatial analysis can be performed by using a formal technique that applies geographical properties to analyze geographic data. Generally, the location and distribution of human activities are determined by an environment of economic which refers to the spatial dimension or the major role of the proximity that may make the external impact in econometrical analysis. The spatial dimension of the differences between the center and periphery can be explained by many spatial theories.

The geographical theories of innovation diffusion, economic development, people interaction, and transportation have been developed to explain the pattern of spatial configurations with respect to a specific area to describe the behavior and activity of an agent in such area because the spatial effect may change the characteristics of people, firms, and markets. Therefore, analyzing only a single economic environment is considered insufficient because other economic environments from an adjacent neighbor may impact the host city. Such impact, which affects both labor and economic activity, may be political, cultural, legal, or involve an abundance of capital, resources, and information, and etc. The influence of these spatial implications on human and economic activities can also be explained by many economic theories.

First, agricultural land use theory describes the productivity at the center, which covers the best land for cultivation and provides the highest returns. However, the expansion of agricultural land use to adjacent areas and higher-order adjacent areas can reduce the productivity in the agricultural sector. Therefore, those labors who are working near the center will earn a higher income than the others.

Second, industrial location theory, which relates to the decision-making process, focuses on raw materials and market seeking, both of which affect the geographical location of an industry that seeks to reach its profit maximization target. Geographical location also affects the income, employment, and migration of workers in both the host and neighboring cities. Those workers who are working in the industrial estate city are also expected to earn a higher income than the others.

Third, geographical location theory discusses why some locations are chosen more often than the others. Some cities are more popular than the others because of their allocation of facilities and infrastructures, such as schools, hospitals, parks, and public utility systems. These public goods and services can provide workers with a high income and excellent quality of life for themselves and their families or provide the necessary production facilities to those industries that depend on the location and planning of the city.

Fourth, central place theory explains the settlement of humans in an urban system. Naturally, some cities are chosen as the main nodes of manufacturing, administration, and transportation, while the other cities are chosen merely as secondary nodes. Therefore, the people, energy, materials, and other information flow to the location that is placed at the center (Golledge, 1996).

On the other hand, the innovation spillover from geographical proximity is also important to the neighboring provinces as well. The new knowledge can be shared among cities within an appropriate distance from one another. This knowledge can also be easily transmitted to nearby areas without the need for social interaction. However, social interaction, which involves culture, norm, and socio-cultural values, affects the characteristics of the labor force and the behavior of workers. An agglomeration of the labor force also contributes to the knowledge and infrastructure in neighborhoods. Thus, the average labor income between two areas increase simultaneously according to the theory of spatial clustering on the surface (Gust-Bardon, 2012).

CHAPTER 3
RESEARCH METHODOLOGY

The Labor Force Survey data from the National Statistical Office of Thailand are combined with the data from GIS to measure the return on education by using a quantitative econometric method. This method is extended step-by-step for the standard Mincer model before applying the empirical econometric function of the spatial Mincer model. The estimation and model specification processes are described in detail in this chapter. After clarifying the econometric approaches, the limitations of the dataset are described along with the data generation process. The last sub-section defines the explanatory and dependent variables as well as reveals summary statistic of the variable set in the individual- and provincial-level models.

3.1 Econometric approach

Generally, the relationship between investment in human capital and personal labor income can be estimated by using microeconomic data. The traditional OLS model is used to estimate the impact of human capital on income distribution as a linear function called the Mincer equation. The regression reveals the economic return from years of completed schooling and work experience by controlling the individual characteristics, socio-economic status, and occupation of workers. The main idea of estimations is based on Mincer (1958), while the regressions are based on the Mincer income function proposed in Mincer (1974). Nevertheless, the empirical econometric models are represented by different specifications in order to assess the robustness of the spatial effect from geographical segregation that is linked to both spatial interaction and spatial structure.

This thesis uses two models that adopt two divergent regression analysis techniques, namely, 1) the standard model that does not apply spatial analysis at the individual and provincial levels, and 2) the spatial model that applies spatial analysis at both levels. The characteristics of these models will be explained step by step in the following sub-sections.

3.1.1 Standard Mincer model

The standard Mincer model, which is a semi-log function of return on education proposed by Mincer (1974), is used as the benchmark model in this thesis to compare the estimated parameter of return on schooling with the other models. The main structure of this function represents that the natural logarithm of labor income is a function of years of schooling, years of work experience, and years of work experience squared as shown in equation (1):

$$\ln y = \ln y0 + \gamma S + \beta_1 X + \beta_2 X^2 + u \tag{1}$$

where ln y denotes the vector of the national logarithm of the worker's monthly income, the constant term ln y0 denotes the average personal income of a worker without any work experience and schooling, S is a vector of schooling years, X is a vector of years of work experience, and X^2 is a vector of work experience squared. However, it can be rewritten as a general matrix form of the extended income function by adding the individual characteristic matrix as shown in equation (2):

$$\ln y = \ln y0 + \gamma S + \beta_1 X + \beta_2 X^2 + O\delta + u \tag{2}$$

where O is the matrix of controlled variables such as individual characteristics and socio-economic status. In general, the error term of income regression, u, is a random disturbance term that is expected to be white noise. The error term of the cross-sectional unit i is not correlated with cross-sectional unit j. However, in this case, it is expected to be spatially correlated with others because of the influence of the geographical dimension. Location may be autocorrelated in the model because the covariation of properties within a geographic space at proximal locations appears to be correlated and the data observations are not truly independent. In this case, the model is affected not only by regressors, such as the standard Mincer model, but also by the location of the workplace and neighbor provinces. The standard Mincer model cannot sufficiently estimate the return on human capital when the geographical dimension is considered. In addition, a standard OLS method cannot produce accurate estimations because the standard linear regression model provides an inconsistent estimator as a result of omitted spatial variable bias (Moser & Schnetzer, 2014). In this case, the spatial autocorrelation or spatial dependence with proximal areas must be solved by using a spatial econometric method. The efficiency and accuracy of the estimators must also be improved by applying the Mincer equation along with spatial econometric methods in the spatial model.

3.1.2 Spatial model

The term "spatial model" relates to the use of GIS data in an analytical process in order to describe the basic properties of a set of spatial features. Different estimation strategies allow the use of a weight matrix in modeling. The spatial weight matrix includes the inverse distance criterion between provinces or the geographical contiguity status of the neighbor set that represents the geographical dimension that may affect a worker's income. This type of model is called a spatial model. However, the structure of this model is still based on the Mincer-type equation. Adding a spatial weight matrix can improve the accuracy of the model because the locations of the worker's provincial residence and neighbor provinces can affect his/her income. The return on education and experience may either increase or decrease after adding a spatial cause. Incidentally, the addition of a spatial weight matrix may also affect the significance level of the estimators, which in turn may change the results from insignificant to significant or vice versa in some explanatory variables. However, in general, the spatial effect from the neighbors decline by a longer distance between the cities. Certainly, those neighbor provinces with a direct border can produce a greater effect than a higher-order neighbor.

The weight matrix represents a spatial structure in the dataset. Each element w_{ij} denotes a weight to illustrate a spatial relationship among different locations. Thus, w_{ij} is a spatial autocorrelation statistic that can be defined in many ways as described in section 3.1.5. The selected weight matrix under the specific definition will be used for both the spatial lag, spatial error, and spatial durbin models. Each of these spatial models uses the same spatial weight matrix, except in the case where the definition of the spatial weight matrix is changed by using other definitions to analyze the relations between new sets of neighbors.

In general, spatial dependence can appear in the traditional linear regression model in two distinct ways. Specifically, spatial dependence appears as an additional independent variable in the form of a spatially lagged dependent variable and as an additional independent variable in the error term (Anselin, 2007) that can be developed to the spatial durbin model (Anselin & Rey, 2014). Thus, the spatial model must be estimated in three approaches which are spatial lag model (SLM), spatial error model (SEM) and spatial durbin model (SDM).

3.1.2.1 Spatial lag model

The spatial lag model denotes how labor income is affected by that of workers in neighbor provinces because income may produce a spillover effect across provinces via interprovincial trade, social capital, financial spillover, and technological and information externalities. The autocorrelation model can be written in its general form as follows:

$$y = \rho Wy + X\beta + u \tag{3}$$

In the spatial lag model, both the independent and dependent variables, y, are the same vector, but the explanatory variable vector y is multiplied by the spatial weight matrix W. Thus, Wy on the right-hand side of the equation represents an additional spatially lagged dependent variable that reflects the effect of the neighboring endogenous variable, such as how a worker's average income is affected by that of a worker in a neighboring location. For example, the spatial weight matrix for three cities in space (the definitions and the meaning of each element are presented in section 3.1.5) and the operation of Wy are presented as follows:

$$Wy = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \times \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} (w_{11}y_1 + w_{12}y_2 + w_{13}y_3) \\ (w_{21}y_1 + w_{22}y_2 + w_{23}y_3) \\ (w_{31}y_1 + w_{32}y_2 + w_{33}y_3) \end{bmatrix} = \begin{bmatrix} (w_{12}y_2 + w_{13}y_3) \\ (w_{21}y_1 + w_{23}y_3) \\ (w_{31}y_1 + w_{32}y_2) \end{bmatrix}$$

where w_{11}, w_{22}, and w_{33} are equal to zero (see section 3.2.5 for details). Therefore, Wy can be included as another vector of the independent variable. Nevertheless, the spatial lag model in equation (3) can be applied to the expanded Mincer function in equation (2), and the combined model can be represented as follows:

$$\ln y = \rho W\ ln\ y + \ln y0 + \gamma S + \beta_1 X + \beta_2 X^2 + O\delta + u \tag{4}$$

where $u \sim N(0, \sigma^2)$, W is the spatial weight matrix size N*N that represents the spatial relations between each pair of locations or provinces in the dataset of Thailand as mentioned earlier, ρ is a spatial autocorrelation coefficient, and the other variables follow the previously mentioned specifications. However, to avoid explosive processes, the estimated spatial coefficient values must be restricted between –1 and 1 ($-1<\rho<1$), because the neighboring factors should not produce a greater effect than the internal factors.

3.1.2.2 Spatial error model

A spatial dependence is working as an omitted variable in the spatial error model as well. The spatial error parameter represents the intensity of spatial correlation which is a nuisance dependence in the use of the spatial data. The spatial error model can be written in its general form as follows:

$$y = X\beta + u \quad ; \quad u = \lambda W u + \varepsilon \tag{5}$$

where the disturbance term u is a function of the neighbor's disturbance. The endogenous u and exogenous u are the same vector, but the exogenous u is multiplied by the spatial matrix W to define the disturbance effect from the neighbor provinces under the same concept as the previous model. Therefore, Wu is a spatial lagged variable. The spatial error model uses the same weight matrix as the spatial lag model. To estimate return on education when spatial error is considered, the spatial error model must be applied to the Mincer function. Therefore, equations (2) and (5) are combined to form equation (6) as follows:

$$\ln y = \ln y0 + \gamma S + \beta_1 X + \beta_2 X^2 + O\delta + u \, ; \quad u = \lambda W u + \varepsilon \tag{6}$$

where $\varepsilon \sim N(0, \sigma^2)$, λ is the spatial error coefficient, and the other parameters follow the previously mentioned specifications. The traditional OLS is no longer efficient and the standard errors are biased.

In general, the spatial lag and spatial error models can be written as nested models in their general form as follows:

$$y = \rho W y + X\beta + u \quad ; \quad u = \lambda W u + \varepsilon \tag{7}$$

The reduced form of equation (7), which is obtained by subtracting matrix ρW from both sides, can be rewritten as follows:

$$y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} u \quad ; \quad u = \lambda W u + \varepsilon \tag{8}$$

The combined Mincer model and spatial model can be written as follows:

$$\ln y = (I - \rho W)^{-1} [\ln y0 + \gamma S + \beta_1 X + \beta_2 X^2 + O\delta + (I - \lambda W)^{-1} \varepsilon] \tag{9}$$

The above model represents a spatial lag model when λ = 0 and represents a spatial error model when ρ = 0. Matrix I denotes the identity matrix, while the other parameters follow the previously mentioned specifications.

3.1.2.3 Spatial durbin model

The geographical impact exerts its influence not only through a spatial lagged dependent variable but also through spatial lagged explanatory variables (WX). The spatial durbin model developed from the spatial error model (Anselin & Rey, 2014) denotes how labor income is influenced by labor income, education and other factors which are explanatory variables in those of the neighbor provinces.

The error process in equation (5) can be written as follows:

$$u = (I - \lambda W)^{-1} \varepsilon \tag{10}$$

Given that equation (10) can be substituted into the original equation, the following constraint is obtained:

$$y = X\beta + (I - \lambda W)^{-1} \varepsilon$$

By subtracting $(I - \lambda W)$ from both sides, the above equation can be written as follows:

$$(I - \lambda W) y = (I - \lambda W) X\beta + \varepsilon \tag{11}$$

Equation (11) is explicitly written out as

$$y - \lambda W y = X\beta - \lambda W X\beta + \varepsilon$$

or written as

$$y = \lambda Wy + X\beta - \lambda WX\beta + \varepsilon \quad (12)$$

The unconstrained form of equation (12) can be written in its general form as follows:

$$y = \theta_1 Wy + X\beta + WX\theta_2 + \varepsilon \quad (13)$$

while the reduced form of equation (13) can be written as

$$y = (I - \theta_1 W)^{-1} X\beta + (I - \theta_1 W)^{-1} WX\theta_2 + (I - \theta_1 W)^{-1}\varepsilon \quad (14)$$

where θ_1 and θ_2 represent the spatial lagged dependent and spatial lagged explanatory variables respectively.

To measure the geographical effect of labor income, the spatial durbin model must be applied by combining equations (2) and (13) as follows:

$$y = \theta_1 W \ln y + \ln y0 + (\gamma_1 S + \beta_1 X + \beta_2 X^2 + O\delta_1)$$

$$+W(\gamma_2 S + \beta_3 X + \beta_4 X^2 + O\delta_2) + \varepsilon \quad (15)$$

The reduced form of equation (15) can be written as

$$y = (I - \theta_1)^{-1}[\ln y0 + (\gamma_1 S + \beta_1 X + \beta_2 X^2 + O\delta_1)$$

$$+W(\gamma_2 S + \beta_3 X + \beta_4 X^2 + O\delta_2)] + (I - \theta_1)^{-1}\varepsilon \quad (16)$$

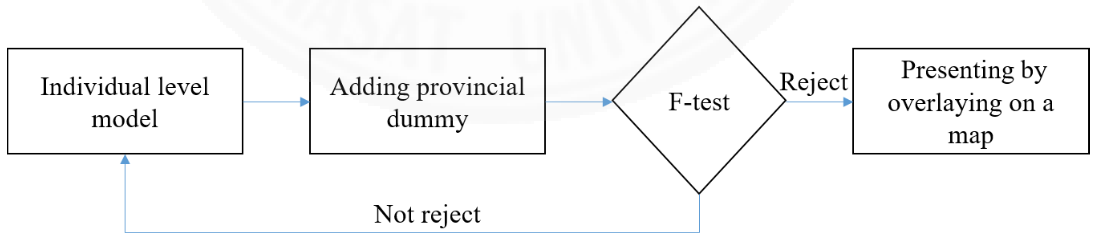
where $\varepsilon \sim N(0, \sigma^2)$, θ_1 is the spatial lagged dependent coefficient, and γ_2, β_3, β_4, and δ_2 are the spatial lagged explanatory coefficients. The other parameters follow the previously mentioned specifications.

In this thesis, the standard Mincer model will be estimated at either the individual or provincial level. At the individual level, the standard Mincer model will be estimated with a full set of control variables that includes provincial dummy variables. If the provincial dummy variables are significant, then they will be classified by quantile and mapped by color in order to reflect the geographical effect on labor income. Figure 3.1 illustrates the model estimation procedure at the individual level.

After explaining the labor wage difference across locations by using the traditional OLS model at the individual level, the standard macro-Mincer equation will be estimated at the provincial level to assess the spatial expression before estimating the spatial models. The error term of the traditional Mincer model is used to test the presence of a spatial effect in the model based on Moran's I index. The test is performed under the null hypothesis of no spatial auto correlation. The specification testing for spatial effect needs a significant Moran's I to confirm the existence of spatial impact before performing a spatial Lagrange multiplier (LM) test to test whether such impact represents as a spatial lag or a spatial error. The LM test can be separated into LM lag and LM error, and at least one of these tests must generate significant results to reveal the pattern of spatial dependency in the model. The robust LM test results must also be evaluated to confirm the spatial pattern if the results of both the LM lag and LM error tests are significant. Further spatial modeling is required when a spatial pattern in the model is clearly evident. Figure 3.2 illustrates the model estimation and specification testing procedures at the provincial level before estimating the spatial Mincer model. The spatial model highlights the importance of the location of concern and identifies the spatial effects in the estimated model.

Figure 3.1

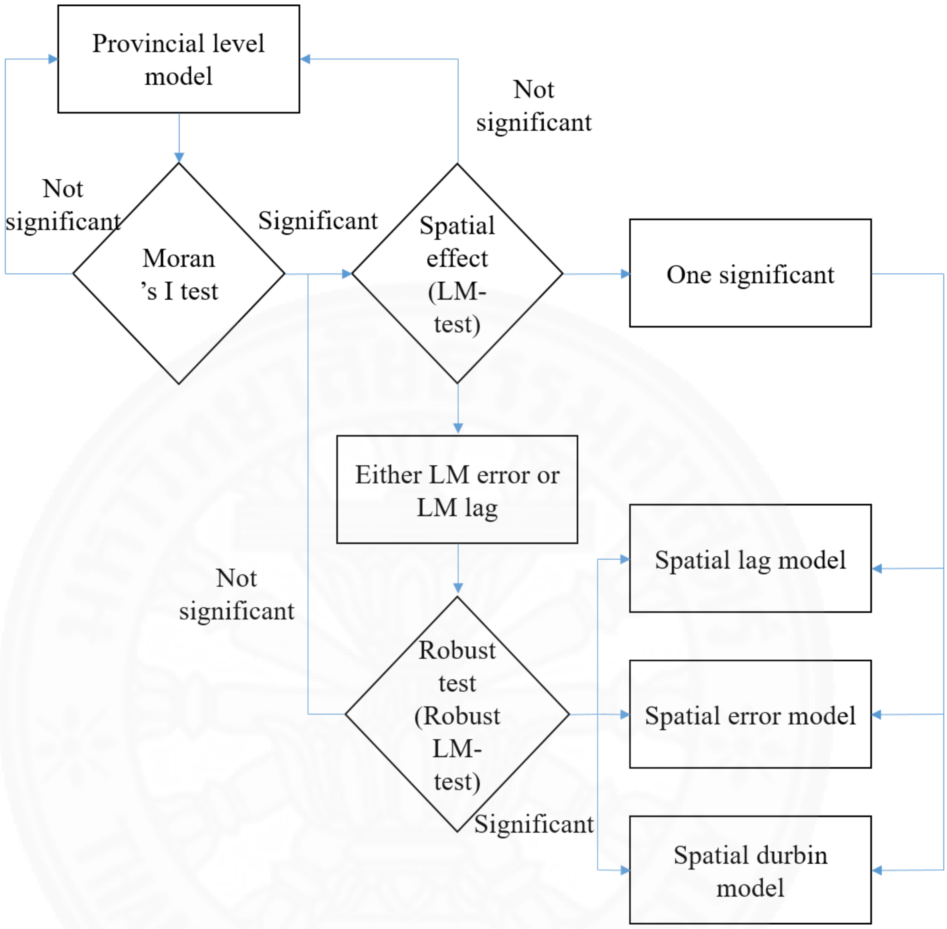
Model estimation procedure at the individual level



Source: Author's diagram

Figure 3.2

Model estimation procedure at the provincial level



Source: Author's diagram

3.1.3 Model specification tests

As a statistical test for spatial specification, Moran's I developed by Moran (1948) applies the regression residual to describe the spatial correlation of a location with its neighbor. This non-constructive or misspecification test is commonly used as a fundamental specification test. The structural form can be represented as follows:

$$I = \frac{e'We/S_0}{e'e/N} \tag{17}$$

where $S_0 = \sum_i \sum_j w_{ij}$ is the sum of the row-standardized weight matrix which value is equal to N. The weight matrix used in the model specification test is exactly the same as the weight matrix in the spatial model as shown in section 3.1.2.

Therefore, equation (17) can be rewritten as

$$I = \frac{e'We}{e'e} \tag{18}$$

Moran's I is tested under the null hypothesis of no spatial autocorrelation. However, rejecting this hypothesis does not mean occurring of the alternative spatial error; instead, such rejection characterizes the correlation among neighbor provinces in more than one space. Therefore, even if the Moran's I test generates significant results, performing LM tests remains necessary for the spatial regression, and all tests must provide consistent results.

The Moran's I test can also be used to measure the spatial autocorrelation of variables or the correlation among nearby locations within a certain space by using the same concept as the specification testing. The functional matrix form is changed as follows:

$$I = \frac{x'Wx}{x'x}$$

where x is an observed variable matrix. In addition, the local Moran's I can be used to represent a spatial correlation on the map and to measure the significance of the geographical correlation.

After estimating the standard regression model, the qualification of the spatial regression model must be tested. If the model is estimated by maximum likelihood (ML), then the result can be assessed by using three main methods, namely, the likelihood ratio (LR) test, Wald test, and Lagrange multiplier (LM) test. However, the first two approaches must be tested after estimating the spatial model. The presence of spatial dependence is generally tested by performing the LM test and Moran's I test because these approaches are based on the restricted model under the null hypothesis of no spatial lag and spatial error effect (Anselin, 2007; Anselin & Bera, 1998).

The formal null hypothesis in the case of spatial lag is H_0: $\rho = 0$ and the alternative is H_1: $\rho \neq 0$.

The LM test is solved from the first derivative of the log likelihood function as follows:

$$LM_{\rho} = \frac{d_{\rho}^{2}}{D} \sim x^{2}(1) \tag{19}$$

where e is the vector of the residual, $d_{\rho} = \frac{e'Wy}{e'e/N}$,

and $D = \frac{(WX\hat{\beta})'[I - X(X'X)^{-1}X'](WX\hat{\beta})}{e'e/N}$+T

The denominator term represents the sum of the square residual in the spatial lagged predicted value $(WX\hat{\beta})$, and the second term is a trace expression for the spatial effects $T = tr(W^2 + W'W)$.

The formal null-hypothesis for the spatial error test is H_0: $\lambda = 0$, and the alternative is H_1: $\lambda \neq 0$.

The LM test for the spatial error is solved from the first derivative of the log likelihood function as follows:

$$LM_{\lambda} = \frac{d_{\lambda}^{2}}{T} \sim x^{2}(1) \tag{20}$$

where e is a vector of the residual, $d_{\lambda} = \frac{e'We}{e'e/N}$, and We is its spatial error term of the neighbor areas.

If the spatial lag and spatial error specification tests both generate statistically significant results, the robust LM test must be performed to assess the robustness of the alternative spatial effects. In general, LM_{ρ} is sensitive to the presence of spatial lagged, while LM_{λ} has power against spatial lagged. The LM test may change the results from significant to insignificant if the original result is incorrect because the robustification is smaller than the original test results and the reduction is large enough (Anselin & Rey, 2014).

The formal robust statistic tests for spatial lagged and spatial error are represented as

$$LM_{\rho}^{*} = \frac{(d_{\rho} - d_{\lambda})^2}{(D - T)} \sim x^2(1) \qquad (21)$$

and

$$LM_{\lambda}^{*} = \frac{\left(d_{\lambda} - TD^{-1}d_{\rho}\right)^2}{[T(1 - TD)]} \sim x^2(1) \qquad (22)$$

Furthermore, the spatial effect can be tested by the joint null hypothesis of no spatial lagged and no spatial error in the model. In this case, rejecting the null hypothesis becomes a necessary condition for the spatial effect model; therefore, at least one of the spatial parameters must not be equal to zero.

H_0: $\rho = \lambda = 0$

The alternative hypothesis is H_1: $\lambda \neq 0$ or H_1: $\rho \neq 0$.

However, the joint null hypothesis is not equal to the sum of the two LM tests, $LM_{\rho\lambda} \neq LM_{\rho} + LM_{\lambda}$, but rather takes a much complex form as follows:

$$LM_{\rho\lambda} = \frac{d_{\lambda}^2}{T} + \frac{(d_{\lambda} - d_{\rho})^2}{(D - T)} \sim x^2(2) \qquad (23)$$

Rejecting the null hypothesis means that the model has at least one type of spatial correlation. However, joint null hypothesis testing is rarely used in practice. Moreover, each specification testing method must provide consistent results.

After describing the model specification test for spatial characteristics, the estimation method is explained thoroughly and the choice of estimation is proposed.

3.1.4 Estimation

For the estimation process, Anselin (2007) and Anselin and Rey (2014) describe the spatial lag and spatial error model that need to be estimated by maximum likelihood estimation (ML) method. This method takes the following general form:

$$\hat{\beta}_{ML} = (X_s' X_s)^{-1}(X_s' y_s) \tag{24}$$

where $X_s = (\mathrm{X} - \lambda\mathrm{WX})$ and $y_s = (\mathrm{y} - \lambda\mathrm{Wy})$ in both the spatial error and spatial durbin model. And $X_s = X$ and $y_s = (\mathrm{y} - \rho\mathrm{Wy})$ in the spatial lag model

In addition, the joint log-likelihood for the multivariate normal distribution of labor income in the spatial regression model is not equal to the sum of log-likelihood of the individual observations in the classic regression model. Therefore, the estimated values of β and σ^2 in the general forms of the spatial error and spatial durbin models are computed by the familiar maximum likelihood based on equation (5) as follows:

$$\hat{\beta}_{ML} = [(\mathrm{X} - \lambda\mathrm{WX})'\,(\mathrm{X} - \lambda\mathrm{WX})]^{-1}(\mathrm{X} - \lambda\mathrm{WX})'\,(\mathrm{y} - \lambda\mathrm{Wy}) \tag{25}$$

and

$$\hat{\sigma}^2_{ML} = [(\nu - \lambda\mathrm{W}\nu)'\,(\nu - \lambda\mathrm{W}\nu)]/N \tag{26}$$

where $\nu = \mathrm{y} - X\hat{\beta}_{ML}$.

The estimated values of β and σ^2 in the general form of the spatial lag model are computed as follows based on equation (3):

$$\hat{\beta}_{ML} = (X'X)^{-1}X'(\mathrm{y} - \rho\mathrm{Wy}) \tag{27}$$

or written as

$$\hat{\beta}_{ML} = \hat{\beta}_0 - \rho\hat{\beta}_1$$

and

$$\hat{\sigma}^2_{ML} = [(\nu_0 - \hat{\rho}\nu_L)'\,(\nu_0 - \hat{\rho}\nu_L)]/N \tag{28}$$

where $\nu_0 = \mathrm{y} - X\hat{\beta}_0,\ \nu_L = \mathrm{Wy} - X\hat{\beta}_L,\ \hat{\beta}_0 = (X'X)^{-1}X'y$
and $\hat{\beta}_1 = (X'X)^{-1}X'Wy$

The estimated parameters in the spatial model cannot be obtained by the traditional OLS method but rather by the maximization of the log-likelihood function. The weight matrix that will be explained in detail in the next section must be non-negative matrix to explain the spatial interaction across different provinces.

Although ML estimation is known for its consistency and asymptotic efficiency, this method is based on strong assumptions, such as the normality distribution and i.i.d. (independent and identically distribution) of the error term, thereby violating the assumptions will prevent some optimal properties from appearing. If strong assumptions are violated, then highly robust methods must be used. The generalized method of moments (GMM) is an alternative method that does not need any normality assumption yet allows the inclusion of additional endogenous explanatory variables in the model (Anselin & Rey, 2014; Lee, 2003).

The spatial weight least squares (SWLS) and feasible generalized least squares (FGLS) methods are applied for the spatial error and spatial durbin models. The estimation results still hold the ML, but the lambda is an inference lambda and the standard deviation is represented as follows:

$$\hat{\sigma}^2_{GMM} = [(\nu - \lambda W\nu)'\ (\nu - \lambda W\nu)]/N \tag{29}$$

where $\nu = y - X\hat{\beta}_{GMM}$

The results for the spatial lag model will be estimated by using the spatial two stage least squares technique (S2SLS), which uses a set of instrument variables (Q) to correct the endogeneity Wy. In this case, the spatial lagged explanatory variables (i.e., the explanatory variables of the neighbor) will be treated as instrument variables for the spatial lagged dependent variables (i.e., endogenous variables of the neighbor), including WX and W^2X. The instrument matrix is Q=[X, WX] or Q= [X, WX, W^2X] depending on the number of instrument variables in the model. The reduced form of equation (3) can be represented as

$$y = Z\varphi + u \tag{30}$$

where φ is a column vector $[B', \rho]'$ and Z is a matrix [X, Wy].

The estimated value of $\hat{\varphi}$ can be represented as

$$\hat{\varphi}_{S2SLS} = (H'\,\hat{Z})^{-1}H'y \text{ ; H=}Q(Q'\,Q)^{-1}Q'\hat{Z} \tag{31}$$

Equation (23) can be rearranged as

$$\hat{\varphi}_{S2SLS} = (\hat{Z}'\,\hat{Z})^{-1}\hat{Z}'y \tag{32}$$

and

$$\hat{\sigma}^2_{S2SLS} = \frac{\mathrm{e}'\mathrm{e}}{\mathrm{N}} \tag{33}$$

where e is a vector of the S2SLS residual.

Before describing the data, the creation of a spatial weight matrix is described in detail and alternative of each definitions are proposed.

3.1.5 The spatial weights matrix

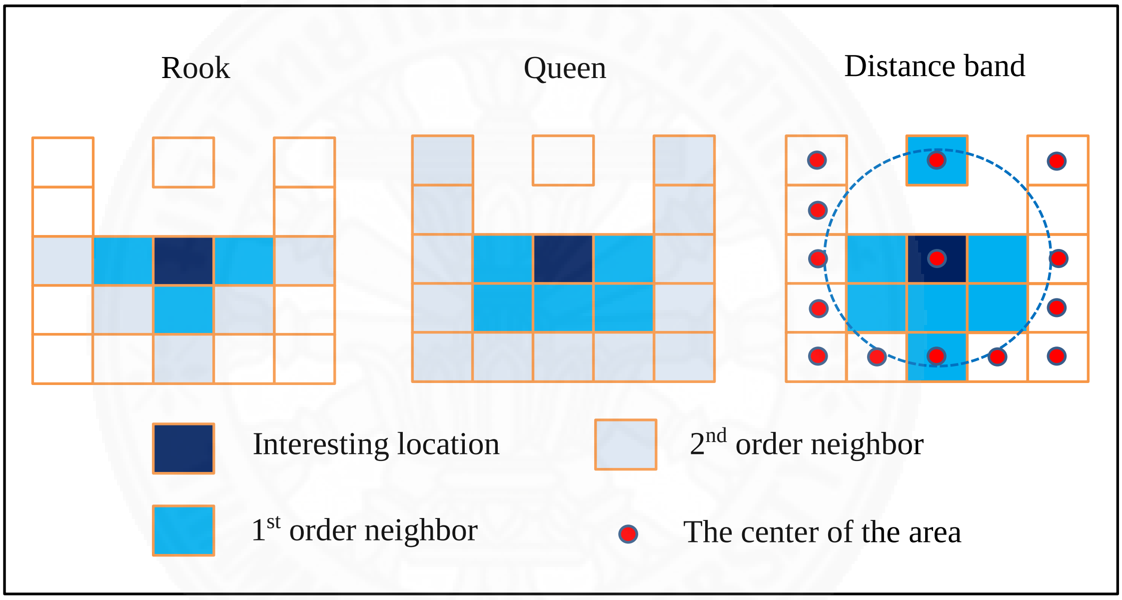
The most important part of spatial modeling is determining the closeness or nearness between the locations of the neighbor. A spatial model estimation with different definitions of the spatial weight matrix generates different model specification results and estimated values. A spatial weight matrix is a non-negative matrix that is created by two spatial-based ideas. the idea of contiguity or adjacent areas separated to be queen and rook contiguity. Another idea is of separating a distance-based spatial weight matrix separated to be distance band weight and k-nearest neighbors. The spatial weight matrix in this thesis is created based on two definitions, namely, rook contiguity and distance band or distance threshold.

Rook contiguity is defined as the adjacent locations which share a common line boundary. However, the rook criterion does not include the corner neighbors. If the definition of queen contiguity is applied, then those locations that share a common line boundary and the corner or a common vertex are defined as adjacent neighbors. Higher order neighbors may also be included to consider the effect from indirect neighbors or the neighbor of direct neighbors (i.e., second order neighbor). However, these criteria imply that those cities that are located on islands have no neighbor locations because they have no direct adjacent boundary.

To eliminate this “lonely island” problem, this thesis applies the distance band criterion, which is based on the distance from the point at the center of a polygon. If the linear distance between cities is less than the imposed threshold, then these cities are classified as neighbors even though some adjacent neighbors are located on the island. Figure 3.3 represents the border and neighbors of the central area and the lonely island under the rook, queen, and distance threshold criterion.

Figure 3.3

Spatial neighbor based on the rook, queen, and distance band criterion



Source: Author’s diagram

The weight matrix for the rook and queen definitions are recorded as 0 and 1. w_{ij} =1 if locations i and j are neighbors, and w_{ij} =0 if otherwise. For the distance criterion, the weight matrix is generated as an inverse distance matrix. In this matrix, each element in the spatial weights for i and j is computed by the following inverse distance function:

$$w_{ij} = f(d_{ij}, \alpha) \tag{34}$$

where w_{ij} is the weight of locations i and j, d_{ij} represents the inverse distance between locations i and j, and α is the decay parameter.

The attribute of distance decay effect decreases the distance function as the distance increases, that is, $\partial w_{ij}/\partial d_{ij} < 0$. The functional form is usually $w_{ij} = 1/d^{\alpha}_{ij}$, where the alpha is a fixed parameter that is equal to 1 for the inverse distance weight and equal to 2 for gravity weight. Nevertheless, this thesis uses an alpha value of 1 for the inverse distance matrix. $w_{ij} = 1/d_{ij}$ if the distance between locations i and j (d_{ij}) is less than the distance threshold ($d_{threshold}$) or if the linear distance remains in the band, and $w_{ij} = 0$ if the distance exceeds the threshold.

In general, the spatial weight matrix typically employs a row-standardized approach, which facilitates the making of interpretations and comparing the parameters of different models. Row-standardization is the given weights w_{ij} which is divided by the row sum as follows:

$$w_{ij(s)} = \frac{w_{ij}}{\sum_j w_{ij}} \tag{35}$$

The summation of each row equal to 1 and the summation of all weights equal to n or total standardization $S_0 = \sum_i \sum_j w_{ij}$ = n denote the number of all the observed locations. As mentioned above, when no island appears, each criterion has an equal number of observations. By contrast, when an island appears, the island is excluded from the adjacent-based spatial weight matrix yet is not eliminated by the distance criterion. An example of a row-standardized weight matrix ($W_{(s)}$) in the case of binary weight is illustrated as follows:

$$W_{(s)} = \begin{bmatrix} 0 & 0 & 1/3 & 1/3 & 1/3 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 1/2 \\ 1/4 & 1/4 & 0 & 0 & 1/4 & 1/4 \\ 1/5 & 1/5 & 1/5 & 0 & 1/5 & 1/5 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

After clarifying the econometric approach above, the use of data and variables will be explained in the next sub-section to reveal the limitations and to discuss the generation of data for the model at each scale. The explanatory and dependent variables at each scale are also defined and described. All this information must be explained before interpreting the results in the next chapter.

3.2 Data description and summary statistics

The dataset used for the analysis combines the data from the Labor Force Survey for the third quarter of 2001 to 2015 with those from GIS. The Labor Force Survey contains microeconomic data that are randomly collected by the National Statistical Office of Thailand. This survey covers more than 200,000 members of the Thai labor force that are sampled in each quarter via two-stage stratified sampling. The first stage enumerates the municipal and non-municipal areas, while the second stage enumerates the households and the persons living in a collective household. The sample units are then presented along with the sample weight to represent the characteristics of the whole country population. The sample, which is collected from 76 provinces at the individual labor level, is thoroughly analyzed to obtain an overview of the whole kingdom. The obtained geographical or spatial information covers 76 provinces in Thailand and supported by the data taken from GIS. These data can be plotted on a map, and the actual physical location of the GIS data can be identified by using geocodes. These data are used to create a spatial weight matrix and to describe the geographical relationships among the variables considered in this thesis.

3.2.1 Data description

The Thai Labor Force Survey data for the third quarter of 2001 to 2015 are pooled for the individual level model. These data include 3,220,988 observations after the dataset has been managed as described in Appendix A. The blank observations in the dataset are eliminated automatically by the program during the estimation process.

The Labor Force Survey data are collected by using a questionnaire that asks the respondents about their address, gender, age, education, occupation, industry, working hours, return on work, estimated income, and household characteristics, etc. However, years of working experience, which is the important variable in the Mincer-type model for analyzing and answering the research questions, is neither asked in the questionnaire nor reported in the database. Thus, years of working must be computed by subtracting the actual experience of a worker from his/her estimated age at the time of completing his/her education (Mincer, 1975). Hanushek et al. (2015), Tangtipongkul

(2015), and Montenegro et al. (2014) calculate the potential years of experience of a worker by subtracting his/her age during the survey from his/her estimated time of schooling and from six years[1] as follow:

$$\text{years of working experience} = \text{age reported at the time of the survey} \\ - \text{years of schooling} - \text{six years}$$

For the education variable, only the highest completed level of education (from pre-primary school to doctoral degree) is collected from the respondents. Years of schooling are not reported in the database. Therefore, the potential years of schooling are calculated based on the National Education Act of 1999 and the definition of the Ministry of Education of Thailand. People must spend six years in elementary or primary education and three years in lower secondary education as a compulsory education. Upper secondary education is a three-year voluntary course. Higher education includes diploma education up to doctoral degree. Generally, Diploma and post-upper secondary education are three-year courses, while undergraduate education is a four-year course. Those people who study at higher levels must spend two years in master's degree and five years in doctoral degree.

The nomination of human capital intensity by education level can be classified into three categories as shown in Table 3.1.

The original dataset is not available for use because the noisy data obtained during the data mining may generate statistical noise and disturb the statistical estimation. These noisy data are meaningless or corrupt data that are produced by either the lack of recorded data or the presence of false data. Therefore, data mining must be cleaned to reduce statistical noise as much as possible. The data must not be used directly because some variables must be generated as new variables. For example, the dependent variable must be generated as a natural logarithm of labor income. Both the standard and spatial provincial-level models also require the transformation of data into the average value of each quantitative variable for each province. Appendix A presents more details regarding the data generation and transformation processes.

[1] Age before entering elementary education.

Table 3.1

Nomination of human capital intensity by level of education

Education Classification	Human Capital Investment Intensity	Education Intensity
None	Low intensity	Low education
Less than elementary	Low intensity	Low education
Primary education	Low intensity	Low education
Lower secondary education	Low intensity	Low education
Upper secondary education	Moderate	Moderate education
Post-secondary education	Moderate	Moderate education
Bachelor degree education	Highly intensive	High education
Master degree level	Highly intensive	High education
Doctoral degree level	Highly intensive	High education
Other education	Not classified	Not classified

Source: Author's modification from International Standard Classification of Occupations by the International Labor Organization

The dataset in the provincial-level model comprises the average value of the quantitative variables and the proportion of the qualitative variables in order to characterize the provincial characteristics. The dataset includes 76 observations for each year. However, the original data must be weighted by the sample weight before generating the provincial variables. At this level, the original observation in the dataset is not eliminated and cleaned because the data will be deleted automatically after generating new variables by using the survey data analysis command. Appendix A presents more details.

The individual model uses the natural logarithm of worker's monthly income as an endogenous variable, while the provincial-level model uses the natural logarithm of mean monthly income of provincial workers as a dependent variable.

The explanatory variables in the individual model include many control variables apart from years of schooling and working experience. These variables include working hours, gender, marital status, type of occupation, type of industry, year of collected data, and geographical provinces. Main working hours refer to the total number of hours that workers spend in their main occupation as recorded in the database. All qualitative control variables are represented as dummy variables. For gender, males take a value of 1 while females take a value of 0. The respondents are split into three groups based on their marital status. Compared with the single respondents, the married ones have higher incentives to work hard, while the divorcees are relatively demotivated in their work, thereby influencing their income. Married labor group is recorded equal to 1 and 0 if otherwise. Divorced, widowed, or separated labor group 1, while the others take a value of 0. A dummy variable for working in the municipal area takes a value of 1. The respondents are also classified into public officials, state enterprise employees, and private employees according to their employment sector. The wage patterns of public officials or state enterprise employees may differ from those of private employees, which are used as the base case in this thesis. The sample is also classified into nine groups according to their occupation characteristics, namely, (1) legislators, senior officials, and managers; (2) professionals; (3) technicians and associated professionals; (4) clerks; (5) service workers and shops and market sales workers; (6) skilled agricultural and fishery workers; (7) craft and related trade workers; (8) plant and machine operators and assemblers; and (9) elementary occupations. The respondents are also divided into eight groups based on their type of industry, namely, (1) agriculture; (2) mining; (3) utilities; (4) construction; (5) low-skill manufacturing; (6) high-skill manufacturing; (7) low-skill services; and (8) high-skill services. The year dummy variables cover the years 2001 to 2015 to capture the time effect. The dummy variables of 76 provinces are also added to capture the differences in the labor income of workers across different provinces.

At the provincial scale, the most interesting explanatory variables include average years of schooling, average years of working experience, and work years squared. All qualitative variables are computed as the mean value of each quantitative variable at the provincial level. The model is used to capture the provincial average return on investment in human capital. Thus, the most important parameters in the macro Mincer function are years of schooling and working experience. The influence of geographical location as captured by a set of dummy variables does not matter at this scale because the geographical influence is described by the additional spatial variable in the model, which in turn reveals the spatial pattern of geographical influence or spatial dependency across neighbor provinces.

Tables B.1 and B.2 in Appendix B describe the variables for the individual- and provincial-level models, respectively.

3.2.2 Summary statistics

The summary statistics are described after introducing the variables for the individual- and provincial-level models. The tables in Appendix B present the summary statistics of each variable in either of these models. Table B.3 shows the descriptive statistics for the individual-level model. The quantitative variables are computed as the statistical mean and standard deviation of the whole population in each year (i.e., the average years of schooling and working experience in each year). However, the qualitative variables, such as gender, marital status, and occupation, cannot be described by statistical mean. Thus, the proportion of these variables must be measured by comparing number considered observation with the whole labor force. At this level, the data are explained after weighting the sample to infer the approximate actual population.

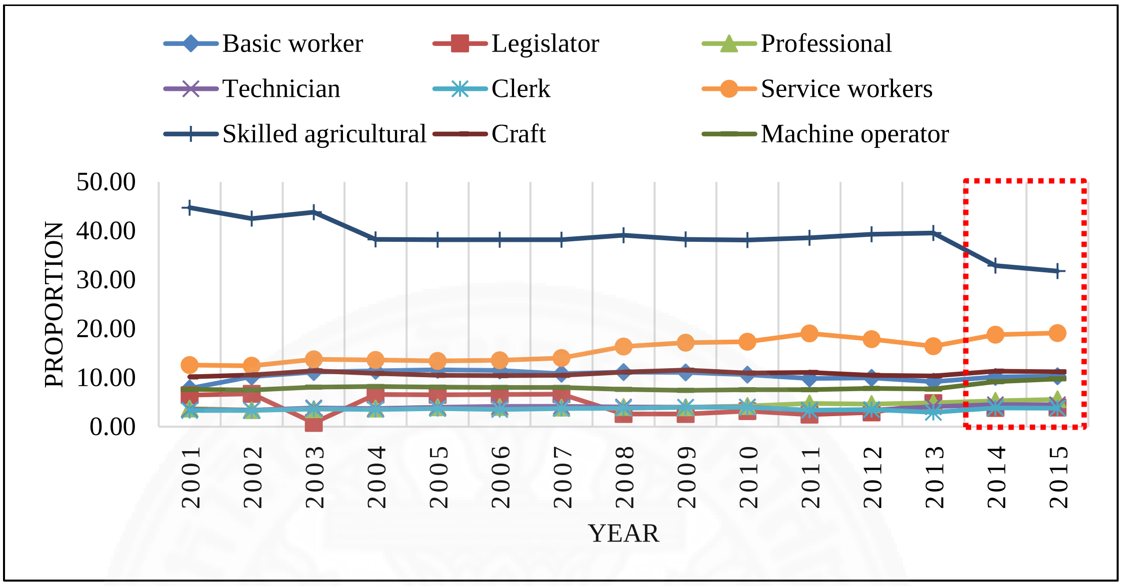
For the provincial level model, the variables must be generated to capture provincial characteristics instead of individual characteristics. Appendix A describes the variable generation process step by step. A total of 76 provincial observations are obtained for each year. The quantitative variables at this level are computed as the average values of each variable in the province. Therefore, the statistical description for each variable is represented as a statistical mean and a standard deviation. The mean value of the natural logarithm of average labor income, mean value of average schooling years, mean value of average years of work experience, and mean value of work experience squared are computed as the average value of each variable in the whole country in each year. Table B.4 shows more details about the statistical summary.

The descriptive statistics reveal that the average natural logarithm of labor income at the individual level increase and mean of the natural logarithm of average labor income at the provincial model increase from 8.4 to 9.3 and 8.6 to 9.4 in 2001 and 2015, respectively. The average schooling year of an individual worker in the whole country also increased from 5.4 years to 6 years in 2011 and 2015, respectively. However, mean of the average years of schooling in each province is spent around 5 years and 5.5 years in 2001 and 2015, respectively. In terms of work experience, the workers at the individual level have been working for 22 years to 28 years on average, while those at the provincial level have been working for 23 years to 25 years. The survey data also reveal the proportion of living area and gender population in Thailand. Most workers live outside the municipality, except for those in Bangkok because the whole Bangkok is a municipal area. In terms of gender, the female respondents have slightly outnumbered the males.

Most of the workers in the sample are employed by the private sector. Almost half of them are working as farmers, hunters, or fishermen, followed by service workers as shown in Figure 3.4. However, the proportional share of agricultural employees decreases every year, while the proportion of service workers significantly increases every year. Unsurprisingly, the agricultural sector emerges as the most popular industry, followed by the low-skilled service industry, high-skilled service industry, and low-skilled manufacturing industry as shown in Figure 3.5. This observation is consistent with the occupation share mentioned above. The tables in Appendix B present more details.

Figure 3.4

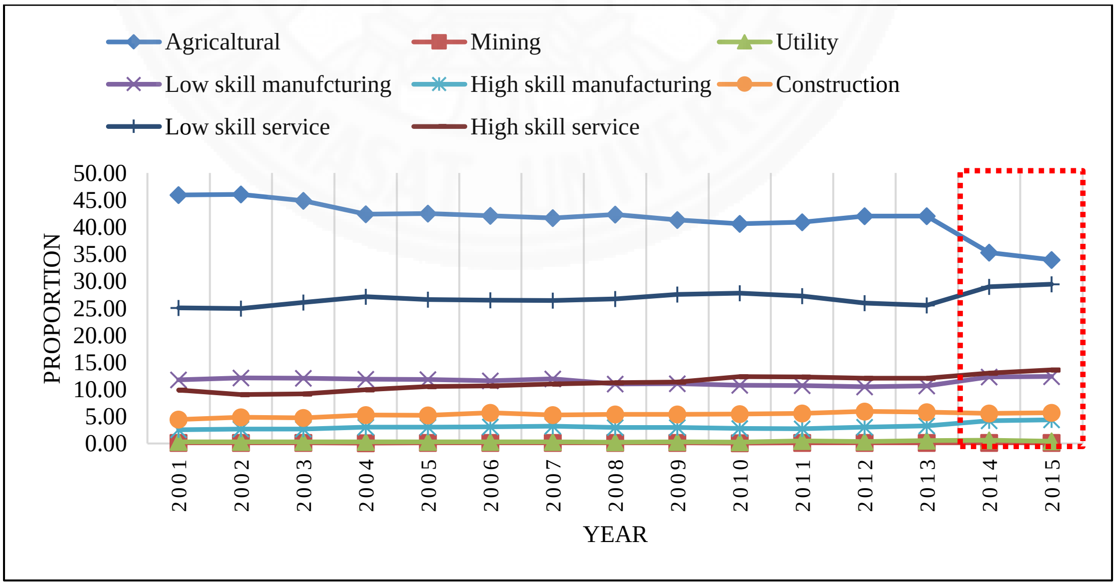
Occupation share of the population in the dataset



Source: Author's calculation from Labor Force Survey

Figure 3.5

Industrial share of the population in the dataset



Source: Author's calculation from Labor Force Survey

CHAPTER 4
RESULTS AND DISCUSSION

This chapter summarizes the empirical results from the traditional Mincer model either at the individual or provincial level and those from the spatial Mincer model. After the standard model of the individual has been estimated, the estimated coefficients of the provincial dummy variables are overlaid on a map to illustrate the distribution and concentration of labor income in Thailand. The spatial correlation is also illustrated by using Moran's I map. And then the spatial model is then applied to illustrate the pattern of spatial dependence. Finally, the spatial effect on the average return on human capital is also described to answer the research question.

4.1 The effect of human capital on individual labor income

The estimated results from the individual level model are tabulated to illustrate the effect of human capital on individual labor income. The locational effect on labor income is then plotted in different colors on the map of Thailand.

4.1.1 Return on education and working experience

Table 4.1 presents the empirical results of the traditional model for individual return on human capital in a semi-log income function under different specifications. In the regression, sampling weights are included in the estimation of all models. Years of schooling, working experience, and squared working experience are statistically significant at the significant level 0.01. Model 1, which was constructed without any control variables, shows that an additional year of education[1] can increase the monthly income of an individual by 11.59% on average if the other variables are fixed according to previous studies as shown in Table 2.1. After adding control variables such as working hours, gender, marital status, living area, working status, occupations, and industries in model 2, the average rate of return on schooling is

[1] One additional total year of education regardless of education level.

reduced to 7.3% per an additional year of schooling. In model 3, the cross-sectional effect is captured by the provincial dummy variables that explain the locational effect on individual labor income. The return on an additional year of education is only 7% on average. After the year dummy variables control the time unit effect in model 4, the average return on an additional year of schooling decreases to 5.77%. An increase in working experience also increases labor income with a diminishing rate. Model 4 shows that 89 years of work experience can lead to the maximum return. In addition, Model 4 also provides the highest R-squared value that presents goodness of fit, and can be used to explain 66.9% of the dataset. Therefore, the estimated coefficient of the provincial dummy variable from model 4 will be used in the next step.

Model 4 provides the lowest coefficient on years of schooling, thereby underscoring the importance of other factors, especially the year effect, in determining labor income. The monetary income of workers increases every year from 2001. Table 4.1 summarizes the results presented in Table Appendix C.1.

Table 4.1

Results of the traditional model at the individual level under different specifications

	Model1	Model2	Model3	Model4
sch	0.1159*** (0.0034)	0.0730*** (0.0020)	0.0704*** (0.0020)	0.0577*** (0.0025)
exp	0.0491*** (0.0008)	0.0404*** (0.0007)	0.0368*** (0.0006)	0.0324*** (0.0005)
expsq	-0.000497*** (0.0000)	-0.000428*** (0.0000)	-0.000380*** (0.0000)	-0.000364*** (0.0000)
constant	7.0649*** (0.0408)	6.6545*** (0.0569)	7.0368*** (0.0450)	6.9152*** (0.0462)
Control variables	No	Yes	Yes	Yes
provinces	No	No	Yes	Yes
Years	No	No	No	Yes
Prob>F	0.0000	0.0000	0.0000	0.0000
R-squared	0.449	0.565	0.607	0.669
Observations	745099	745098	745098	745098

Note: ***, ** and * are significant at 1%, 5% and 10% respectively; Standard Error is in parenthesis, and dependent variable is Lninc

Source: Author’s calculation from Labor Force Survey

The rate of return on schooling in model 1, which was constructed without any control variables, is similar to that obtained in other studies in Thailand that have been conducted by using the same functional form as mentioned earlier in the literature review (Table 2.1). The return on an additional year of schooling ranges between 9% and 16% depending on the number of years and the employed estimating method. However, the functional form of models 2 to 4 differs from that of other studies. By comparing with the rate of return around the world, Montenegro et al. (2014) found that the world average rate of return on another year of schooling from 2000 to 2013 remained at 9.7% . This rate is nearly similar to that recorded in high-income economies (10%) and in countries within East Asia and the Pacific (9.4%). Sub-Saharan Africa shows the highest return (12.4%), while the Middle Eastern and North African regions show the lowest return (7.3%). EU and South Asia have returns of 7.4% and 7.7%, respectively, which are lower than the world average. The rate of return in Thailand is also similar to that of other ASEAN countries such as Malaysia (12% in 2010), Indonesia (10.4% in 2010), and Singapore (12.5% in 1998).

In addition, the average rate of return per additional year of schooling in model 1 is also lower than that recorded in Japan (13.2%) and Korea (13.5%) in 2004. However, the average rate of return per additional year of schooling is higher that in the UK (6.8%) and the US (10.0%). OECD countries showed an average return of 10% per additional year of schooling in 2004 (Psacharopoulos, 2006). Generally, the rate of return in most developed countries is lower than that in Thailand and other developing countries. For example, the returns on investing in an additional year of schooling in Austria, Germany, Italy, and Sweden are 6.5%, 5.0%, 5.1%, and 4.4%, respectively (Glocker & Steiner, 2011). Meanwhile, low-income countries, such as sub-Saharan Africa, still greatly benefit from an additional year of schooling because of the relative scarcity of human capital in these countries (Michaelowa, 2000). Although the other models in this thesis have been structured differently from those in previous studies, the importance of other factors that affect labor income, including marital status, occupation, gender, and workplace location, remains to be seen

.

4.1.2 The map of income distribution and concentration

The estimated coefficients of the provincial dummy variable in model 4 are classified by quantile into six groups before they are plotted on a map by colors.

As the base case, Bangkok has an estimated coefficient of zero. Given that workers in Bangkok receive the highest income among other workers in Thailand, the estimated coefficient of the other provinces should negative. A low negative coefficient means that the income of workers in a specific province is close to that of workers in Bangkok, while a high negative coefficient indicates that the income of these workers is much lower than that of workers in Bangkok.

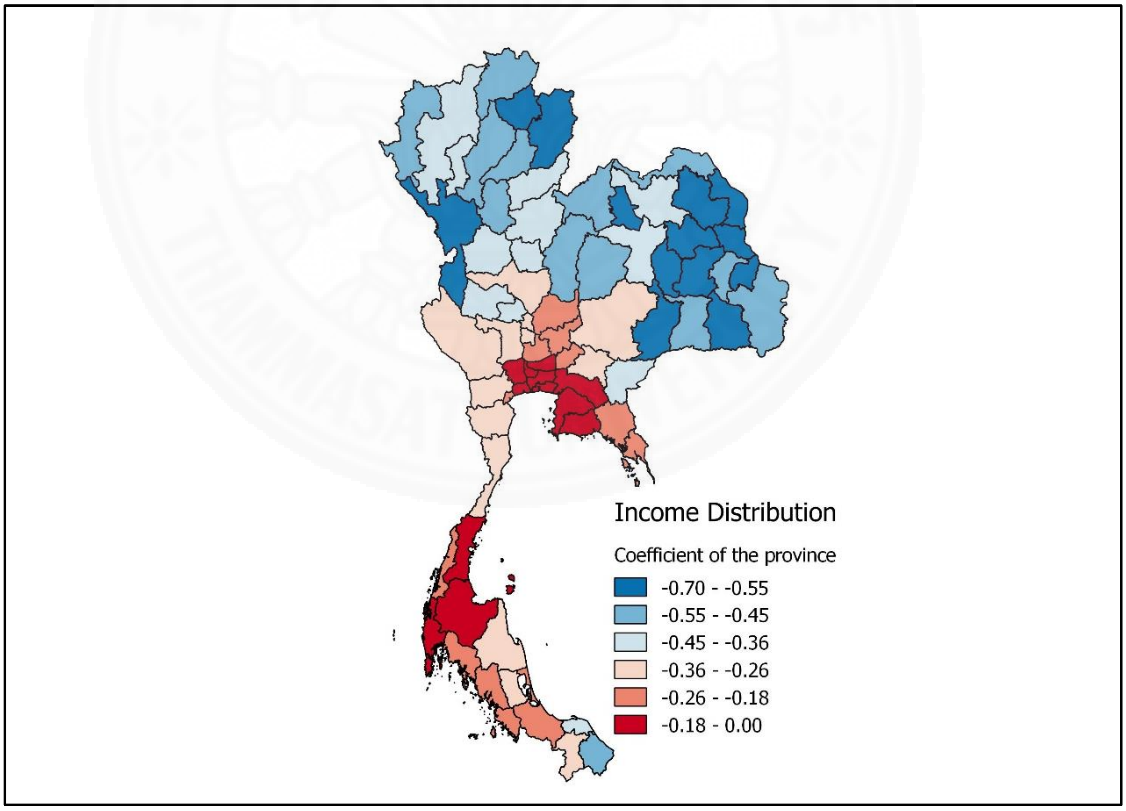
Figure 4.1 presents the locational effect on labor income in Thailand. Different income levels are classified by quantile into six groups, with each group represented by a specific color. The darkest shade of red represents the province where workers receive the highest income, the lighter shades of red and blue represent those provinces with a lower income, and the darkest shade of blue represents the provinces with the lowest labor income. Figure 4.1 displays the distribution of labor income across all provinces in Thailand. The highest labor income is highly clustered in Bangkok Metropolitan and surrounding areas such as Bangkok, Samut Prakan, Nonthaburi, Pathum Thani, Nakhon Pathom, and Samut Sakhon, as well as in industrial estate provinces, such as Chon Buri and Rayong. In the southern region, the concentration of the highest labor income is observed in Phuket, Phangnga, Suratthani, and Chumphon. Almost all provinces with the lowest labor income concentration are located in Northeastern Thailand.

For the highest income province group, those laborers who are working outside Bangkok receive a lower income than those who are working in Bangkok in rank between 2% and 17% when the other factors are fixed. For example, those laborers who are working in Nonthaburi, Samut Prakan, Nakhon Pathom, Chon Buri, Phuket, Phangnga, and Suratthani earn about 2%, 11%, 16%, 12%, 9%, 17%, and 6% less than those working in Bangkok, respectively. Meanwhile, those laborers who are working in the second (including Pra Nakhon Si Ayutthaya and Krabi), third (including Nakhon Ratchasima), fourth (including Chiang Mai, Khon Kaen, and Lampang), fifth (Loei and Phrae), and sixth province groups (including Nan, Tak, Maha Sarakham, and Si Sa Ket) earn 18%–26%, 26%–28%, 39%–45%, 46%–54%, and 56%–70% lower than those who are working in Bangkok, respectively.

Figure 4.1 shows that the labors in main cities, especially Bangkok Metropolitan and the adjacent provinces, receive a higher income compared with the other workers if the other factors are fixed. This finding implies that workplace location has an important role in determining individual labor income in Thailand. The clustering of provinces with the highest labor income also implies a spatial externality or spatial correlation between these provinces. Therefore, the geographical effect, which may be spatially omitted or treated as a nuisance in the estimation process, must be considered when using the data.

Figure 4.1

The map of the distribution of the provincial effect on labor income



Source: Author's calculation from Labor Force Survey

4.2 The results of Moran's I test

In statistics, a spatial autocorrelation between proximal locations in space can be measured by global Moran's I, which evaluates whether the expressed pattern is clustered or dispersed. Each spatial weight matrix that is constructed under a specific spatial definition affects the Moran's I index differently because of different sets of neighborhoods. The local indicator of spatial association (LISA) is used to evaluate the existence of clusters in a specific location. Those variables with high and low values are investigated by comparing their statistical means. In this case, the average values of variables at the provincial level are used to measure the spatial influence across provinces. The results presented in this section are estimated using the GeoDa software.

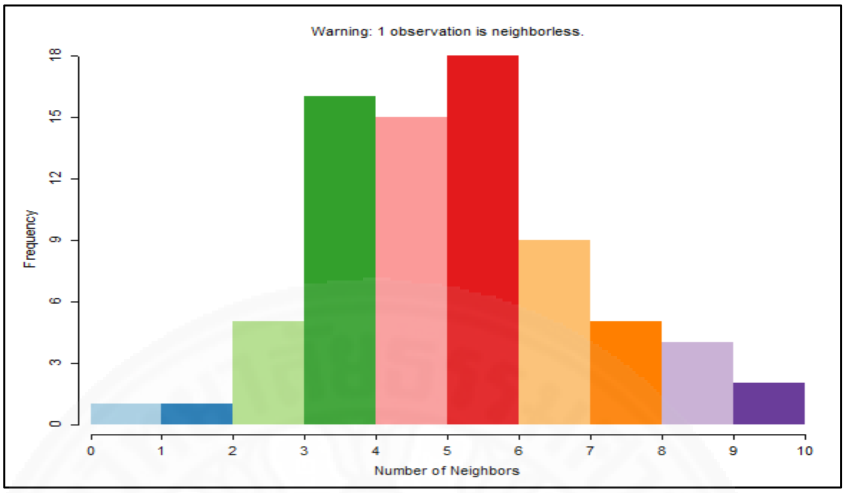
4.2.1 Spatial autocorrelation

The spatial weight matrix is defined by a rook contiguity and a spatial distance band criterion at 150 kilometers of the threshold to confirm the spatial pattern of the geographical relationship between provinces. The proximity areas can reasonably be called that the neighbors if they are defined by sharing a boundary. However, previous studies argue that the geography of population flow across the provinces and the intercity area for the agglomeration of economic and social activities also interact with in the areas of 150 kilometers of a radius from the city (Bazzi, Gaduh, Rothenberg, & Wong, 2016; Hutchison, 2009; Rigotti, 2006).

Figure 4.2 presents a histogram of the number of connected neighbors based on the definition of spatial rook contiguity. Figure 4.3 presents the number of connected neighbors when the set of spatial neighbors in the spatial weight matrix is defined by using the spatial inverse distance with the 150 kilometers band criterion. The histogram does not have neighborless province after the neighbor set is modified by using spatial distance idea.

Figure 4.2

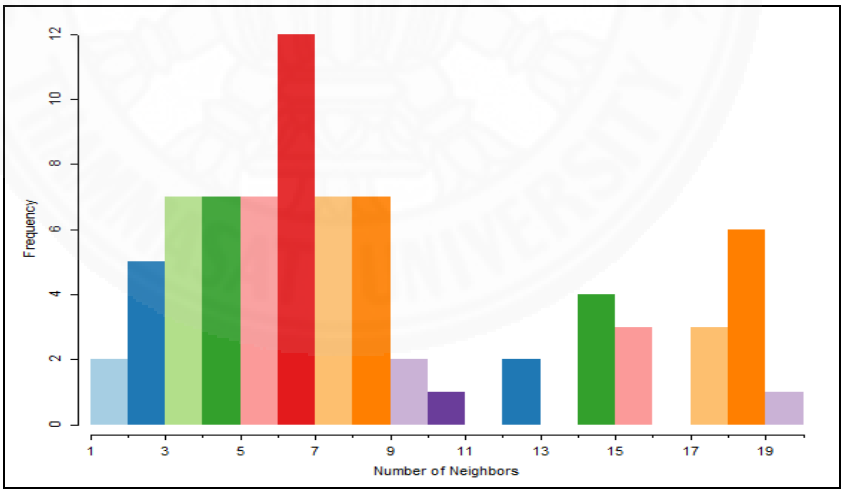
The number of connected neighbors based on the definition of rook contiguity



Source: Author's calculation from Labor Force Survey

Figure 4.3

The number of connected neighbors based on the 150 kilometers criterion



Source: Author's calculation from Labor Force Survey

Figure 4.4 reveals a spatial autocorrelation in the labor income of neighbor provinces. Using the average labor income of a province or its natural logarithm reveals a similar spatial correlation that is measured by the global Moran's I index if the spatial correlations are considered under the same weight matrix. Meanwhile, given the varying definitions of the spatial weight matrix, some differences are observed in the results of both spatial definitions. The Moran's I index under spatial rook contiguity reveals a higher spatial correlation between neighbors than the Moran's I index under the spatial inverse distance of a radius located 150 km away from the center of the host province. The index under rook contiguity shows that the rank of spatial correlation varies between 0.3 and 0.5, thereby indicating a positive linear spatial correlation of average labor income and the natural logarithm of average labor income between neighbor provinces.

Such positive spatial correlation implies that those provinces with similar average labor income are clustered together on the map. For instance, if the labor income in a particular province is high, then the labor income in its neighbor province is also high. Conversely, if the labor income in a particular province is low, then the labor income in its neighbor province is also low. The positive spatial autocorrelation also reveals that the spatial distribution of high or low labor income provinces is spatially clustered. However, Moran's I only shows the spatial linear correlation of variables between neighbors and does not imply causation. A high positive correlation merely indicates a strong positive linear relationship and may also imply that the high labor income provinces are more clustered and low labor income provinces are more clustered as well.

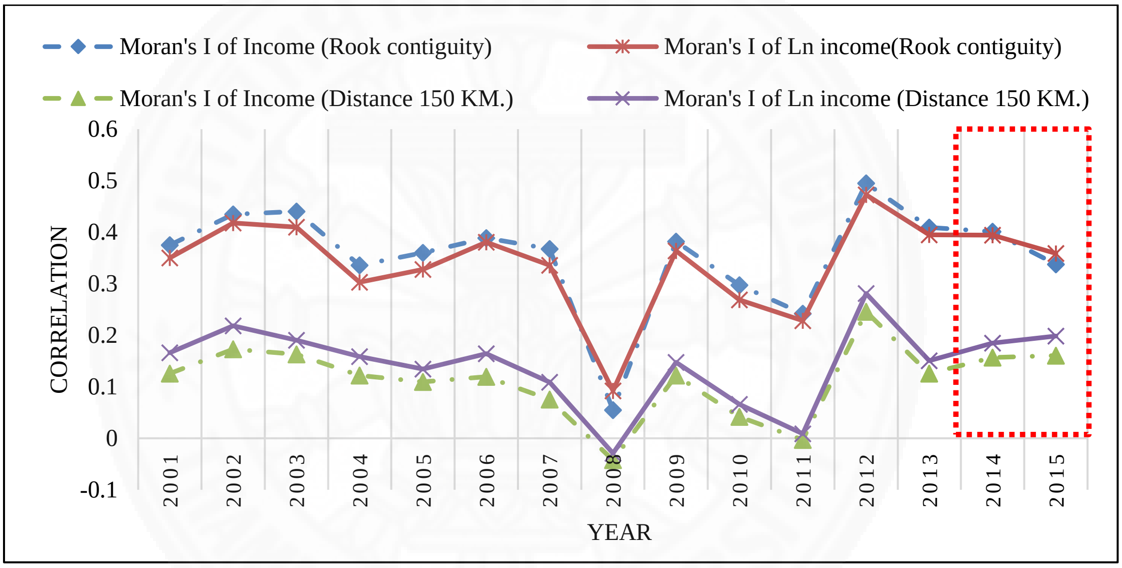
By contrast, a negative spatial autocorrelation implies that the spatial distribution of high and low labor income provinces is spatially dispersed, and such dispersion often reflects a spatial competitive process.

Figure 4.4 shows that the index is lower than 0.3 in 2008 and 2011, during which the world was hit by a global financial crisis and Thailand suffered from a great flood, respectively. These values imply that the geographical relationship between provinces is sensitive to international and domestic economic conditions. For instance, these provinces may show a positive linear relationship when the economic conditions are normal but may show a weak linear relationship during crises. The apparent decline of the index during the global financial crisis indicates that those provinces with a similar average labor income are either less spatially clustered or more spatially scattered at the time. A near-zero index may also imply the random spatial process of labor income during the crisis.

Both spatial definitions of the spatial weight matrix show that the Moran's I index moves along the same direction in each year even though they have the different groups of the neighborhood in spatial weight matrix. These findings reflect the robustness of the spatial correlation and the consistent movement of spatial dependency in each year. The next section measures and plots the linear relationship of the average provincial labor income between neighbors on the map of Thailand.

Figure 4.4

Spatial autocorrelation of average labor income



Source: Author's calculation from Labor Force Survey

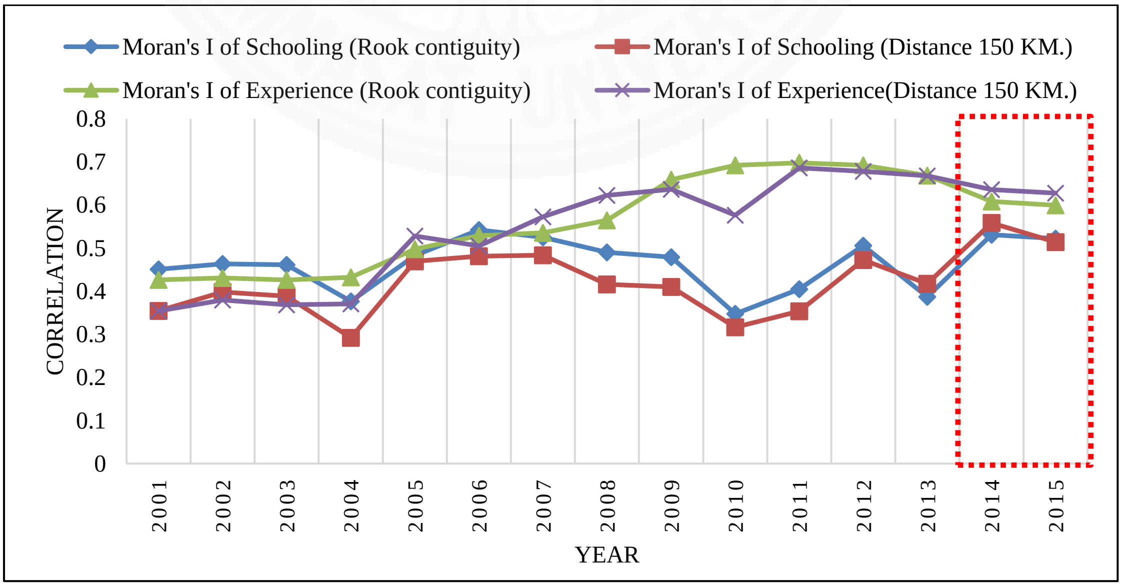
For the average years of schooling and working experience of workers in the province, Moran's I index is also measured under either the spatial rook or the spatial distance weight matrix. Figure 4.5 presents the spatial autocorrelation of average years of schooling and working experience between neighbor provinces. The index for average years of schooling is moderately positive between 0.3 and 0.56 depending on the year. Compared with another spatial weight matrix, the spatial rook shows that the Moran's I index reaches a high value in most years. The positive correlation indicates that those provinces with a similarly educated labor force are spatially clustered together.

Figure 4.5 shows a positive spatial correlation between provinces in terms of average years of work experience. Such correlation shows an increasing trend since 2004, which may be attributed to the changes in survey methodology. Such tendency also points toward a highly positive linear relationship between those provinces where workers have similar years of work experience. A stronger relationship implies that experienced workers are more concentrated and inexperienced workers are more concentrated together as well, that is, these workers live together with workers who have the same level of experience.

Statistically, such correlation is ranked between 0.4 and 0.8 depending on the year. The positive relationship implies that those provinces with highly (lowly) experienced labors are surrounded by neighbor provinces with highly (lowly) experienced labors. The spatial rook weight matrix provides a higher value of the Moran’s I index than the spatial distance criterion. However, Figure 4.5 shows that the spatial autocorrelation of average working experience under the spatial rook contiguity of the neighbor sets is smoother than that under the spatial inverse distance criterion of spatial weight matrix.

Figure 4.5

Spatial autocorrelation of the average years of schooling and working experience



Source: Author’s calculation from Labor Force Survey

4.2.2 Local Moran's I (LISA)

Figures 4.4 and 4.5 show that the impact of rook contiguity neighbors is larger and smoother than that of another weight matrix. Thus, in this section, the spatial autocorrelation of the natural logarithm of average labor income, average years of schooling, and average years of working experience in the province is measured by using local indicators of spatial association (LISA), which indicates the spatial relationship between each pair of provinces and neighbors. The positive index reported in the previous section indicates that those provinces with high/low income, education, and experience are clustered together on the map.

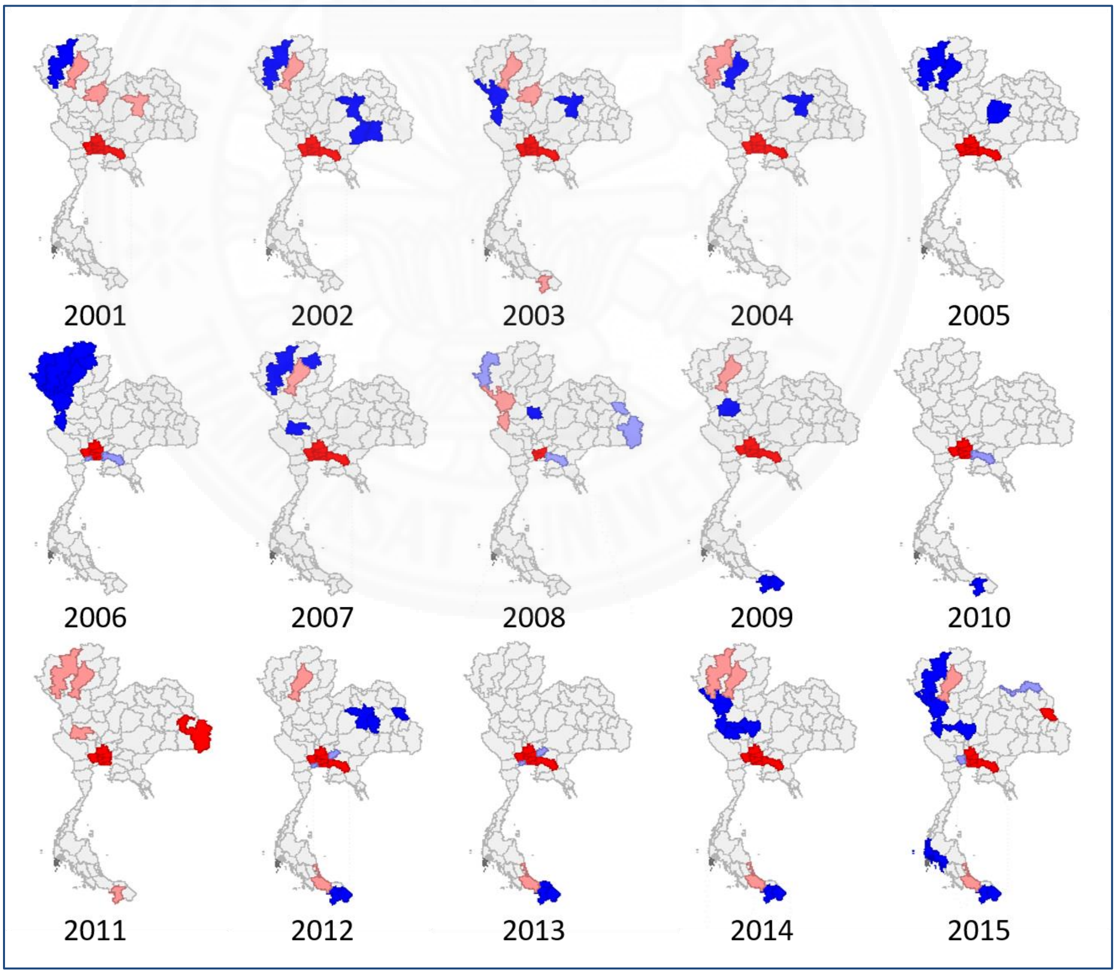
Figure 4.6 presents the local autocorrelation between neighbors. Each color on the map has a specific meaning and reflects the spatial significance of the LISA index at least at the 0.05 level. The red color (high–high) represents a spatial cluster of high labor income provinces and indicates that a province with a high natural logarithm of average income is spatially surrounded by neighbors with a similarly high labor income. The dark blue color (low–low) represents a cluster of provinces with low labor income and indicates that a province with a low natural logarithm of average labor income is spatially surrounded by neighbors with a similarly low labor income. The pink color (high–low) indicates that a province with a high labor income is spatially surrounded by neighbors with a low labor income. The light blue color (low–high) indicates that a province with a low labor income is surrounded by neighbors with a high labor income. The gray color denotes the insignificant spatial autocorrelation of labor income across locations and implies that the distribution of labor force income in a province is spatially random. Those provinces that are marked in gray are surrounded by neighbors with mixed levels of income.

The high and low labor income provinces are identified by comparing with the mean natural logarithm of their average labor income in each year. The statistical means for years 2001 to 2015 are 8.59, 8.59, 8.61, 8.67, 8.74, 8.82, 8.84, 8.96, 8.93, 9.00, 9.09, 9.15, 9.24, 9.33, and 9.35, respectively.

The red color in Figure 4.6 shows the same pattern in each year. This color shows that the cluster of those provinces with a high labor income which are all surrounded by the same type of neighbors are Bangkok Metropolitan areas, including

Samut Prakan, Nonthaburi, Pathum Thani, Phra Nakhon Si Ayutthaya, Nakhon Pathom, and Samut Sakhon. The other colors show varying patterns in each year. The provinces marked in dark blue are mostly located in the northern and northeastern parts of Thailand. In addition, the violence in the deep south over the past four years may have affected the spatial distribution pattern of average labor income as well. This observation confirms the spatial influence of labor income distribution as described in section 4.1.2.

Figure 4.6
Local Moran's I test of the natural logarithm of average labor income from 2001 to 2015



Source: Author's calculation from Labor Force Survey

As mentioned earlier, spatial correlation is sensitive to either international or domestic economic conditions. For example, those provinces with a high labor income which have similar neighbors have a low density and are less spatially clustered in 2008 and 2011 than in the other years. Some provinces, such as Phra Nakhon Si Ayutthaya and Nakhon Pathom, which used to be spatially surrounded primarily by neighbors with similar labor force characteristics, were spatially surrounded by provinces with varying labor income (random type of neighbors) in 2008, thereby confirming that labor income is either more scattered or less clustered during a crisis than in periods of normal economic conditions according to the global Moran's I index.

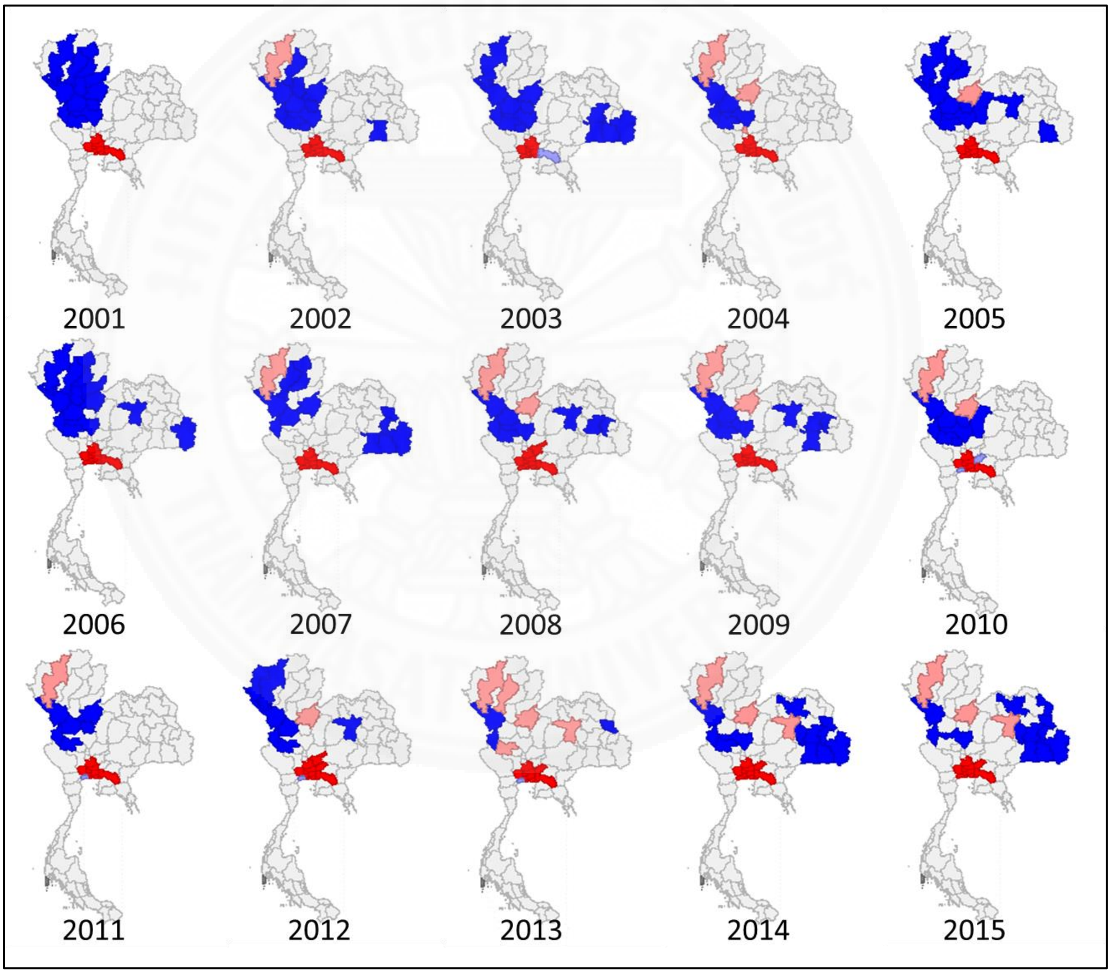
For the local Moran's I index of average years of schooling under spatial rook contiguity, each color in Figure 4.7 has the same meaning as that of average income in Figure 4.6. For example, red represents those provinces with highly educated workers which are spatially surrounded by the similar neighbors, blue represents those provinces with lowly educated workers which are spatially surrounded by the similar neighbors. Pink color represents those provinces with highly educated workers which are spatially surrounded by the dissimilar neighbors, light blue color represents those provinces with lowly educated workers which are spatially surrounded by the dissimilar neighbors as well.

Those provinces with high and low average years of schooling, which are used to analyze the local spatial correlation between neighbors, are identified by comparing the average years of schooling with its statistical mean. The statistical mean of average years of education is 4.99, 5.13, 5.33, 5.54, 5.60, 5.68, 4.56, 4.72, 4.89, 5.08, 5.18, 5.28, 5.48, 5.36, and 5.47 years for 2001 to 2015, respectively.

Figure 4.7 also shows that those provinces with highly educated workers which have similar neighbors are spatially clustered in the capital city and the surrounding areas. Those highly educated workers who have spent much time in school have emigrated to Bangkok Metropolitan areas, such as Bangkok, Samut Prakan, Nonthaburi, Pathum Thani, Phra Nakhon Si Ayutthaya, Nakhon Pathom, and Samut Sakhon, to look for work. These workers also receive higher returns than the lowly educated workers. Therefore, the high-income workers are spatially clustered in these provinces as well. This result is consistent with those reported in the literature, which contend that highly educated workers want to migrate to large cities and receive high income. Meanwhile, the lowly educated workers are spatially clustered in the northern and northeastern regions, including the main provinces such as Chiang Mai and Khon Kaen. Chiang Mai is colored pink in Figure 4.7 in several years, which implies that its labor force is highly educated while that of its neighbor provinces is lowly educated. Khon Kaen shows the same pattern as Chiang Mai in the last three years as well.

Figure 4.7

Local Moran's I test of the average schooling years of workers from 2001 to 2015



Source: Author's calculation from Labor Force Survey

The local Moran's I of the provincial average years of working experience in Figure 4.8 is also represented by colors, with each color having the same meaning as those in Figures 4.6 and 4.7. Specifically, the hot spot (red) or high–high and the cool spot (dark blue) or low–low indicate that those provinces with a highly and lowly experienced workforce and are spatially surrounded by neighbors with a highly and lowly experienced workforce, respectively. The pink and light blue colors on the map, also called high–low and low–high, indicate that those provinces with a highly and lowly experienced workforce are geographically surrounded by neighbors with a lowly and highly experienced workforce, respectively.

Those provinces with high and low average years of work experience, which are used to analyze the local spatial correlation between neighbor provinces (i.e., high–high, low–low, high–low, and low–high), have been identified by comparing the average years of work experience in the province with the corresponding statistical mean in each year. The statistical mean for years 2001 to 2015 is 27.50, 27.54, 27.67, 27.72, 28.24, 28.41, 22.57, 23.03, 23.24, 23.44, 23.66, 23.83, 24.02, 25.52, and 25.75 years, respectively.

The red color in Figure 4.8 shows that most provinces are spatially clustered in the northern region especially after 2004, while some provinces are spatially clustered in the central region. Those provinces with a highly experienced workforce which is surrounded by similar neighbors are clustered together on the map. For instance, after 2004, those provinces with a highly experienced workforce are clustered together, including Chiang Mai, Chiang Rai, Lampang, Uttaradit, Phrae, Nan, Phayao, Nakhon Sawan, Uthai Thani, Kamphaeng Phet, Sukhothai, Phitsanulok, Phichit, Phetchabun, Suphan Buri, Ang Thong, Lop Buri, Sing Buri, and Chai Nat.

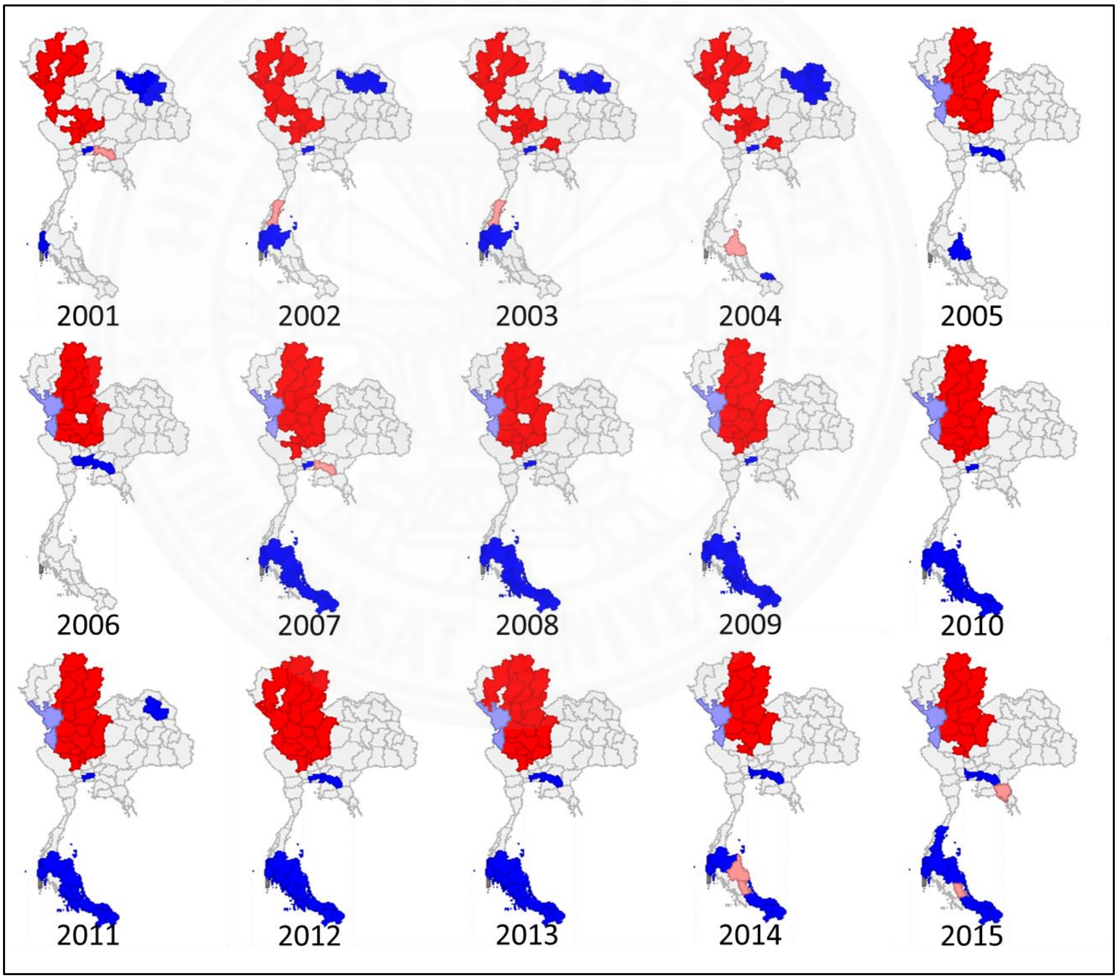
As mentioned earlier, the positive spatial autocorrelation of highly experienced workforce shows an increasing tendency especially after 2004, and such autocorrelation affects the intensity of the local spatial pattern. Those provinces with lowly experienced workers which have similar neighbors are more spatially clustered after 2004 than in the previous years. The local spatial pattern changes due to the increased focus of the National Statistics Office of Thailand on informal workers beginning from its launch of the first informal employment survey in 2005.

Meanwhile, Bangkok is always marked blue on the maps, thereby indicating that the city has a lowly experienced workforce and is geographically surrounded by neighbors with the same type of workers. This observation is also evident among provinces in the lower southern region, including in Phangnga, Surat

Thani, Krabi, Songkhla, Satun, Trang, Phatthalung, Pattani, Yala, and Narathiwat, especially after 2006. The spatial distribution of highly experienced workers differs from the spatial distribution of highly educated workers and highly income of workers. However, one cannot conclude that all workers with high working experience are living in areas with low average income because in return, these areas are not all of the provinces with lowest average labor income.

Figure 4.8

Local Moran's I test of the average years of work experience from 2001 to 2015



Source: Author's calculation from Labor Force Survey

Nevertheless, all results in this part are consistent with those previously reported by the Moran's I index. Therefore, both LISA and Moran's I help highlight the importance of the spatial influence and geographical clustering of labor force income, education, and working experience in Thailand.

The global and local Moran's I indices discussed in this chapter are both statistically significant. The null hypothesis of no spatial dependence is rejected at the 0.05 significance level in all areas, except for the gray ones. The effect of spatial dependency on return on human capital will be discussed further in the next sub-section.

The results of the local Moran's I index reveal the spatial clustering of provinces with high average labor force income, high average years of schooling, and high average years of work experience. The index also shows a weak spatial relationship among provinces in some years that can be attributed to economic conditions.

In the empirical evidence in the second half of 2008, oil prices around the world dramatically increased to a record high of $147 per barrel in July before sharply decreasing afterward. The economic performance and exports of Thailand were also gloomy at the time because of the spillover effect from the global financial crisis. The domestic political issues in Thailand also contributed to its economic troubles. The poor exports of Thailand mainly affected the country's primary exporting industries, especially the target industry, including electrical and electronics, agro processing, automotive, textile and garment, and petrochemicals (Economic Research Team, 2008).

In 2008, the Moran's I index recorded a significant decrease in spatial correlation according to LISA. Those provinces with average labor income are either less spatially clustered or more dispersed in 2008 than in other years. The number of high labor income provinces that are surrounded by high labor income neighbors has also been reduced. Many provinces, such as Phra Nakhon Si Ayutthaya and Nakhon Pathom, show statistically insignificant results in 2008. These provinces are mainly focused on the production industry, which is also the target industry of Thailand, because they are located in metropolitan areas and have many production facilities. The main industries in Nakhon Pathom include agro processing, plastics, metals, chemicals, and textiles, while those in Phra Nakhon Si Ayutthaya include electrical and electronics and automotive.

Before 2008, those provinces which have high average labor income used to be statistically surrounded by the similar neighbor. However, after the onset of the global financial crisis in 2008, the spatial correlation sharply declined and the local Moran’s I index became statistically insignificant, thereby implying that these provinces which were once surrounded by high labor income neighbors are now spatially surrounded by neighbors with various levels of income. Therefore, industries, particularly those that are directly related to exports, must learn to effectively adapt to the economic conditions. An insufficient spillover effect is also observed during a crisis.

However, the clustering of provinces with highly educated and experienced workforce does not change sharply over time because the effects of economic conditions are only temporary. Therefore, instead of changing their labor force and dismissing their workers, the industries in these provinces have only reduced their production capacity, which has subsequently affected the income of workers only during the crisis.

4.3 Evidence from the spatial model on the return on human capital at the provincial level

This section applies spatial regressions to explore the return on human capital at the provincial level and to explore the effect of spatial dependency on these returns. Spatial models are also used to reveal a spatial pattern and spatial influence in the estimated model. These models are then compared to answer the research questions, achieve the research objectives, and generate key findings.

4.3.1 Traditional Mincer model and specification testing

Tables 4.2 and 4.3 present the results of the standard Mincer model at the provincial level for the years 2001 to 2015. These tables also present the results of the spatial specification testing, which is performed to check for the existence of a spatial effect and to identify a spatial pattern in the model before estimating the spatial model. The Jarque–Bera test is performed to check the normality distribution of the error term, the Moran's I index—under the null hypothesis of no spatial autocorrelation of the error term—is used to check for the existence of a spatial autocorrelation or spatial effect, and the Lagrange multiplier test of spatial lag and spatial error is performed to identify whether the spatial pattern represents a spatial lag or a spatial error effect and to identify the most appropriate spatial econometric model.

The results in Tables 4.2 and 4.3 are the same because they are estimated under the same specifications. However, these tables obtain different results for the spatial specification testing because these results have been evaluated under different spatial weight matrices. In these tables, the rate of return on average years of schooling ranges from 10% to 23% per additional year of schooling. These findings are strongly and statistically significant at the 0.01 level. The average years of work experience is statistically insignificant before 2009 and becomes statistically significant starting from 2009. The rate of return on work experience is positively significant at a diminishing rate. On average, those provinces with laborers having an average work experience of 23 years to 27 years receive the highest economic return depending on the year. Almost all reported R-squared values are higher than 0.6, which indicates that more than 60% of the observed data can be explained by the traditional Mincer model.

Given that different spatial weight matrices have been used for the model specification testing, the results of the spatial specification testing in Tables 4.2 and 4.3 are also different. The neighbor sets in the spatial weight matrix for specification testing in Table 4.2 are defined by rook contiguity. The Moran's I index in the same table represents the significant spatial autocorrelation of the error term in the standard Mincer model and implies a spatial influence on the use of geographical survey data. Nevertheless, the intensity of such spatial effect varies across each year. For example, 2001 and 2008 do not show any spatial effect according to the spatial testing results presented in the previous section. The pattern of an omitted spatial variable in the model must be identified by performing a Lagrange multiplier test for both spatial lag and spatial error. Performing this test is also necessary to determine the most appropriate spatial pattern model before estimating the spatial models. The test results reveal a spatial lag effect in some years, but the spatial error that acts as a nuisance in the use of spatial data can be observed in several years. Therefore, the spatial error model is considered appropriate for application, while the spatial lag model needs to be subjected to further model comparisons. The spatial model in 2008 and 2011, during which the world was hit by a financial crisis and Thailand suffered from a great flood, must be estimated by using other GMM estimation techniques because the null hypothesis of normal distribution was rejected in the Jarque–Bera test. The results in this part strongly confirm the important role of spatial influence in the estimation.

To assess its robustness, the spatial effect must be repeatedly measured by different sets of neighbors. As mentioned earlier, the traditional Mincer model in Table 4.3 is exactly the same as that in Table 4.2, but the spatial specification testing yields varying results because different spatial weight matrices have been used. Table 4.3 reveals that the Moran's I index is insignificant in 2001, 2008, and 2009, and that the spatial error effect is more significant than the spatial lag effect in several years more than testing under spatial weight matrix of spatial rook criteria, but others still follow above qualification.

Table 4.2

Standard model with a spatial specification testing of the spatial rook contiguity weight matrix

	2001	2002	2003	2004	2005	2006	2007	2008
sch	0.2204*** (0.0280)	0.2252*** (0.0235)	0.2311*** (0.0235)	0.2135*** (0.0220)	0.1957*** (0.0236)	0.1490*** (0.0230)	0.1553*** (0.0166)	0.1685*** (0.0390)
exp	-0.0638 (0.1757)	0.0296 (0.1491)	0.1025 (0.1545)	0.0200 (0.1542)	0.1330 (0.1058)	0.0205 (0.1016)	0.1585 (0.1128)	0.4402* (0.2564)
expsq	0.0017 (0.0031)	0.0001 (0.0026)	-0.0012 (0.0027)	0.0001 (0.0027)	-0.0019 (0.0019)	-0.0001 (0.0018)	-0.0036 (0.0025)	-0.0097 (0.0055)
constant	7.9286*** (2.5880)	6.5953*** (2.1938)	5.5068** (2.2784)	6.8387*** (2.2723)	5.4504*** (1.5766)	7.4915*** (1.5360)	6.3880*** (1.2982)	3.2186 (3.0152)
R-squared	0.6633	0.7025	0.6972	0.7055	0.6323	0.5522	0.6074	0.2546
Log likelihood	50.5860	59.6830	60.4250	62.5930	53.4780	54.8580	59.1990	10.2010
AIC	-93.1720	-111.3670	-112.8510	-117.1860	-98.9550	-101.7170	-110.3970	-12.4020
SIC	-83.8490	-102.0440	-103.5280	-107.8630	-89.6330	-92.3940	-101.0740	-3.0800
Jarque-Bera (p-value)	0.7906	0.8176	0.3678	0.5950	0.8801	0.9400	0.3520	0.0000***
Moran's I (p-value)	0.1396	0.0003***	0.0021***	0.0054***	0.000***	0.0001***	0.0232**	0.3827
LM (lag) (p-value)	0.1862	0.0332**	0.1001	0.1762	0.2609	0.5811	0.7557	0.6888
LM robus (lag) (p-value)	0.2185	0.0606*	0.1490	0.2414	0.4509	0.7916	0.8577	0.6457
LM (error) (p-value)	0.2841	0.0017***	0.0098***	0.0200**	0.0000***	0.0004***	0.0845*	0.7026
LM robus (error) (p-value)	0.3391	0.0029***	0.0139**	0.0261**	0.0000***	0.0005***	0.0880*	0.6574

Table 4.2 (continued)

	2009	2010	2011	2012	2013	2014	2015
sch	0.2296*** (0.0201)	0.2026*** (0.0219)	0.1905*** (0.0240)	0.2134*** (0.0159)	0.1834*** (0.0164)	0.1229*** (0.0135)	0.1010*** (0.0156)
exp	0.3159** (0.1337)	0.4837*** (0.1481)	0.4701*** (0.1762)	0.2595* (0.1335)	0.4921*** (0.1268)	0.2085*** (0.0787)	0.2047** (0.0883)
expsq	-0.0067* (0.0029)	-0.0103*** (0.0031)	-0.0100*** (0.0037)	-0.0054* (0.0027)	-0.0101*** (0.0026)	-0.0037** (0.0015)	-0.0037** (0.0017)
constant	4.0938** (1.5821)	2.3489 (1.7544)	2.6080 (2.1040)	4.9367*** (1.6132)	2.2945 (1.5388)	5.8087*** (1.0187)	5.9865*** (1.1489)
R-squared	0.6587	0.5828	0.4898	0.7278	0.6528	0.5476	0.3757
Log likelihood	62.349	58.388	50.402	70.707	68.555	61.746	52.637
AIC	-116.699	-108.776	-92.805	-133.414	-129.109	-115.493	-97.274
SIC	-107.376	-99.453	-83.482	-124.091	-119.787	-106.17	-87.951
Jarque-Bera (p-value)	0.4406	0.3664	0.0001***	0.1031	0.6984	0.5026	0.2365
Moran's I (p-value)	0.0692**	0.0018***	0.0139**	0.0002***	0.0216**	0.000***	0.000***
LM (lag) (p-value)	0.9893	0.1063	0.3756	0.0042***	0.0457**	0.4782	0.2931
LM robus(lag) (p-value)	0.9155	0.0706*	0.2970	0.0023***	0.0346**	0.2919	0.1588
LM (error) (p-value)	0.2074	0.0131**	0.0642*	0.003***	0.0870*	0.0000***	0.0000***
LM robus (error) (p-value)	0.2058	0.0090***	0.0535*	0.0016***	0.0651*	0.0000***	0.0000***

Note: ***, ** and * are significant at 1%, 5% and 10% respectively; Standard Error is in parenthesis, and dependent variable is Lninc

Source: Author’s calculation from Labor Force Survey

Table 4.3

Standard model with a spatial specification testing of the spatial inverse distance 150 km weight matrix

	2001	2002	2003	2004	2005	2006	2007	2008
sch	0.2204*** (0.0280)	0.2252*** (0.0235)	0.2311*** (0.0235)	0.2135*** (0.0220)	0.1957*** (0.0236)	0.1490*** (0.0230)	0.1553*** (0.0166)	0.1685*** (0.0390)
exp	-0.0638 (0.1757)	0.0296 (0.1491)	0.1025 (0.1545)	0.0199 (0.1542)	0.1330 (0.1058)	0.0205 (0.1016)	0.1585 (0.1128)	0.4402* (0.2564)
expsq	0.0017 (0.0031)	0.0003 (0.0026)	-0.0012 (0.0027)	0.0001 (0.0027)	-0.0019 (0.0019)	-0.0001 (0.0018)	-0.0036 (0.0025)	-0.0097* (0.0055)
constant	7.9286*** (2.5879)	6.5953*** (2.1938)	5.5068** (2.2784)	6.8387*** (2.2723)	5.4504*** (1.5766)	7.4915*** (1.5360)	6.3880*** (1.2982)	3.2186 (3.0152)
R-squared	0.6633	0.7025	0.6972	0.7055	0.6323	0.5522	0.6074	0.2546
Log likelihood	50.5860	59.6830	60.4250	62.5930	53.4780	54.8580	59.1990	10.2010
AIC	-93.1720	-111.367	-112.851	-117.186	-98.9550	-101.717	-110.397	-12.4020
SIC	-83.8490	-102.0440	-103.5280	-107.8630	-89.6330	-92.3940	-101.0740	-3.0800
Jarque-Bera (p-value)	0.7906	0.8176	0.3678	0.5950	0.8801	0.9400	0.3520	0.0000***
Moran's I (p-value)	0.1696	0.0002***	0.0128**	0.0105**	0.0000***	0.0001***	0.0356**	0.2426
LM (lag) (p-value)	0.2562	0.0773*	0.2358	0.1793	0.0731*	0.0734*	0.8079	0.9847
LM robus (lag) (p-value)	0.4960	0.8115	0.9494	0.9160	0.0729*	0.2635	0.0811*	0.1953
LM (error) (p-value)	0.3416	0.0015***	0.0471**	0.0385**	0.0000***	0.0006***	0.1388	0.5371
LM robus (error) (p-value)	0.7789	0.0079***	0.1110	0.1147	0.0000***	0.0017***	0.0229**	0.1514

Table 4.3 (continued)

	2009	2010	2011	2012	2013	2014	2015
sch	0.2296*** (0.0201)	0.2026*** (0.0219)	0.1905*** (0.0240)	0.2133*** (0.0159)	0.1834*** (0.0165)	0.1229*** (0.0135)	0.1010*** (0.0156)
exp	0.3159** (0.1337)	0.4837*** (0.1481)	0.4701*** (0.1762)	0.2595* (0.1335)	0.4921*** (0.1267)	0.2085*** (0.0787)	0.2047** (0.0883)
expsq	-0.0067* (0.0029)	-0.0103*** (0.0031)	-0.0100*** (0.0037)	-0.0054* (0.0028)	-0.0101*** (0.0026)	-0.0037** (0.0015)	-0.0037** (0.0017)
constant	4.0938** (1.5821)	2.3490 (1.7544)	2.6080 (2.1040)	4.9367*** (1.6132)	2.2946 (1.5388)	5.8087*** (1.0187)	5.9865*** (1.1489)
R-squared	0.6587	0.5828	0.4898	0.7278	0.6528	0.5476	0.3757
Log likelihood	62.3490	58.3880	50.4020	70.7070	68.5550	61.7460	52.6370
AIC	-116.6990	-108.7760	-92.8050	-133.4140	-129.1090	-115.4930	-97.2740
SIC	-107.376	-99.453	-83.482	-124.091	-119.787	-106.17	-87.951
Jarque-Bera (p-value)	0.4406	0.3664	0.0001***	0.1031	0.6984	0.5026	0.2365
Moran's I (p-value)	0.1822	0.0029***	0.083*	0.0023***	0.0919*	0.0000***	0.0000***
LM (lag) (p-value)	0.4164	0.3629	0.6855	0.0599*	0.4672	0.1116	0.0112**
LM robus (lag) (p-value)	0.6682	0.3309	0.0512*	0.5342	0.9588	0.0798*	0.4186
LM (error) (p-value)	0.4483	0.0231**	0.2650	0.0217**	0.2857	0.0001***	0.0002***
LM robus (error) (p-value)	0.7540	0.0216**	0.0271**	0.1455	0.4332	0.0001***	0.0055***

Note: ***, ** and * are significant at 1%, 5% and 10% respectively; Standard Error is in parenthesis, and dependent variable is Lninc

Source: Author’s calculation from Labor Force Survey

4.3.2 Evidence of the spatial lag model

As mentioned earlier, the spatial lag model is significant in some years, such as in 2002, 2012, and 2013, under the spatial weight matrix of rook contiguity. However, the spatial lag effect under the spatial inverse distance definition is significant in more years.

Table 4.4 presents the results of the spatial lag model that uses the same spatial weight matrix as that in Table 4.2 (rook contiguity). In this case, W_LNINC is an endogenous explanatory variable that represents the natural logarithm of the average labor income of neighbors. The spatial lagged dependent variable has a positive effect on labor income before showing a negative relationship since 2008. However, this variable is statistically and significantly different from zero at the 0.05 level in some years, such as in 2002, 2012, and 2013. The severity of labor income externality from neighbor provinces has a low influence on the measurement of average labor income in a particular province. For example, if the average labor income in proximal areas increases by 100%, then the average income in a particular province significantly increases by only 2.9% in 2002 and decreases by 3% and 2.2% in 2012 and 2013, respectively.

The positive relationship between neighbor provinces in terms of income externality indicates that the labor income in a province is influenced by that in its neighbors. On the surface, this finding is consistent with theory of clustering growth. Surprisingly, the labor income in a particular province is negatively influenced by that in its neighbors since 2008. Such negative relationship reveals the scattered growth of neighbor provinces in terms of labor income. De Vreyer and Spielvogel (2005) offered an alternative explanation to these results by mentioning that average labor income in the province point out the labor productivity in a province. The productivity growth in one locality tends to be driven by the labor and capital in the neighboring localities. And then point toward the negative externality of labor income in these areas according central place theory which shows that labor and capital tend to be attracted to large cities or main nodes. In addition, the rate of return on an additional year of schooling is strongly significant and the rate of return ranges between 12% and 29% which is higher than that captured in the traditional Mincer model. The statistical significance of work experience Table 4.5 presents the results of the spatial lag model that are estimated by using the spatial distance band criterion at 150 km, which is the same weight matrix applied in Table 4.3. The table shows that the spatial lag effect can influence the use of data in 2002, 2005, 2006, 2012, and 2015. Therefore, the average labor income in a province is predominantly and positively influenced by the labor income in its neighbor provinces, except for the years 2007, 2008, and 2011, which were previously known as abnormal periods for the world economy and the Thai economy. The statistically significant results of the spatial lag dependent explanatory variable under a spatial inverse distance weight matrix confirm the presence of clustering growth in these provinces. One of these provinces has been spatially influenced by the positive labor income externality in the neighbors. Economic conditions, production technology, and other knowledge can be shared across provinces within an appropriate distance. The spillover effect also encourages a high labor productivity in the province (Maskell, 2001). This finding is consistent with the findings of theory of cluster growth or those of geographical cluster theory.

As shown in Table 4.5, the average financial return on education ranges between 8% and 22% per an additional year of schooling. This percentage is higher than that obtained by the traditional Mincer model (Table 4.3) because some omitted factors that can strongly affect the educational variables in the traditional Mincer model can be explained by the spatially lagged variables in this model. The labor force gives the highest economic return when the work experience of labors ranges between 24 years and 29 years depending on the year.

Table 4.4

Spatial lag model of rook contiguity

	2001	2002	2003	2004	2005	2006	2007	2008
w_lninc	0.0199 (0.0149)	0.0291** (0.0134)	0.0218* (0.0131)	0.0179 (0.0131)	0.0162 (0.0143)	0.0076 (0.0137)	0.0040 (0.0128)	-0.0095 (0.0243)
sch	0.2394*** (0.0280)	0.2228*** (0.0222)	0.2305*** (0.0225)	0.2115*** (0.0212)	0.1971*** (0.0228)	0.1491*** (0.0224)	0.1554*** (0.0161)	0.1680*** (0.0380)
exp	-0.1118 (0.1727)	-0.0568 (0.1468)	0.0486 (0.1514)	-0.0447 (0.1557)	0.1093 (0.1045)	0.0126 (0.0998)	0.1530 (0.1112)	0.4549* (0.2521)
expsq	0.0026 (0.0030)	0.0015 (0.0026)	-0.0003 (0.0027)	0.0013 (0.0027)	-0.0015 (0.0018)	0.00004 (0.0017)	-0.0035 (0.0024)	-0.0100* (0.0054)
constant	8.4670*** (2.5217)	7.5976*** (2.1272)	6.0983*** (2.2096)	7.6250*** (2.2607)	5.6366*** (1.5329)	7.5422*** (1.4956)	6.4174*** (1.2665)	3.1284 (2.9407)
R-squared	0.6699	0.7161	0.7053	0.7126	0.6344	0.5522	0.6075	0.2562
Log likelihood	51.4690	61.9930	61.7910	63.5150	54.1130	55.0110	59.2470	GMM
AIC	-92.9380	-113.9860	-113.5830	-117.0300	-98.2260	-100.0210	-108.4940	GMM
SIC	-81.2840	-102.3330	-101.9290	-105.3760	-86.5730	-88.3670	-96.8410	GMM

Table 4.4 (continued)

	2009	2010	2011	2012	2013	2014	2015
w_lninc	-0.0002 (0.0123)	-0.0206* (0.0125)	-0.0139 (0.0139)	-0.0303*** (0.0099)	-0.0217** (0.0106)	-0.0090 (0.0125)	-0.0153 (0.0142)
sch	0.2296*** (0.0196)	0.2026*** (0.0209)	0.1917*** (0.0233)	0.2897** (0.1231)	0.1842*** (0.0156)	0.1231*** (0.0131)	0.2411*** (0.0912)
exp	0.3162* (0.1323)	0.5164*** (0.1429)	0.4847*** (0.1713)	-0.0060** (0.0026)	0.5275*** (0.1211)	0.22878*** (0.0810)	-0.0044*** (0.0018)
expsq	-0.0067* (0.0028)	-0.0110*** (0.0030)	-0.0103*** (0.0036)	0.2167*** (0.0147)	-0.0108*** (0.0025)	-0.0041*** (0.0016)	0.1003*** (0.0151)
constant	4.0911*** (1.5497)	2.1318 (1.6821)	2.5440 (2.0384)	4.8063*** (1.4828)	2.0452 (1.4612)	5.6259*** (1.0169)	5.6556*** (1.1469)
R-squared	0.6587	0.6005	0.497	0.7609	0.673	0.5535	0.3908
Log likelihood	62.3490	59.7250	GMM	75.0670	70.6140	62.0020	53.2050
AIC	-114.6990	-109.449	GMM	-140.1330	-131.227	-114.004	-96.4090
SIC	-103.0450	-97.7950	GMM	-128.4800	-119.5740	-102.3500	-84.7560

Note: ***, ** and * are significant at 1%, 5% and 10% respectively; Standard Error is in parenthesis, and dependent variable is Lninc

Source: Author’s calculation from Labor Force Survey

Table 4.5

Spatial lag model of spatial inverse distance 150 km

	2001	2002	2003	2004	2005	2006	2007	2008
w_lninc	0.1246 (0.1121)	0.1890* (0.1021)	0.1261 (0.1077)	0.1602 (0.1030)	0.2225** (0.1080)	0.2374** (0.1182)	-0.0397 (0.1271)	-0.0273 (0.3435)
sch	0.2083*** (0.0291)	0.2085*** (0.0246)	0.2195*** (0.0248)	0.2038*** (0.022)	0.1840*** (0.0240)	0.1349*** (0.0233)	0.1582*** (0.0188)	0.1699*** (0.0417)
exp	-0.0508 (0.1696)	0.0624 (0.1422)	0.1220 (0.1494)	0.0463 (0.1488)	0.1176 (0.1007)	0.0109 (0.0963)	0.1674 (0.1116)	0.4473* (0.2654)
expsq	0.0014 (0.0030)	-0.0006 (0.0025)	-0.0016 (0.0026)	-0.0004 (0.0026)	-0.0017 (0.0018)	0.0003 (0.0017)	-0.0038 (0.0025)	-0.0099* (0.0057)
constant	6.7773** (2.6900)	4.6536** (2.2918)	4.2460* (2.4230)	5.1616** (2.4306)	3.7694** (1.6827)	5.6181*** (1.7287)	6.6278*** (1.5243)	3.3756 (3.5381)
R-squared	0.6646	0.6905	0.6925	0.6992	0.5949	0.5226	0.6111	0.2545
Log likelihood	51.1900	61.2130	61.0930	63.5800	55.1500	56.4860	59.2370	GMM
AIC	-92.3810	-112.4270	-112.1850	-117.1600	-100.3000	-102.9720	-108.4740	GMM
SIC	-80.7270	-100.7730	-100.5320	-105.5060	-88.6460	-91.3180	-96.8200	GMM

Table 4.5 (continued)

	2009	2010	2011	2012	2013	2014	2015
w_lninc	0.0988 (0.1117)	0.1315 (0.1210)	-0.3117 (0.3044)	0.1904* (0.0999)	0.0918 (0.1131)	0.2154* (0.1180)	0.3447*** (0.1213)
sch	0.2197*** (0.0221)	0.1950*** (0.0226)	0.2056*** (0.0277)	0.1907*** (0.0187)	0.1754*** (0.0183)	0.1112*** (0.0153)	0.0875*** (0.0162)
exp	0.2936** (0.1294)	0.4515*** (0.1439)	0.5326*** (0.1826)	0.2248* (0.1271)	0.4660*** (0.1240)	0.1589** (0.0762)	0.1125 (0.0823)
expsq	-0.0062** (0.0028)	-0.0097*** (0.0030)	-0.0113*** (0.0038)	-0.0047* (0.0027)	-0.0096*** (0.0026)	-0.0028* (0.0015)	-0.0019 (0.0016)
constant	3.5271** (1.8004)	1.5842 (1.9048)	4.6301 (2.8500)	3.7457** (1.7033)	1.8133 (1.6843)	4.5069*** (1.3030)	4.0429*** (1.4119)
R-squared	0.6584	0.5701	0.4865	0.7244	0.6508	0.5078	0.312
Log likelihood	62.6990	58.8590	GMM	72.3800	68.8410	63.0880	55.7510
AIC	-115.3980	-107.7180	GMM	-134.7590	-127.6820	-116.1750	-101.5020
SIC	-103.7440	-96.0650	GMM	-123.1060	-116.0290	-104.5220	-89.8480

Note: ***, ** and * are significant at 1%, 5% and 10% respectively; Standard Error is in parenthesis, and dependent variable is Lninc

Source: Author’s calculation from Labor Force Survey

4.3.3 Evidence form the spatial error model

The spatial error model in Tables 4.6 and 4.7 are estimated under the spatial rook contiguity definition and the spatial distance band criterion, respectively. The spatial weight for the model in Tables 4.6 and 4.7 is exactly the same as the spatial weight matrix in Tables 4.2 and 4.3, respectively.

The most important variable for the spatial error model is the spatial error coefficient (lambda), which is strongly significant either under the spatial rook contiguity or the spatial distance band criterion. However, this variable is insignificant for the rook spatial weight matrix in 2001 and 2008 and for the inverse distance weight matrix in 2001, 2008, and 2009 resulting from economic conditions and according to the spatial specification testing. Spatial error dependence is treated as a spatial nuisance in the use of data due to the omission of some unobserved spatial variables in nearby locations. Thus, the inferential errors from ignoring such effects should be assessed by using the spatial error model.

The spatial regression results highlight a spatial effect that is not induced by the average labor income externality in neighbors yet occurs on account of other reasons that are not examined in the traditional model (i.e., quality of infrastructure, quality of schools, and social values in a province) and other unobserved factors that are commonly found in nearby locations.

The spatial errors of the unaccounted variables in neighbor provinces are considered in the estimation. Table 4.6 shows that the financial rate of return on an additional year of schooling is positively and strongly significant. The rate of return ranges between 10% and 23%, which is close to the results of the traditional Mincer model. However, if the annual rate of return on schooling is considered, then the effect of education on the average return is slightly higher than that recorded in the traditional model for some years, such as in 2002, 2004–2007, 2010, 2011, and 2014–2015. Meanwhile, the work experience of labors in the province has a statistically significant effect on the rate of return in few years and has a statistically insignificant effect in many years. Those provinces which labors have 22 years to 33 years of average work experience will receive the highest returns.

Table 4.7 presents the robustness of the results of the spatial error model that are estimated by using the spatial distance band criterion at 150 km. Lambda is strongly significant as mentioned earlier. The rate of return on schooling ranges from 11% to 23% per an additional year of schooling depending on the year. A positive rate of return on schooling is higher than that in Table 4.6 in 2005, 2007, 2010, 2011, 2014, and 2015. The average years of work experience of labors in the province is statistically significant in few years, and those provinces which labors have 22 years to 33 years of average working experience obtain the highest returns.

These findings highlight the importance of the spatial influence that is treated as a nuisance in the use of spatial data due to the omission of some neighbor variables that strongly affect the economic activities in a particular province. The productivity and labor income in the province may be influenced by the spatial effect of other factors in neighbor areas that are not included in the model, such as quality of infrastructure and amenities. The abundance of living facilities and production facilities in the neighbor areas may also affect the labor productivity and income of a particular province. These observations may be explained by industrial location theory, which underscores the importance of neighboring areas as a source of raw materials and shipping goods (Hanink, 2016). The results of the spatial error model are consistent with those reported in previous studies that find a spatial error effect in geographical data, such as Kemeny and Storper (2012) and Gibson, Kim, and Olivia (2011).

Table 4.6

Spatial error model of rook contiguity

	2001	2002	2003	2004	2005	2006	2007	2008
lambda	0.1784 (0.1554)	0.4958*** (0.1200)	0.3844*** (0.1347)	0.3470** (0.1390)	0.5310*** (0.1148)	0.4564*** (0.1255)	0.2668* (0.1475)	0.0655 (0.1534)
sch	0.2195*** (0.0280)	0.2291*** (0.0229)	0.2309*** (0.0235)	0.2164*** (0.0213)	0.2157*** (0.0212)	0.1571*** (0.0216)	0.1593*** (0.0178)	0.1712*** (0.0390)
exp	-0.0620 (0.1698)	0.1254 (0.1310)	0.1553 (0.1432)	0.0231 (0.1438)	0.1911** (0.0904)	0.0596 (0.0908)	0.1650 (0.1121)	0.4208* (0.2551)
expsq	0.0017 (0.0030)	-0.0017 (0.0023)	-0.0022 (0.0025)	0.0002 (0.0025)	-0.0029* (0.0016)	-0.0008 (0.0016)	-0.0037 (0.0025)	-0.0093* (0.0055)
constant	7.9031*** (2.4981)	5.2132*** (1.9123)	4.7523** (2.1012)	6.7340*** (2.1090)	4.4449*** (1.3004)	6.8607*** (1.3343)	6.2934*** (1.2868)	3.4389 (2.9999)
R-squared	0.6633	0.7006	0.6965	0.7049	0.6306	0.5511	0.6074	0.2545
Log likelihood	51.1670	64.9010	63.6000	65.1380	61.4490	60.1810	60.6260	GMM
AIC	-94.3340	-121.8020	-119.2010	-122.2750	-114.8990	-112.3610	-113.2520	GMM
SIC	-85.0110	-112.4790	-109.8780	-112.9520	-105.5760	-103.0380	-103.9290	GMM

Table 4.6 (continued)

	2009	2010	2011	2012	2013	2014	2015
lambda	0.2681* (0.1473)	0.4335*** (0.1285)	0.2792** (0.1426)	0.7144*** (0.0830)	0.3557*** (0.1380)	0.5413*** (0.1133)	0.5585*** (0.1106)
sch	0.2248*** (0.0215)	0.2045*** (0.0222)	0.1967*** (0.0251)	0.1945*** (0.0172)	0.1711*** (0.0173)	0.1268*** (0.0141)	0.1081*** (0.0166)
exp	0.1981 (0.1404)	0.3820** (0.1626)	0.4175** (0.1874)	0.1171 (0.1298)	0.3786*** (0.1344)	0.1000 (0.0829)	0.0464 (0.0913)
expsq	-0.0042 (0.0030)	-0.0083** (0.0034)	-0.0089** (0.0040)	-0.0029 (0.0027)	-0.0079*** (0.0028)	-0.0016 (0.0016)	-0.0006 (0.0018)
constant	5.5219*** (1.6568)	3.6102* (1.9154)	3.2045 (2.2338)	7.0343*** (1.5579)	3.7784** (1.6290)	7.1570*** (1.0570)	7.9617*** (1.1738)
R-squared	0.6541	0.5752	0.4888	0.6702	0.647	0.5339	0.345
Log likelihood	63.4040	61.9060	GMM	79.1490	70.5370	70.5930	60.9550
AIC	-118.8080	-115.8130	GMM	-150.2990	-133.0730	-133.1870	-113.9100
SIC	-109.4860	-106.4900	GMM	-140.9760	-123.7500	-123.8640	-104.5870

Note: ***, ** and * are significant at 1%, 5% and 10% respectively; Standard Error is in parenthesis, and dependent variable is Lninc

Source: Author’s calculation from Labor Force Survey

Table 4.7

Spatial error model of spatial inverse distance 150 km

	2001	2002	2003	2004	2005	2006	2007	2008
lambda	0.1478 (0.1706)	0.4488*** (0.1345)	0.2637* (0.1588)	0.3253** (0.1515)	0.5687*** (0.1148)	0.4724*** (0.1309)	0.2665* (0.1585)	0.1290 (0.1832)
sch	0.2187*** (0.0276)	0.2251*** (0.0226)	0.2295*** (0.0233)	0.2112*** (0.0211)	0.2093*** (0.0208)	0.1489*** (0.0214)	0.1586*** (0.0176)	0.1725*** (0.0398)
exp	-0.0497 (0.1694)	0.0952 (0.1332)	0.1284 (0.1470)	0.0310 (0.1455)	0.1548* (0.0921)	0.0286 (0.0932)	0.1360 (0.1152)	0.3959 (0.2590)
expsq	0.0015 (0.0030)	-0.0012 (0.0024)	-0.0017 (0.00258)	-0.00008 (0.0026)	-0.0024 (0.0017)	-0.0004 (0.0017)	-0.0031 (0.0025)	-0.0088 (0.0056)
constant	7.7535*** (2.4917)	5.7420*** (1.9438)	5.1821** (2.1588)	6.7057*** (2.1341)	5.0909*** (1.3331)	7.4389*** (1.3771)	6.6536*** (1.3229)	3.7180 (3.0467)
R-squared	0.6632	0.6996	0.6966	0.7055	0.6313	0.5499	0.6065	0.2541
Log likelihood	50.9770	64.0170	61.9160	64.5310	62.0350	59.7850	60.3250	GMM
AIC	-93.9530	-120.0330	-115.8320	-121.0620	-116.0700	-111.570	-112.6490	GMM
SIC	-84.6300	-110.7100	-106.5090	-111.7390	-106.7470	-102.2470	-103.3260	GMM

Table 4.7 (continued)

	2009	2010	2011	2012	2013	2014	2015
lambda	0.2030 (0.1653)	0.4236*** (0.1382)	0.1972 (0.1543)	0.6061*** (0.1080)	0.2738* (0.1577)	0.5437*** (0.1192)	0.5712*** (0.1144)
sch	0.2251*** (0.0209)	0.2063*** (0.0223)	0.1981*** (0.0246)	0.1914*** (0.0178)	0.1789*** (0.0175)	0.1347*** (0.0146)	0.1122*** (0.0165)
exp	0.2137 (0.1385)	0.3778** (0.1618)	0.4383** (0.1839)	0.1336 (0.1391)	0.4149*** (0.1331)	0.0828 (0.0808)	-0.0092 (0.0881)
expsq	-0.0046 (0.0030)	-0.0082** (0.0034)	-0.0093** (0.0039)	-0.0032 (0.0029)	-0.0086*** (0.0028)	-0.0013 (0.0016)	0.0004 (0.0017)
constant	5.3355*** (1.6377)	3.6595* (1.9165)	2.9572 (2.1961)	6.7959*** (1.6856)	3.2979** (1.6200)	7.3581*** (1.0391)	8.6840*** (1.1422)
R-squared	0.6551	0.5729	0.4891	0.68	0.6484	0.5246	0.3188
Log likelihood	62.7830	61.2520	GMM	75.2720	69.3790	68.9530	60.1000
AIC	-117.5650	-114.5030	GMM	-142.5440	-130.7590	-129.9070	-112.2000
SIC	-108.2430	-105.1800	GMM	-133.2210	-121.4360	-120.5840	-102.8770

Note: ***, ** and * are significant at 1%, 5% and 10% respectively; Standard Error is in parenthesis, and dependent variable is Lninc

Source: Author's calculation from Labor Force Survey

4.3.4 Evidence from the spatial durbin model

As mentioned earlier, spatial dependence does not exert its influence only via the neighbor average income variable but also via the average schooling years and work experience of labors in neighboring provinces. If they are common factors, then the spatial durbin model is equivalent to the spatial error model as shown in Chapter 3.

Tables 4.8 and 4.9 present the results of the spatial durbin model that uses rook contiguity and inverse distance band criterion in the spatial weight matrix, respectively. The most important variables in this model include the spatial lagged explanatory variables of the neighbor, namely, w_sch, w_exp, and w_expsq, which represent the average schooling years of laborers in neighbor provinces, the average years of work experience of laborers in the neighbor provinces, and work experience squared, respectively. W_LNINC is an endogenous explanatory variable that represents the natural logarithm of the average labor income in neighbors. The lagged labor income of neighbors has a significant positive effect on the model for each year except in 2001, 2007–2009, 2011, and 2013 for rook contiguity and has an insignificant effect in 2001, 2003, 2007–2009, 2011, and 2013 for spatial inverse distance band. For spatial rook contiguity, the significant elasticity of the labor income ranges from 0.27 to 0.5, which indicates that a 100% increase in the average labor income in neighbors can increase the average labor income in the host province by 27% to 50% depending on the year. For the spatial inverse distance band, the externality of labor income has a significant elasticity of 0.25 to 0.53 depending on the year.

An increase in the average schooling years in the neighbor provinces also negatively affects the average labor income in a particular province. However, such effect is insignificant in several years according to Rodríguez-Pose and Vilalta-Bufí (2005). An additional average year of schooling in the neighbor areas reduces the financial labor returns by 5% to 17% under the spatial rook weight matrix, and such effect is slightly lower under the spatial inverse distance band. An additional year of schooling also reduces the average labor income in a particular province by 5% to 13% depending on the year. In addition, the years of work experience in the neighbor does not help increase the labor income in the host province.

The negative schooling externality can be attributed to two reasons. On the one hand, the negative effect of schooling years in nearby locations on the labor income in a particular province indicates the existence of a displacing effect, in which the high value added activities move away from the province to the surrounding areas where highly educated workers are concentrated because an increase in the average schooling year in the neighbor implies that the neighbor has a pool of highly educated workers (Morone, 2013). Such negative effect can lead to the local displacement of jobs and economic activities across the treatment area in the clustering neighbor province. However, if the economic activity cannot be easily shifted, then the labor income in the host province tends to drive the highly educated workers into moving and concentrating in some other provinces. In this case, the neighbors begin to pool lowly educated workers, thereby resulting in the segregation of spatial income in the clustering area. The concentration of highly educated workers may increase the wage premium similar to the case of the US and to the findings of Group (2017) and Rabinovitz (2016). This observation is consistent with many geographical theories, such as industrial location theory, theory of clustering growth, and central place theory.

On the other hand, the negative effect of schooling externality may imply a shortage of job vacancies and job creation for the highly educated workers in the cluster provinces because an increase in the average years of schooling implies the concentration of labors with more years of schooling. Both the oversupply of and the limited demand for highly educated laborers can reduce the labor income in the host province because these types of workers can be easily replaced by the labors in the neighbor provinces. These reasons can be also used to explain why the work experience of labors in the neighbors takes a negative sign.

The rate of return on an additional year of schooling is positively significant and ranges from 12% to 14% under the spatial rook contiguity and from 13% to 24% under the spatial distance of the spatial weight matrix. The average work experience is statistically significant after 2008 under the spatial rook contiguity but is significant only in 2013 under the spatial distance band. A province obtains the highest average labor income if the average work experience of its labors ranges between 23 years and 27 years under the spatial rook and 23 years under the spatial distance band criterion of spatial weight matrix.

Table 4.8

Spatial durbin model of rook contiguity

	2001	2002	2003	2004	2005	2006	2007	2008
w_lninc	0.1757 (0.1510)	0.4802*** (0.1190)	0.3773*** (0.1331)	0.3114** (0.1381)	0.4994*** (0.1154)	0.4528*** (0.1188)	0.1705 (0.1388)	-0.0168 (0.1651)
sch	0.2197*** (0.0326)	0.2294*** (0.0252)	0.2291*** (0.0269)	0.2208*** (0.0226)	0.2230*** (0.0235)	0.1530*** (0.0252)	0.1870*** (0.0248)	0.2292*** (0.0532)
exp	-0.1117 (0.1740)	0.0147 (0.1381)	0.1009 (0.1472)	-0.0484 (0.1560)	0.1067 (0.0982)	0.0031 (0.1014)	0.1750 (0.1119)	0.3730 (0.2515)
expsq	0.0026 (0.0031)	0.0003 (0.0024)	-0.0012 (0.0026)	0.0014 (0.0027)	-0.0014 (0.0018)	0.0002 (0.0018)	-0.0040 (0.0025)	-0.0090* (0.0054)
w_sch	-0.0501 (0.0606)	-0.1322*** (0.0469)	-0.0915* (0.0513)	-0.0987 (0.0470)	-0.1772*** (0.0462)	-0.1204** (0.0473)	-0.0871** (0.0434)	-0.1297 (0.0867)
w_exp	-0.0674 (0.0787)	-0.2108*** (0.0638)	-0.1698** (0.0698)	-0.1203 (0.0734)	-0.1954*** (0.0665)	-0.2080*** (0.0692)	-0.0900 (0.1036)	0.0340 (0.1287)
w_expsq	0.0010 (0.0014)	0.0034*** (0.0012)	0.0027** (0.0013)	0.0017 (0.0013)	0.0029** (0.0013)	0.0033** (0.0013)	0.0019 (0.0023)	0.0008 (0.0030)
constant	8.4361*** (2.5104)	6.5809*** (1.9627)	5.3904** (2.1110)	7.5502*** (2.2343)	5.4412*** (1.3845)	7.6612*** (1.4670)	5.9971*** (1.2682)	3.9878 (2.9076)
R-squared	0.6781	0.7237	0.707	0.7127	0.6672	0.641	0.6246	0.2928
Log likelihood	52.0280	67.2160	64.8720	66.1860	63.5390	61.1990	61.5110	GMM
AIC	-88.0560	-118.4310	-113.7440	-116.3720	-111.0780	-106.3980	-107.0220	GMM
SIC	-69.4100	-99.7850	-95.0980	-97.7260	-92.4320	-87.7520	-88.3760	GMM

Table 4.8 (continued)

	2009	2010	2011	2012	2013	2014	2015
w_lninc	0.0972 (0.1550)	0.3619*** (0.1358)	0.7979 (0.5417)	0.3714*** (0.1237)	0.2218 (0.1483)	0.3544*** (0.1226)	0.2723** (0.1317)
sch	0.2418*** (0.0282)	0.2033*** (0.0233)	0.2071*** (0.0320)	0.1971*** (0.0178)	0.1620*** (0.0181)	0.1391*** (0.0171)	0.1281*** (0.0211)
exp	0.2938** (0.1362)	0.5254*** (0.1376)	0.5169*** (0.1790)	0.3457*** (0.1088)	0.5354*** (0.1166)	0.2707*** (0.0852)	0.2483*** (0.0951)
expsq	-0.0065** (0.0029)	-0.0114*** (0.0029)	-0.0110*** (0.0037)	-0.0078*** (0.0023)	-0.0112*** (0.0024)	-0.0050*** (0.0017)	-0.0045** (0.0019)
w_sch	-0.0513 (0.0574)	-0.0746 (0.0485)	-0.1716 (0.2171)	-0.0501 (0.0390)	0.0062 (0.0412)	-0.0752** (0.0308)	-0.0831** (0.0352)
w_exp	-0.0672 (0.1075)	-0.2755*** (0.0958)	-0.5541** (0.2772)	-0.3218*** (0.0838)	-0.2084** (0.1022)	-0.2307*** (0.0849)	-0.1653* (0.0930)
w_expsq	0.0018 (0.0024)	0.0062*** (0.0021)	0.0118 (0.0118)	0.0075*** (0.0018)	0.0047** (0.0021)	0.0044*** (0.0017)	0.0031* (0.0018)
constant	4.4056*** (1.5705)	2.1517 (1.6131)	2.1038 (2.1175)	4.5519*** (1.3022)	2.2115 (1.4005)	5.0459*** (1.0476)	5.3887*** (1.1723)
R-squared	0.6638	0.6033	0.5100	0.7971	0.6898	0.5319	0.3822
Log likelihood	63.5620	63.1520	GMM	85.4820	74.2370	67.2680	57.4400
AIC	-111.1230	-110.3050	GMM	-154.9630	-132.4740	-118.5360	-98.8800
SIC	-92.4770	-91.6590	GMM	-136.3170	-113.8280	-99.890	-80.2340

Note: ***, ** and * are significant at 1%, 5% and 10% respectively; Standard Error is in parenthesis, and dependent variable is Lninc

Source: Author’s calculation from Labor Force Survey

Table 4.9

Spatial durbin model of spatial inverse distance 150 km

	2001	2002	2003	2004	2005	2006	2007	2008
w_lninc	0.1287 (0.1712)	0.3662** (0.1447)	0.2155 (0.1632)	0.3163** (0.1523)	0.5332*** (0.1204)	0.4524*** (0.1335)	0.1316 (0.1676)	0.0045 (0.1759)
sch	0.2027*** (0.0304)	0.2156*** (0.0242)	0.2219*** (0.0255)	0.2049*** (0.0226)	0.2139*** (0.0229)	0.1598*** (0.0240)	0.1776*** (0.0200)	0.2278*** (0.0448)
exp	-0.0452 (0.1684)	0.0721 (0.1330)	0.1337 (0.1464)	0.0350 (0.1472)	0.1688* (0.0991)	0.0664 (0.0987)	0.1585 (0.1196)	0.0095 (0.2828)
expsq	0.0012 (0.0030)	-0.0009 (0.0024)	-0.0020 (0.0026)	-0.0002 (0.0026)	-0.0026 (0.0018)	-0.0010 (0.0017)	-0.0038 (0.0027)	-0.0007 (0.0060)
w_sch	-0.0496 (0.0672)	-0.1291** (0.0510)	-0.0797 (0.0570)	-0.0756 (0.0493)	-0.1288*** (0.0490)	-0.0504 (0.0513)	-0.0524 (0.0362)	-0.1255* (0.0693)
w_exp	-0.4599 (0.4020)	-0.5826* (0.3191)	-0.4546 (0.3406)	-0.2456 (0.3204)	-0.1019 (0.2464)	0.0774 (0.2489)	0.2372 (0.2008)	0.9280** (0.4464)
w_expsq	0.0083 (0.0071)	0.0104* (0.0056)	0.0081 (0.0060)	0.0042 (0.0057)	0.0017 (0.0042)	-0.0011 (0.0042)	-0.0048 (0.0044)	-0.0196** (0.0096)
constant	13.3849** (6.6695)	11.8660** (5.2601)	10.1354* (5.5005)	7.9317 (5.1613)	2.4971 (4.2699)	1.8852 (4.4214)	2.5788 (2.2152)	-2.3241 (4.3206)
R-squared	0.6738	0.7351	0.7127	0.7086	0.6540	0.5707	0.6493	0.3323
Log likelihood	52.0950	66.1680	63.1620	64.8660	62.6520	60.6700	63.4190	GMM
AIC	-88.1910	-116.3360	-110.3230	-113.7320	-109.3050	-105.3390	-110.8370	GMM
SIC	-69.5450	-97.6900	-91.6770	-95.0860	-90.6590	-86.6930	-92.1910	GMM

Table 4.9 (continued)

	2009	2010	2011	2012	2013	2014	2015
w_lninc	-0.0448 (0.1795)	0.3188** (0.1499)	0.0189 (0.1758)	0.3847*** (0.1403)	0.1486 (0.1680)	0.3002** (0.1470)	0.2464* (0.1492)
sch	0.2394*** (0.0221)	0.2155*** (0.0227)	0.2282*** (0.0264)	0.1954*** (0.0181)	0.1818*** (0.0191)	0.1504*** (0.0151)	0.1305*** (0.0163)
exp	-0.0253 (0.14875)	0.2748 (0.1770)	0.2438 (0.2124)	0.1164 (0.1363)	0.2943** (0.1424)	0.0616 (0.0805)	-0.0232 (0.0839)
expsq	0.0004 (0.0032)	-0.0063* (0.0037)	-0.0055 (0.0044)	-0.0030 (0.0028)	-0.0064** (0.0029)	-0.0009 (0.0016)	0.0008 (0.0016)
w_sch	0.0086 (0.0512)	-0.0975** (0.0458)	-0.1051** (0.0491)	-0.0564 (0.0398)	-0.0281 (0.0407)	-0.0808*** (0.0250)	-0.0779*** (0.0247)
w_exp	0.8504*** (0.2377)	0.0933 (0.2642)	0.3306 (0.3155)	0.0585 (0.1992)	0.2723 (0.2262)	0.2546** (0.1239)	0.4756*** (0.1320)
w_expsq	-0.0178*** (0.0051)	-0.0014 (0.0056)	-0.0066 (0.0066)	-0.0004 (0.0042)	-0.0052 (0.0047)	-0.0050** (0.0024)	-0.0093*** (0.0026)
constant	-1.5858 (2.1176)	1.1848 (2.3392)	1.4940 (2.8390)	2.7063 (2.0685)	0.1194 (2.0956)	1.9955 (1.3544)	0.8552 (1.3920)
R-squared	0.717	0.6263	0.5433	0.7769	0.6917	0.6728	0.5897
Log likelihood	69.661	63.454	GMM	80.533	72.929	73.792	68.181
AIC	-123.322	-110.908	GMM	-145.066	-129.858	-131.585	-120.361
SIC	-104.676	-92.262	GMM	-126.42	-111.212	-112.939	-101.716

Note: ***, ** and * are significant at 1%, 5% and 10% respectively; Standard Error is in parenthesis, and dependent variable is Lninc

Source: Author's calculation from Labor Force Survey

4.3.5 The model comparison

This part compares the rate of return on an additional average year of schooling in each provincial-level model to show how the spatial dimension affects the economic returns to human capital. The differences between each spatial model and the traditional model are also explained.

Before analyzing the key findings, a criterion for model selection should be established. The Akaike information criterion (AIC) and Schwarz information criterion (SIC) are used in this thesis to measure the relative quality of the model. A smaller AIC and SIC indicate that the model has a better fit to the spatial data. Therefore, the model with the smallest AIC and SIC must be selected. However, these criteria only identify the model with the best fit to the truth, and the results do not necessarily indicate that the other useful models that can express spatial relationships in other ways must be eliminated.

In regression analysis, the spatial error model provides the lowest AIC and SIC in both types of the spatial weight matrix and among all models examined in this thesis, including the traditional Mincer model, the spatial lagged model, and the spatial durbin model, for each year. As in the spatial autocorrelation test, the spatial error models which neighbor set has been defined by spatial rook contiguity have a better fit to the data than models that have been defined by spatial distance band criterion.

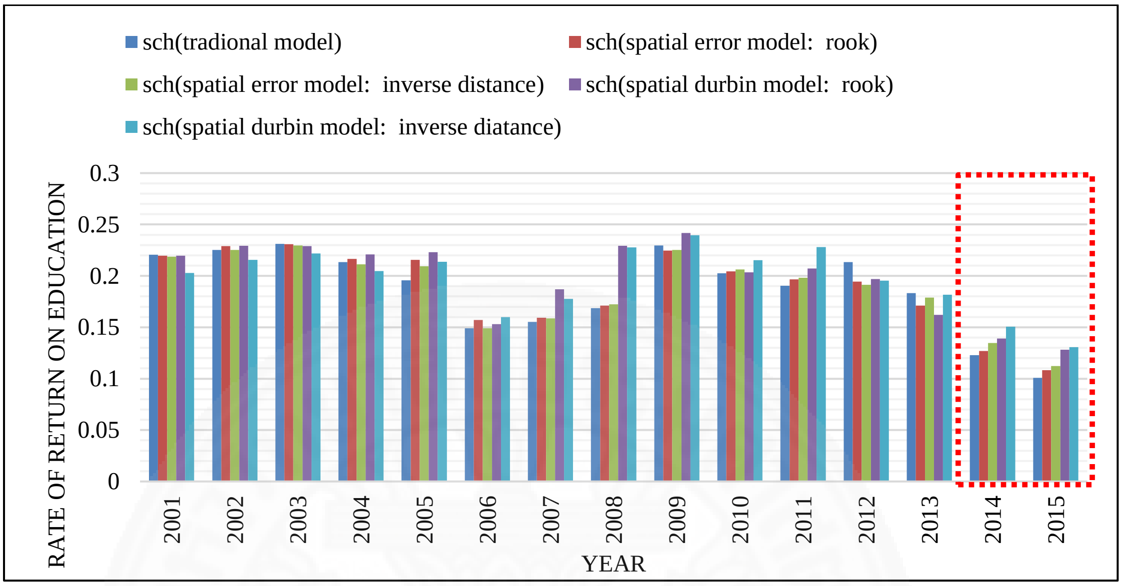
Three main findings are obtained from the model comparison. First, the spatial specification testing and spatial regression results are consistent and altogether confirm the existence of a spatial influence. The strongly significant spatial variable in the model also indicates a strong spatial interaction among neighbor provinces within a space. Second, the spatial error model is considered most appropriate for analyzing and describing the rate of return on human capital at the provincial level. This finding echoes the arguments of many studies which find that the spatial error model provides the best fit to the dataset (Brunow & Hirte, 2009; Jiafeng, 2014; Silveira-Neto & Azzoni, 2006). Third, the model comparison generates mixed results, and the positive effect of schooling years is always significant and robust in every model.

Figure 4.9 shows the rate of return on average years of schooling in the province as estimated by different models. Given that the spatial error derived from the spatial error model is significant in more years compared with those derived from the spatial lag effect and spatial durbin models, Figure 4.9 only shows the value of parameters derived from the traditional Mincer model, spatial error model, and spatial durbin model under the spatial rook and spatial inverse distance weight matrix. The rate of return generated in each model moderately differs from that generated in the traditional model, except for the agglomeration effect from the spatial model. Figure 4.9 also shows that the changes in the economic return on average schooling years differ every year. However, spatial influence does not statistically exist in some years if the neighbor set is defined by the spatial rook (i.e., in 2001 and 2008) and by the spatial distance band (i.e., in 2001, 2008, and 2009). Therefore, although schooling year is statistically significant in the spatial models, using another type of model in certain years would be more appropriate (i.e., using only the traditional model in 2001 and 2008, and using the traditional model and the spatial model under the spatial rook weight matrix in 2009). Figure 4.9 also reveals that the rate of return on schooling in Thailand follows a decreasing trend, especially after 2012.

The relationship between geographical influence and economic and social conditions warrants further discussion. The 2008 global financial crisis and 2011 great flood highlight that the status of spatial relationships is affected by both domestic and international economic conditions. Economic conditions have also changed the error term distribution in 2008 and 2011 from normal to non-normal (i.e., the skewness of the distribution is not equal to zero), thereby necessitating the adoption of other GMM estimation techniques, such as two stages least square for the spatial lag model and spatial weighted least square. However, the spatial omitted variable indicates that 2001 and 2008 do not have any spatial influence. In other words, the 2008 global financial crisis may have affected the target industries as mentioned earlier. Industry adaptation does not show any spatial effect in 2008 because geographical externality follows a random pattern during the crisis. The labor income and other spatial factors in the neighbor areas do not have any influence on the average labor income in another province.

Figure 4.9

Rate of return on education in each model



Source: Author's calculation from Labor Force Survey

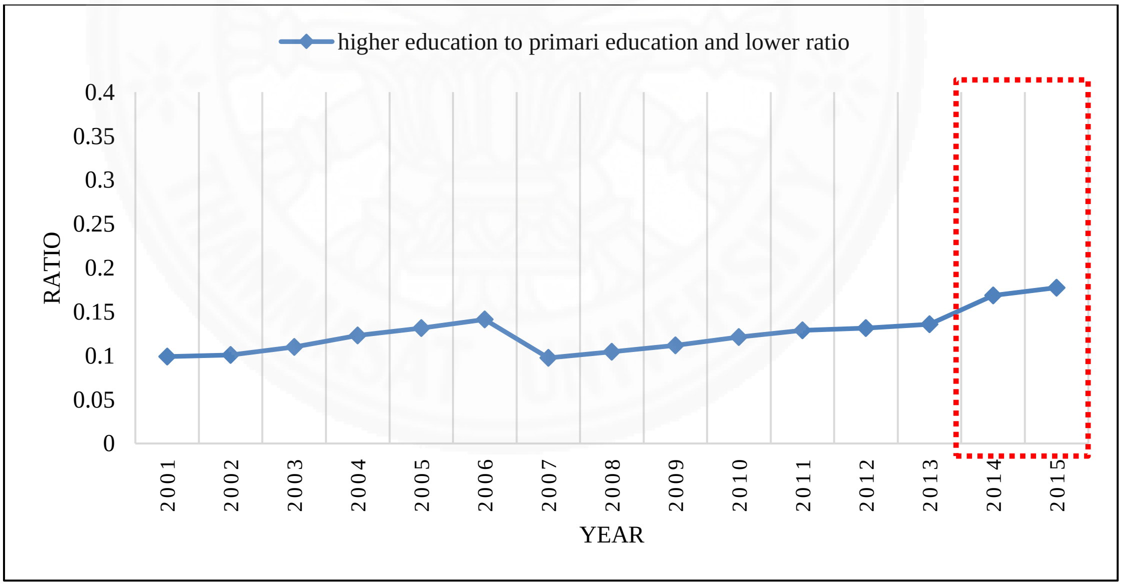
The rate of return on schooling in Thailand shows a decreasing trend, especially after 2012. Such trend may be explained by three main linked factors, namely, the minimum wage policy, the increasing supply of tertiary-educated workers, and the decreasing income gap among workers with different education levels. These reasons are linked and can be explained altogether.

First, Thailand has implemented a minimum wage policy in early 2013 during which the price floor was set to 300 baht at a flat rate. This policy specifically targeted the lowly educated workers instead of the highly educated ones, thereby explaining the increase in minimum wage for certain types of workers. As the wages for the lowly educated workforce are increased while those for the other workers remain the same, the rate of return on an additional year of schooling decreased after 2012.

Second, workers aim to find a secure job by investing in a higher level of education. Figure 4.10 shows that the ratio of tertiary-educated workers to primary-educated workers and lower has increased from 2001 to 2015 because of the increasing supply of university graduates and the decreasing supply of lowly educated workers. This ratio decreased in 2007 following a change in the survey methodology. The supply of university graduates in Thailand also outsized the demand for highly educated workers. The increasing supply of highly educated workers has been more than compensated by the increasing demand for skills. Either an increase in the supply of highly educated workers or the tight labor market of lowly educated workers has decreased the rate of return on schooling because the labor income for highly educated workers has not changed much yet that of the lowly educated workers has increased at high rates. Consequently, the rate of return on an additional year of schooling has decreased over time. This argument supports the reason mentioned in the previous paragraph and is consistent with the findings of previous studies, such as Fersterer and Winter-Ebmer (2003), Senkrua (2015), and Paweenawat and Vechbanyongratana (2015).

Figure 4.10

The ratio of highly educated workers to lowly educated workers

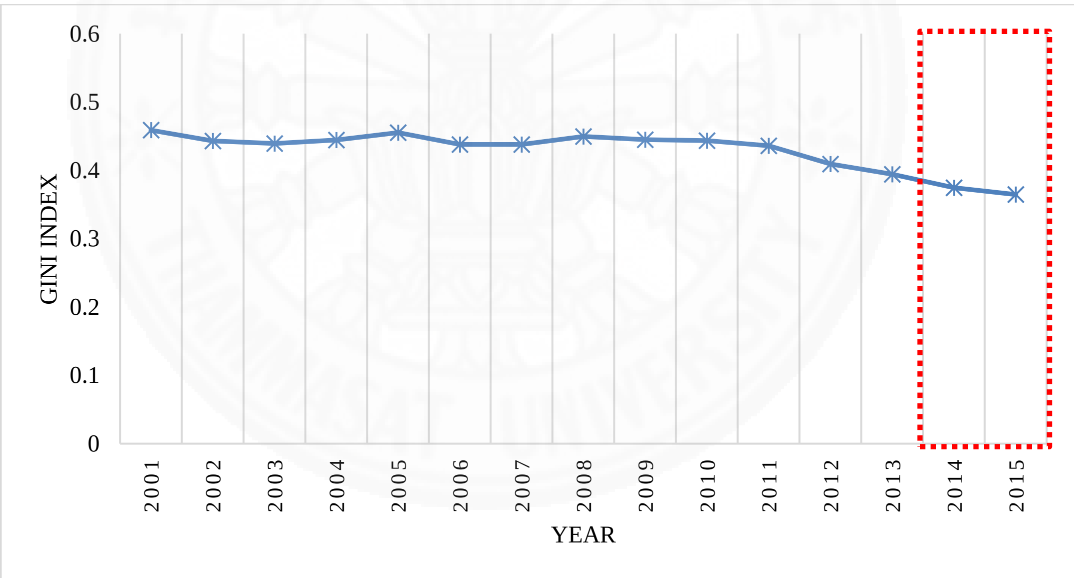


Source: Author’s calculation from Labor Force Survey

Finally, the continuous decrease in the return on an additional year of schooling also corresponds to the decreasing income gap among workers with different education levels. This observation also reflects the decreasing income inequality in Thailand. The income distribution in a country is generally measured by the Gini index with a value ranging between 0 and 1, where 0 indicates perfect income equality and 1 indicates maximum income inequality. Although the underlying reasons for decreasing income inequality are not apparent, the empirical evidence obtained by the Gini index by using the data from the Thai Labor Force Survey suggests that the efforts to reduce inequality in Thailand are effective and successful. Figure 4.11 shows that the Gini coefficient tends to decrease especially after 2012. The continuous reduction in labor income inequality or income gap as measured by the Gini index confirms the reduction in the rate of return on schooling for workers with different education levels. However, the actual cause for such declining returns may also be correlated with the two reasons mentioned in the previous paragraphs.

Figure 4.11

Gini coefficient of the labor income of the whole population



Source: Author’s calculation from Labor Force Survey

CHAPTER 5
CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

Studies in Thailand on the effect of human capital have not considered the influence of spatial relationship. Therefore, this thesis aims to analyze the effect of human capital on income in Thailand with the inclusion of spatial influence by combining the data from GIS with those from the Thai Labor Force Surveys of years 2001 to 2015.

Six main findings are obtained from the regression at the individual or provincial level using both the standard and spatial Mincer models.

First, the estimations show that the mechanism of income determination in Thailand is compatible with Mincer's theory. The key influencing factors that determine labor income include years of education and work experience, with education showing a strong significance in each model and year. The results of the standard Mincer model at the individual level without controlling any variables are similar to those reported in the literature.

Second, workplace location affects the returns for workers. Those provinces in Thailand with a high labor income are spatially clustered together around the Bangkok Metropolitan and proximal areas, the industrial estate provinces, and some provinces in the south. The workers from several northeastern provinces earn the lowest income among all quantile groups. Those provinces with high-value labor income are spatially surrounded by neighbors with similarly high-value labor income and are spatially clustered together around Bangkok Metropolitan. However, those provinces which laborers have the same work experience are spatially clustered in some provinces in northern and central Thailand. Those provinces with highly educated workers are concentrated in Bangkok and its surrounding areas because those major sectors with high value-added production need highly educated workers, require excellent infrastructure and facilities, and are mostly located in Bangkok Metropolitan, its neighboring areas, and industrial estate provinces. In this case, the workers in these provinces receive high returns and may motivate other highly educated workers in Thailand to migrate to these areas. Either the high value-added businesses or highly educated workers may benefit from such concentration via the positive externality or the neighboring spillover.

Third, spatial influence does matter. The criteria for model selection indicate that spatial effect influences the spatial model by acting as a nuisance in the use of data. The spatial error model shows the best fit if both AIC and SIC are used in the model selection. The labor income in a province may be influenced by the spatial effect of other factors in neighbor areas. Therefore, spatial autocorrelation does not affect neighbor income externality directly but rather through other factors that have not been considered in the examined model, including infrastructure, utilities, inconvenience of the area, and education quality of the workforce in each location and the proximal areas.

Fourth, the spatial durbin model shows that the average years of schooling in nearby areas negatively affects the average labor income in the host province. Such negative effect indicates the existence of a displacing effect in which activities with high returns move to surrounding areas in search for highly educated workers. Consequentially, highly educated workers migrate to other locations until they become concentrated together in some location. Such negative effect may also imply the shortage of jobs for highly educated workers in the clustering areas in Thailand. Nevertheless, such effect is small and insignificant in several years.

Fifth, spatial dependence is most likely affected by either domestic or international economic conditions. This finding also coincides with the changes in the economic activities in Thailand during the global financial crisis. Specifically, in 2008, the high labor income provinces in Thailand became less spatially clustered and were surrounded by neighbors with random workforce characteristics because the economic conditions at that time primarily influenced the target industry which is the main exporting industry in Thailand. Economic conditions also produce a spillover effect on labor income in a province and its neighbors because industries must reduce their production capacity and exports in response to a low international demand. Therefore, industries must learn to effectively adapt to the economic conditions. An insufficient spillover effect is also observed during a crisis.

Sixth, the decreasing rate of return on schooling in Thailand after 2012 may be attributed to three factors, namely, the minimum wage policy, the increasing supply of highly educated workers, and the decreasing wage gap among workers with different education levels.

These findings confirm the importance of geographical influence in determining the returns to human capital in Thailand. These studies show that apart from education and work experience, geographical externality may also affect labor income. Consistent with many geographical theories such as the theory of spatial clustering growth, central place theory, and industrial location theory, these findings point toward a spatial effect between neighboring provinces.

5.2 Policy implications

The spatial analysis of the effects of human capital on income sheds light on a few issues related to infrastructure and urban planning. The following policy implications focus on the spatial impact in the clustering zones in Thailand.

Years of schooling remain a key determinant of labor income. Therefore, both the quality and quantity of education must be promoted equally throughout the whole country. The geographical effect does matter in this case. Some empirical evidence reveals the spatial impact of neighbors. Promoting super cluster economic zones can encourage economic growth and result in the spatial concentration of highly educated workers in these clusters, thereby increasing labor income and producing a positive spillover effect to other areas. However, locating the cluster zone near Bangkok Metropolitan areas, such as the Eastern Economic Corridor, may result in a monocentric growth in Thailand. Alternatively, super cluster economic zones may be built in other regions to promote a multicenter development by taking advantage of the spatial effect and spatial spillover from the cluster. This strategy can encourage an equal development and promote high labor income in other provinces as a result of the positive spillover in the clusters.

5.3 Limitations and recommendation for the future study

Future studies may consider the 1997 Thailand financial crisis to further show the impact of economic conditions on spatial dependence. They may also devise a new estimation method by using individual data to measure spatial impact and to confirm the influence of area on labor income. Future studies must also examine the factors that underlie the region-specific influences on the wage premium in Bangkok and other main provinces in Thailand. The actual cause of the spatial effect that represents the spatial error model in this thesis also presents an interesting research direction.

REFERENCES

Books and Book Articles

Anselin, L. (2007). Spatial Econometrics *A Companion to Theoretical Econometrics* (pp. 310-330): Blackwell Publishing Ltd.

Anselin, L., & Rey, S. J. (2014). *Modern Spatial Econometrics in Practice: A Guide to Geoda, Geodaspace and Pysal*: GeoDa Press LLC.

Hanink, D. M. (2016). Industrial Location Theory *International Encyclopedia of Geography: People, the Earth, Environment and Technology*: John Wiley & Sons, Ltd.

Hutchison, R. (2009). *Encyclopedia of urban studies*: Sage Publications.

Krongkaew, M., Chamnivickorn, S., & Nitithanprapas, I. (2006). *Economic growth, employment, and poverty reduction linkages: The case of Thailand*: Recovery and Reconstruction Department, International Labour Office.

Mincer, J. (1974). *Schooling, experience, and earning*: New York, National Bureau of Economic Research (1974).

Mincer, J. (1975). Education, Experience, and the Distribution of Earnings and Employment: An Overview *Education, Income, and Human Behavior* (pp. 71-94): NBER.

Morone, P. (2013). *Knowledge, Innovation and Internationalisation: Essays in Honour of Cesare Imbriani*: Routledge.

Moser, M., & Schnetzer, M. (2014). *The geography of average income and inequality : spatial evidence from Austria*. Vienna: Wien : Wirtschaftsuniv., Dep. of Economics.

Articles

Anselin, L., & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. *Statistics Textbooks and Monographs, 155*, 237-290.

Bazzi, S., Gaduh, A., Rothenberg, A. D., & Wong, M. (2016). Skill transferability, migration, and development: Evidence from population resettlement in indonesia. *The American Economic Review, 106*(9), 2658-2698.

Becker, G. S. (1962). Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy, 70*(5), 9-49.

Brixi, H. P. (2012). *Leading with ideas : skills for growth and equity in Thailand*. Washington, DC: World Bank Retrieved from http://documents.worldbank.org/curated/en/2012/01/15940701/leading-ideas-skills-growth-equity-thailand.

Brunow, S., & Hirte, G. (2009). The age pattern of human capital and regional productivity: A spatial econometric study on German regions. *Papers in Regional Science, 88*(4), 799-823.

Case, A. C. (1991). Spatial Patterns in Household Demand. *Econometrica, 59*(4), 953-965. doi:10.2307/2938168

De la Fuente, A., & Jimeno, J. F. (2005). The private and fiscal returns to schooling and the effect of public policies on private incentives to invest in education: A general framework and some results for the EU.

De Vreyer, P., & Spielvogel, G. (2005). *Spatial externalities between Brazilian municipios and their neighbours*. Retrieved from

Fersterer, J., & Winter-Ebmer, R. (2003). Are Austrian returns to education falling over time? *Labour Economics, 10*(1), 73-89. doi:https://doi.org/10.1016/S0927-5371(02)00105-7

Gebremariam, G. H., Gebremedhin, T. G., & Schaeffer, P. V. (2011). Employment, income, and migration in appalacia: a spatialL simultaneous equation approach*. *Journal of Regional Science, 51*(1), 102-120. doi:10.1111/j.1467-9787.2010.00681.x

Gibson, J., Kim, B., & Olivia, S. (2011). Spatial correlation in household choices in rural indonesia*. *Asian Economic Journal, 25*(3), 271-289. doi:10.1111/j.1467-8381.2011.02063.x

Glocker, D., & Steiner, V. (2011). *Returns to education across Europe: A comparative analysis for selected EU countries*. Retrieved from http://EconPapers.repec.org/RePEc:zbw:fubsbe:201115

Golledge, R. G. (1996). Geographical theories. *International Social Science Journal, 48*(150), 461-476. doi:10.1111/1468-2451.00049

Gust-Bardon, N. I. (2012). *The role of geographical proximity in innovation: Do regional and local levels really matter?* Retrieved from

Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review, 73*, 103-130. doi:http://dx.doi.org/10.1016/j.euroecorev.2014.10.006

Harmon, C., Oosterbeek, H., & Walker, I. (2002). The returns to education : A review of evidence, and deficiencies in the literature. *LSE in its series CEE Discussion Papers*(no 5).

Hawley, J. D. (2003). Comparing the payoff to vocational and academic credentials in Thailand over time. *International Journal of Educational Development, 23*(6), 607-625. doi:http://dx.doi.org/10.1016/S0738-0593(03)00061-0

Hawley, J. D. (2004). Changing returns to education in times of prosperity and crisis, Thailand 1985–1998. *Economics of Education Review, 23*(3), 273-286. doi:http://dx.doi.org/10.1016/j.econedurev.2003.11.010

Hering, L., & Poncet, S. (2010). Income Per Capita Inequality in China: The Role of Economic Geography and Spatial Interactions. *World Economy*, *33*(5), 655-679. doi:10.1111/j.1467-9701.2010.01241.x

Jiafeng, G. (2014). Spatial dynamics, vocational education and Chinese economic growth. *Journal of Systems Science and Information, 2*(5), 385-400.

Kemeny, T., & Storper, M. (2012). The Sources of Urban Development: Wages, Housing, and Amenity Gaps across American Cities *. *Journal of Regional Science, 52*(1), 85-108. doi:10.1111/j.1467-9787.2011.00754.x

Krongkaew, M., & Kakwani, N. (2003). The growth–equity trade-off in modern economic development: the case of Thailand. *Journal of Asian Economics, 14*(5), 735-757. doi:http://dx.doi.org/10.1016/j.asieco.2003.10.003

Lathapipat, D., & Chucherd, T. (2013). *Labor market functioning and Thailand's competitiveness*.

Lee, L. f. (2003). Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances. *Econometric Reviews*, *22*(4), 307-335. doi:10.1081/ETC-120025891

Maskell, P. (2001). Towards a Knowledge-based Theory of the Geographical Cluster. *Industrial and Corporate Change, 10*(4), 921-943. doi:10.1093/icc/10.4.921

Mehta, A., Felipe, J., Quising, P., & Camingue, S. (2013). Where have all the educated workers gone? Services and Wage Inequality in Three Asian Economies. *Metroeconomica*, *64*(3), 466-497.

Michaelowa, K. (2000). *Returns to education in low income countries: Evidence for Africa.* Paper presented at the annual meeting of the Committee on Developing Countries of the German Economic Association.

Miller, H. J., & Wentz, E. A. (2003). Representation and Spatial Analysis in Geographic Information Systems. *Annals of the Association of American Geographers, 93*(3), 574-594.

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, *66*(4), 281-302.

Montenegro, C. E., Montenegro, C. E., & Patrinos, H. A. (2014). Comparable Estimates of Returns to Schooling around the World. *Policy Research Working Paper Series, no. WPS 7020*. doi:10.1596/1813-9450-7020

Moran, P. A. P. (1948). The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society. Series B (Methodological), 10*(2), 243-251.

Paweenawat, S. W., & Vechbanyongratana, J. (2015). Wage Consequences of Rapid Tertiary Education Expansion in a Developing Economy: The Case of Thailand. *The Developing Economies, 53*(3), 218-231. doi:10.1111/deve.12078

Psacharopoulos, G. (2006). The Value of Investment in Education: Theory, Evidence, and Policy. *Journal of Education Finance, 32*(2), 113-136.

Psacharopoulos, G., & Patrinos, H. A. (2004). Returns to investment in education: a further update. *Education Economics*, *12*(2), 111-134. doi:10.1080/0964529042000239140

Reardon, S. F., & Bischoff, K. (2011). Income inequality and income segregation. *The American Journal of Sociology*(4), 1092.

Rigotti, J. I. R. (2006). Geography of population flow according to the Migrants' Educational level. *Estudos avançados*, *20*(57), 237-254.

Rodríguez-Pose, A., & Vilalta-Bufí, M. (2005). Education, migration, and job satisfaction: the regional returns of human capital in the EU. *Journal of Economic Geography, 5*(5), 545-566. doi:10.1093/jeg/lbh067

Rodríguez-Pose, A., & Tselios, V. (2009). Education and income inequality in he regions of the European union. *Journal of Regional Science, 49*(3), 411-437. doi:10.1111/j.1467-9787.2008.00602.x

Rodríguez-Pose, A., & Tselios, V. (2010). Returns to migration, education and externalities in the European Union. *Papers in Regional Science, 89*(2), 411-434. doi:10.1111/j.1435-5957.2010.00297.x

Senkrua, A. (2015). *The Mismatch in Thai labor market: Overeducation.* National Institute of Development Administration.

Silveira-Neto, R., & Azzoni, C. R. (2006). Location and regional income disparity dynamics: The Brazilian case. *Papers in Regional Science, 85*(4), 599-613.

Srinang, J. (2014). Vocational and general secondary education : the rate of return across regions in Thailand. *SOUTHEAST ASIAN JOURNAL OF ECONOMICS, 2*(2, (12)), 103-115.

Tangtipongkul, K. (2015). Rates of Return to Schooling in Thailand. *Asian Development Review, 32*(2), 38.

Up, L., & Donghyun, K. (2015). Toward Sustainable Economic Growth: A Spatial Panel Data Analysis of Regional Income Convergence in US BEA Economic Areas. *Sustainability (2071-1050), 7*(8), 9943.

Wannakrairoj, W. (2013). The effect of education and experience on wages : the case study of Thailand in 2012. *SOUTHEAST ASIAN JOURNAL OF ECONOMICS, 1*(1, (12)), 27-48.

Warunsiri, S., & McNown, R. (2010). The returns to education in Thailand: A pseudo-panel approach. *World Development, 38*(11), 1616-1625.

Watson, T. (2009). Inequality and the measurement of residential segregation by income in American neighborhoods. *Review of Income and Wealth*, *55*(3), 820-844. doi:10.1111/j.1475-4991.2009.00346.x

Yano, K. (2000). GIS and quantitative geography. *GeoJournal, 52*(3), 173-180.

Yildirim, J., Öcal, N., & Özyildirim, S. (2009). Income Inequality and Economic Convergence in Turkey: A Spatial Effect Analysis. *International Regional Science Review, 32*(2), 221.

Other Materials

Becker, G. S. (2008). Human Capital. *The Concise Encyclopedia of Economics.* Retrieved from http://www.econlib.org/library/Enc/HumanCapital.html

Economic Research Team, R. D. B. o. A. (2008). Economic Review 2008 and Outlook 2009 Retrieved from https://www.krungsri.com/bank/getmedia/61c34199-55a8-49ea-a189-5d892ef79b35/Research-econ-31012009.aspx

Group, T. A. M. (Producer). (2017). Mapping the Achievement Gap. *Mapping the Achievement Gap*. Retrieved from https://www.citylab.com/housing/2017/03/a-us-dot-map-of-educational-achievement-segregation/519063/

Rabinovitz, J. (2016). Local education inequities across U.S. revealed in new Stanford data set. from Stanford University https://ed.stanford.edu/news/local-education-inequities-across-us-revealed-new-stanford-data-set

APPENDICES

APPENDIX A
Data cleaning and generation

In general, the original Labor Force Survey is unavailable for direct use because noisy data in data mining may disturb statistical estimation as statistical noise. For example, noisy data are meaningless or corrupt data caused by the lack of record, false data, and qualitative data codes. Therefore, data must be managed to reduce statistical noise as much as possible. In addition, the original variables must be modified to create the most appropriate variables for the model at each level.

In this section, data cleaning and generation are described and illustrated.

The first part presents the statistical sample weight of the survey. In general, survey data collect information from a sample group to infer the properties of a population. Therefore, the sample weight must be used to weight the sample back and infer population size. From 2001 to 2011, sample weight has been recorded without any a decimal. In reality, however, sample weight must have four digits of a decimal number to represent the actual sample weight. Therefore, the sample weight obtained during these years must be divided by the ten-thousand and replaced as the actual weight. However, the sample weight for the other years is already reported as a four-digit decimal value. An example of the STATA code for generating the true sample weight in 2011 can be represented as follows:

```
. replace weight=weight/10000
weight was long now double
(216757 real changes made)
```

Other issues are the maximum value and conditional code of financial data, such as the monthly income called “approx.” (approximate monthly income). From 2001 to 2007, the observed data 0–99,997 represent the monthly income of a laborer (in baht). Recorded data 99,998 represent laborer’s income of 99,998 baht and higher. In addition, the code 99,999 is recorded to represent unknown and blank data. Nevertheless, monthly income has been recorded as 0–999,997 baht since 2008 and 999,998 is recorded to represent laborer’s income of 99,998 baht and higher. The code

999,999 is unknown and blank. Thus, the code 99,999 or 999,999 must be deleted because it is a qualitative code in a qualitative variable. For example, the first case is from 2001 to 2007 and the second case is from 2008 and beyond.

In the year 2001

```
. drop if approx==99999
(127 observations deleted)
```

In the year 2008

```
. drop if approx==999999
(223 observations deleted)
```

This process does not delete unrecorded or blank data because the provincial level model requires the average labor income, schooling years, and years of working experience of the entire population. Then, these data will be eliminated automatically by using STATA software to compute the mean value process.

At the individual level, the dependent variable in the Mincer type model is the natural logarithm of the monthly income. Thus, labor income must be transformed by using the natural logarithm. In this case, blank data remain the missing value. An example for 2011 can be shown as follows:

```
. gen lnwage=log(approx)
(175054 missing values generated)
```

The most important variables in the Mincer function are the human capital variables, which are the years of schooling and work experience. However, these two variables were not asked and recorded during the survey. Therefore, the potential years of these variables must be generated.

In the database, labor education is recorded as the statistical code for the details of the educational program at each educational level that can be disaggregated into ten categories as a qualitative dummy variable. Since 2007, the code has been recorded by a three-digit number. However, it had been recorded by a two-digit code in previous years. The ten levels of education must be generated under the same definition but different educational codes, where grade=0 is no education, grade=1 is early childhood education, grade=2 is elementary education, grade=3 is lower secondary education, grade=4 is upper secondary education, grade=5 is post-secondary education, vocational and diploma education, grade=6 is bachelor degree, grade=7 is a master degree, grade=8 is a doctoral degree, and grade=9 is others education. An unknown degree is represented by blank data. An example can be presented as follows:

In the years 2001–2006	In the years 2007–2015
. gen grade =0 if grade_b==00	. gen grade =0 if grade_b==000
. replace grade=1 if grade_b==1	. replace grade=1 if grade_b==110
. replace grade=1 if grade_b==2	. replace grade=1 if grade_b>=211 |
. replace grade=1 if grade_b==3	grade_b<=215
. replace grade=1 if grade_b==4	. replace grade=1 if grade_b>=241 |
. replace grade=1 if grade_b==5	grade_b<=242
. replace grade=1 if grade_b==6	. replace grade=1 if grade_b>=251 |
. replace grade=1 if grade_b==7	grade_b<=255
. replace grade=1 if grade_b==8	. replace grade=2 if grade_b==210
. replace grade=1 if grade_b==9	. replace grade=2 if grade_b==240
. replace grade=1 if grade_b==84	. replace grade=2 if grade_b==250
. replace grade=1 if grade_b==85	. replace grade=3 if grade_b==310
. replace grade=1 if grade_b==86	. replace grade=3 if grade_b==320
. replace grade=2 if grade_b==10	. replace grade=3 if grade_b==330
. replace grade=2 if grade_b==12	. replace grade=3 if grade_b==340
. replace grade=2 if grade_b==13	. replace grade=3 if grade_b==350
. replace grade=2 if grade_b==14	. replace grade=4 if grade_b==410
. replace grade=2 if grade_b==35	. replace grade=4 if grade_b==420
. replace grade=2 if grade_b==37	. replace grade=4 if grade_b==430
. replace grade=2 if grade_b==38	. replace grade=4 if grade_b==440
. replace grade=2 if grade_b==71	. replace grade=4 if grade_b==450
. replace grade=2 if grade_b==72	. replace grade=4 if grade_b==460
. replace grade=2 if grade_b==87	. replace grade=5 if grade_b==510
. replace grade=3 if grade_b==15	. replace grade=5 if grade_b==520
. replace grade=3 if grade_b==17	. replace grade=6 if grade_b==610
. replace grade=3 if grade_b==18	. replace grade=6 if grade_b==630
. replace grade=3 if grade_b==36	. replace grade=6 if grade_b==640
. replace grade=3 if grade_b==39	. replace grade=6 if grade_b==650
. replace grade=3 if grade_b==40	. replace grade=6 if grade_b==660
. replace grade=3 if grade_b==41	. replace grade=7 if grade_b==710
. replace grade=3 if grade_b==73	. replace grade=7 if grade_b==730

```
. replace grade=3 if grade_b==74
. replace grade=3 if grade_b==75
. replace grade=3 if grade_b==88
. replace grade=3 if grade_b==89
. replace grade=4 if grade_b==19
. replace grade=4 if grade_b==21
. replace grade=4 if grade_b==23
. replace grade=4 if grade_b==24
. replace grade=4 if grade_b==25
. replace grade=4 if grade_b==26
. replace grade=4 if grade_b==27
. replace grade=4 if grade_b==46
. replace grade=4 if grade_b==48
. replace grade=4 if grade_b==51
. replace grade=4 if grade_b==52
. replace grade=4 if grade_b==53
. replace grade=4 if grade_b==54
. replace grade=4 if grade_b==60
. replace grade=4 if grade_b==62
. replace grade=4 if grade_b==66
. replace grade=4 if grade_b==67
. replace grade=4 if grade_b==76
. replace grade=4 if grade_b==78
. replace grade=4 if grade_b==79
. replace grade=4 if grade_b==80
. replace grade=4 if grade_b==90
. replace grade=4 if grade_b==91
. replace grade=4 if grade_b==92
. replace grade=4 if grade_b==93
. replace grade=4 if grade_b==11
. replace grade=4 if grade_b==16
. replace grade=4 if grade_b==20
. replace grade=4 if grade_b==29
. replace grade=4 if grade_b==42
. replace grade=4 if grade_b==43
. replace grade=4 if grade_b==44
. replace grade=4 if grade_b==45
. replace grade=4 if grade_b==47
. replace grade=4 if grade_b==49
. replace grade=4 if grade_b==63
. replace grade=4 if grade_b==64
. replace grade=4 if grade_b==65
. replace grade=4 if grade_b==58
. replace grade=7 if grade_b==750
. replace grade=7 if grade_b==760
. replace grade=8 if grade_b==810
. replace grade=8 if grade_b==830
. replace grade=8 if grade_b==850
. replace grade=8 if grade_b==860
. replace grade=8 if grade_b==870
. replace grade=9 if grade_b==911
. replace grade=9 if grade_b==912
. replace grade=9 if grade_b==919
. replace grade=9 if grade_b==921
. replace grade=9 if grade_b==929
```

```
. replace grade=5 if grade_b==22
. replace grade=5 if grade_b==50
. replace grade=5 if grade_b==55
. replace grade=5 if grade_b==56
. replace grade=5 if grade_b==82
. replace grade=5 if grade_b==59
. replace grade=5 if grade_b==61
. replace grade=5 if grade_b==77
. replace grade=6 if grade_b==30
. replace grade=6 if grade_b==81
. replace grade=6 if grade_b==57
. replace grade=6 if grade_b==81
. replace grade=6 if grade_b==68
. replace grade=6 if grade_b==94
. replace grade=7 if grade_b==31
. replace grade=7 if grade_b==32
. replace grade=7 if grade_b==69
. replace grade=7 if grade_b==83
. replace grade=8 if grade_b==33
. replace grade=8 if grade_b==70
. replace grade=9 if grade_b==95
. replace grade=9 if grade_b==97
```

As mentioned in Chapter 3, schooling years is developed based on the definition of the Thai educational system by the Ministry of Education. Labors who finished lower elementary school are classified under no education because formal cognitive skill is emphasized after pre-elementary school and compulsory education begins at primary school. Primary school is a six-year course. For lower and upper secondary education, students complete a three-year course. For post-secondary and diploma education, students must spend three years to complete. Most undergraduate courses or tertiary education require four years of studying. Master's and doctoral degrees are two-year and five-year courses, respectively. In addition, students who study other forms of education, such as religious schools and Pondok school, spend a total of nine years in school as required by law.

Therefore, the accumulated years of schooling are 6, 9, 12, 15, 16, 18, 22, and 9 years for primary, lower secondary, upper secondary, diploma, undergraduate, master's, doctoral, and other forms of education, respectively.

The STAT command can be illustrated as follows:

```
. gen y_school=0 if grade==0
. replace y_school=0 if grade==1
. replace y_school=6 if grade==2
. replace y_school=9 if grade==3
. replace y_school=12 if grade==4
. replace y_school=15 if grade==5
. replace y_school=16 if grade==6
. replace y_school=18 if grade==7
. replace y_school=22 if grade==8
. replace y_school=9 if grade==9
```

In addition, the potential years of working experience must be generated by the formula as the reported age minus schooling years minus six, which is the age for starting elementary class. The STATA command does not differ across the years and can be represented as follows:

```
. gen exp=age-6 if grade==0
. replace exp=age-6 if grade==1
. replace exp=age-6-6 if grade==2
. replace exp=age-6-9 if grade==3
. replace exp=age-6-12 if grade==4
. replace exp=age-6-15 if grade==5
. replace exp=age-6-16 if grade==6
. replace exp=age-6-18 if grade==7
. replace exp=age-6-22 if grade==8
. replace exp=age-6-9 if grade==9
. replace exp=0 if exp<0
```

However, the models require a quadratic term for working experience to explain a diminishing rate of return on that effect. The squared work experience of each year can be generated by the following command:

```
. gen expsq=exp^2
```

Gender, marital status, working status, and living area are important variables to determine income difference based on personal characteristics. These variables are generated as qualitative dummy variables. The gender variable is equal to 1 for male and 0 otherwise. The living area variable is equal to 1 for the municipal area and 0 otherwise. Marital status is divided into three groups, namely, married, divorced, and single, which is the base case. A married worker is equal to 1 and 0 otherwise. A divorced worker is equal to 1 and 0 otherwise. Working status is classified into three groups of employment types, namely, public sector, state enterprise, and private sector, which is the base case.

An example of generating the marital status variable is as follows:

```
. gen marit =1 if marital==2
. replace marit =0 if marital==1
. replace marit =0 if marital==3
. replace marit =0 if marital==4
. replace marit =0 if marital==5
. replace marit =0 if marital==6
. replace marit =0 if marital==.
. gen divorced =1 if marital==3
. replace divorced =1 if marital==4
. replace divorced =1 if marital==5
. replace divorced =1 if marital==6
. replace divorced =0 if marital==1
. replace divorced =0 if marital==2
. replace divorced =0 if marital==.
```

An example of generating the gender variable is as follows:

```
. gen male =1 if sex==1
. replace male=0 if sex==2
```

An example of generating the living area variable is as follows:

```
. gen municipal =1 if area==1
. replace municipal =0 if area==2
```

An example of generating the working status variable is as follows:

```
. gen public=1 if status==4
. replace public=0 if status==1
. replace public=0 if status==2
. replace public=0 if status==3
. replace public=0 if status==5
. replace public=0 if status==6
. replace public=0 if status==7
. replace public=0 if status==8
. replace public=0 if status==.
.
. gen state_en=1 if status==5
. replace state_en=0 if status==1
. replace state_en=0 if status==2
. replace state_en=0 if status==3
. replace state_en=0 if status==4
. replace state_en=0 if status==6
. replace state_en=0 if status==7
. replace state_en=0 if status==8
. replace state_en=0 if status==.
```

Occupation and industry variables must be recorded as occupation codes. The original occupation code is a four-digit code according to the standard of the International Standard Classification of Occupations (ISCO). However, such code is inappropriate for the estimation process. Therefore, nine occupation types are classified by the STATA command as follows:

```
. gen occupation=1 if occup>=1100
. replace occupation=2 if occup>=2100
. replace occupation=3 if occup>=3100
. replace occupation=4 if occup>=4100
. replace occupation=5 if occup>=5100
. replace occupation=6 if occup>=6100
. replace occupation=7 if occup>=7100
. replace occupation=8 if occup>=8100
. replace occupation=0 if occup>=9100
. replace occupation=. if occup==.
```

where 0–8 are the new codes that represent a dummy variable of occupation, with 0 as basic worker. Then, 1–8 represent legislators, professionals, technicians, clerks, service workers, skilled agricultural workers, craft workers, and machine operators, respectively. In this case, blank data mean unknown or unidentified data.

The classification of nine types of occupation is based on the definition of the National Statistical Office of Thailand and ISCO. The element of each category can be represented as shown in Table Appendix A.1.

Table Appendix A.1

Nine-type of occupation classification

Occupation (ISCO)	Description
Managers, Senior officials, and Legislators	- Senior manager, Senior officials, and Legislators - Manager of production and services and Hotel managers
Professionals	- Professionals in Science and Engineering, Health, Teaching, business, and business administration and Professional in legal, social and cultural.
Technicians and associated professionals	- Professionals involved in the physical sciences and engineering, the field of health, the business and management, the legal, social, cultural and other aspects involved
Clerk	- The customer service clerk - Clerk recording the merchandising.
Service workers, and Shop sales worker	- Personal Assistant - Dealers - Personal service worker - Housekeeping and Related Service Supervisors

Table Appendix A.1 (continued)

Occupation (ISCO)	Description
Skilled agricultural, and Fishery worker	- Skilled workers in agriculture, forestry, fishery, and hunting for trading. - Skilled workers in agriculture, fishing, hunting and collecting fruits for a living
Craft, and related worker	- Construction and related workers. - Electrician and Electronic technician - Metal crafts and practitioners involved. - Artisan, Printers, and related workers - Food, wood, and costumes processors.
Machine operator, and Assembler worker	- Stationary Plant operators and Related Equipment Operators - Drivers, and mobile machinery controller, and Transport Equipment Operators
Basic worker	- Common Laborers, Unskilled Workers - House workers and Cleaners - Workers in agriculture, fisheries and forestry. - Workers in mining, construction, manufacturing and transportation. - Cook assistant - waste operator

Source: Author's compilation from Labor Force Survey based on ISCO

The industry variable is recorded by the code proposed by the International Standard Industrial Classification of All Economic Activities. However, the industry code can be disaggregated into eight groups as mentioned earlier. Numbers 1–8 of the variables are recorded as codes for qualitative dummy variables. where number 1 represents workers in the agricultural, fishing, and hunting sectors, 2 represents workers in the mining sector, 3 represents workers in the utility sector, 4 and 5 respectively represent low- and high-skilled manufacturing sector, 6 represents workers in the construction sector, 7 and 8 respectively represents low- and high-skilled service sector.

Nevertheless, the code recorded for each year differs. The code had been recorded as 3- to 4-digit numbers in 2001–2010 before several details regarding sub-industries changed in 2011. Since 2012, the code has been recorded as 4- to 5-digit numbers according to ISIC. Therefore, the STATA command for each year can be presented as follows:

In the years 2001-2010	In the years 2011	In the years 2012-2015
`. gen industry=1 if indus>=100`	`. gen industry=1 if indus>=100`	`. gen industry=1 if indus>=1000`
`. replace industry=2 if indus>=1000`	`. replace industry=2 if indus>=500`	`. replace industry=2 if indus>=5000`
`. replace industry=5 if indus>=1100`	`. replace industry=5 if indus>=600`	`. replace industry=5 if indus>=6000`
`. replace industry=2 if indus>=1200`	`. replace industry=2 if indus>=700`	`. replace industry=2 if indus>=7000`
`. replace industry=4 if indus>=1500`	`. replace industry=4 if indus>=1000`	`. replace industry=4 if indus>=10000`
`. replace industry=5 if indus>=2100`	`. replace industry=5 if indus>=1700`	`. replace industry=5 if indus>=17000`
`. replace industry=4 if indus>=2500`	`. replace industry=4 if indus>=2200`	`. replace industry=4 if indus>=22000`
`. replace industry=5 if indus>=2700`	`. replace industry=5 if indus>=2400`	`. replace industry=5 if indus>=24000`
`. replace industry=4 if indus>=2800`	`. replace industry=4 if indus>=2500`	`. replace industry=4 if indus>=25000`
`. replace industry=5 if indus>=2900`	`. replace industry=5 if indus>=2600`	`. replace industry=5 if indus>=26000`
`. replace industry=4 if indus>=3400`	`. replace industry=4 if indus>=2900`	`. replace industry=4 if indus>=29000`
`. replace industry=3 if indus>=4000`	`. replace industry=5 if indus>=3300`	`. replace industry=5 if indus>=33000`
`. replace industry=6 if indus>=4500`	`. replace industry=3 if indus>=3500`	`. replace industry=3 if indus>=35000`
`. replace industry=7 if indus>=5000`	`. replace industry=6 if indus>=4100`	`. replace industry=6 if indus>=41000`
`. replace industry=8 if indus>=6500`	`. replace industry=7 if indus>=4500`	`. replace industry=7 if indus>=45000`

. replace industry=7 if indus>=9000	. replace industry=5 if indus>=5800	. replace industry=5 if indus>=58000
. replace industry=8 if indus>=9900	. replace industry=8 if indus>=5900	. replace industry=8 if indus>=59000
. replace industry=. if indus==.	. replace industry=7 if indus>=8700	. replace industry=7 if indus>=87000
	. replace industry=8 if indus>=9900	. replace industry=8 if indus>=99000
	. replace industry=. if indus==.	. replace industry=. if indus==.

The classification of industries into eight types is recorded according to the study of Tangtipongkul (2015). In addition, the element of each category can be described as shown in Table Appendix A.2.

Table Appendix A.2

Eight-type of industry classification

Sector	Description
Agriculture	- Agriculture - Animal husbandry, Fisheries and Hunting - Forestry
Mining & Quarrying	- Mining - Quarrying.
Utilities	- Electricity, Gas, and Water supply. - Wastewater Management
Construction	- Construction and Civil engineering
Low-skill manufacturing	- Food products - Beverage production - Tobacco Manufacturing - Textile and Apparel production - Leather production and related products - products of wood and cork. (Except furniture) - Manufacture of plaiting materials.

Table Appendix A.2 (continued)

Sector	Description
Low-skill manufacturing (continued)	- Production of rubber - Production of other products made of metals. - Fabricated metal products (Except machinery) - Production of furniture
High-skill manufacturing	- Production of basic metals - Manufacture of paper and paper products - Printing and reproduction of recorded media. - Crude petroleum and natural gas drilling - Manufacture of coke, refined petroleum product and nuclear fuel - Chemical products - Production of pharmaceuticals, Chemicals - Electrical Equipment Manufacturing - Repair and installation of machinery
Low-skill services	- Repairing of motor vehicles and motorcycles - Wholesale trade and Retail trade - Transportation - Postal and parcel delivery activities. - Personal and Household service activities - Social work activities - Creative arts and entertainment activities. - Library, Archives, Museum And other cultural activities - Gambling Activities - Hotels and restaurants - Sanitary and similar activities
High-skill services	- The programming - Telecommunications - Information service

Table Appendix A.2 (continued)

Sector	Description
High-skill services (continued)	- Financial service - Insurance, insurance and pension funds - Real estate activity - Legal and accounting activities - Education, Scientific and Research - Human health and medical service - Government administration, and Defense - Social work and compulsory social security - Service management consulting organization. - Business activities, including renting

Source: Author's compilation from LFS based on Tangtipongkul (2015)

Other variables, such as working hours, do not differ across the years and do not need to be transformed. In addition, the provincial ID called cwt and the year variables are used as dummy variables to capture the cross-sectional effect and time effect, respectively.

To develop the Mincer model at the provincial level, the quantitative variables of this model must explain the overall population characteristics in the entire province as an average value of such variables. Qualitative variables, such as gender and marital status, must be converted into their proportion to the entire population. In general, survey data including the Labor Force Survey collects data from a sample group that represents the whole.

The sample must be a representative of the population in the province. The Labor Force Survey uses a stratified two-stage sampling technique under a province stratum, which can be divided into municipal and non-municipal areas in the first stage sample unit before sample households from all the households are selected via systematic sampling. Therefore, each sample requires an individual sample weight or probability weight to weight the sample back and infer the population. In general, the weight is the inverse of the probability to be selected or the original selection probability of a respondent from a population based on sampling design. In the formula, sample weight can be calculated with N/n, where N is the number of the population and n is the number of the sample. Therefore, a sample weight should be higher than one.

Before calculating the average values and proportions of quantitative and qualitative variables, the sample must be weight back to infer the population because average values and proportions with and without weight are different.

In STATA, the data must be set according to the design of the survey, including sample weight. In general, the average value of each quantitative variable over the province must be computed by the summation of multiplying the sample weight with the observed value before dividing by the summation of the sample weight to represent the average value of the variable in the province. The proportion of a qualitative variable over the province is also calculated based on population size, such as the proportion of males compared with the entire population of the province. In this thesis, average values and proportions do not cover missing data, which are eliminated automatically in the process.

In STATA, the survey data are introduced as follows:

```
      .  svyset  cwt  [pweight=weight],  strata(area)  vce(linearized)
singleunit(missing) || id

      Note: Stage 1 is sampled with replacement; all further stages
will be ignored
            pweight: weight
                VCE: linearized
        Single unit: missing
           Strata 1: area
               SU 1: cwt
              FPC 1: <zero>
```

where svyset is the command for introducing the stratum and sample unit to STATA.

Individuals aged 15 years old and above, before and after management, can be represented by the Labor Force Survey of 2001–2015 by weighting back to represent the population of Thailand as shown in Table Appendix A.3.

In addition, after setting the survey data, an example of the STATA command for generating an average value of the quantitative variable and the command for generating a proportion of the qualitative variable can be represented as follows:

```
. svy linearized : mean Schooling_year , over(cwt)
. svy linearized : proportion Male , over(cwt)
```

where "Schooling_year" and "Male" are quantitative and qualitative variables, respectively. However, a qualitative variable must be generated as a binary variable 0 and 1, where 1 = yes and 0 = no instead of the average value.

Table Appendix A.3

Expected Thailand population aged 15 years old and above by weighting back from the sample in the survey data

Year	Before cleaning	After cleaning
2001	63,001,140	62,914,163
2002	63,526,908	63,422,307
2003	64,062,602	63,908,949
2004	65,197,160	65,031,164
2005	64,884,045	64,726,169
2006	65,199,806	65,199,806
2007	65,800,080	65,676,629
2008	66,511,667	66,439,864
2009	66,933,897	66,870,762
2010	67,333,139	67,248,208
2011	67,621,973	67,589,020
2012	67,932,487	67,885,603
2013	68,276,530	67,987,829
2014	67,012,058	66,741,447
2015	67,244,098	67,090,445

Source: Author's calculation from Labor Force Survey

In this thesis, income, earning, schooling years, years of work experience, and working hours are generated as average values of the labor force in the province. Proportion is generated to represent the gender, living area, and working status variables, along with eight industry sectors and nine occupation types. These variables in the provincial level model are generated from different population sizes because several respondents did not answer the questionnaire, and thus, the survey has missing data, which are eliminated automatically by the program. Consequently, each variable is calculated from a different population size as indicated in Table Appendix A.4.

Table appendix A.4

Expected number of population used to calculate the average value and proportion of each variable in each year

Year	**Income and Earning**	**Schooling years and Experience**	**Working hour**	**All of qualitative variables**
2001	13,416,009	49,497,135	33,396,748	62,914,163
2002	13,572,230	47,538,979	34,157,798	63,422,307
2003	13,904,881	48,117,980	34,522,737	63,908,949
2004	15,476,757	49,116,212	35,545,333	65,031,164
2005	15,689,562	49,663,091	36,144,484	64,726,169
2006	15.728,106	50,193,826	36,202,442	65,199,806
2007	16,052,187	65,676,629	36,999,242	65,676,629
2008	16,269,366	66,439,864	37.764,755	66,439,864
2009	16,387,377	66,879,762	38,308,391	66,870,762
2010	16,400,395	67,248,208	38,606,651	67,248,208
2011	16,543,203	67,589,020	38,284,283	67,589,020
2012	16,398,312	67,885,603	39,538,833	67,885,603
2013	15,907,236	67,987,829	38,823,699	67,987,829
2014	17,255,021	66,731,447	38,150,510	66,741,447
2015	17,669,429	67,090,445	38,176,766	67,090,445

Source: Author’s calculation from Labor Force Survey

The meaning of each variable can be described as the average value and proportion of that of the labors in the province. In this case, the observation in each year is equal to 76 or the number of the province in Thailand. The average income is the average monthly income of the laborers in the province. The average schooling years, years of work experience, and main working hours are the mean values of those in the province. All the qualitative variables are generated in proportion form. Therefore, they must be described as the proportion of qualified workers to the total number of workers in the entire province. For example, the variable male is the proportion of men in the province.

APPENDIX B
Variable and summary statistic tables

Table Appendix B.1

Variables description in the individual model

variable	Description
Lninc	Natural logarithm of monthly income
grade	Level of education Grade=0 is no education Grade=1 is finished pre-primary education Grade=2 is finished primary education Grade=3 is finished lower secondary education Grade=4 is finished upper secondary education Grade=5 is finished post-secondary education, vocational or diploma Grade=6 is finished bachelor degree Grade=7 is finished master degree Grade=8 is finished doctoral degree Grade=9 is finished others education
sch	Years of schooling
exp	Years of working experience
expsq	Years of working experience squared
hr	The main working hour
male	Gender, if male=2 and otherwise=0
munc	Living in municipal area=1 and otherwise=0
marit	Marital status=1 and otherwise=0
divorced	Divorced, Widowed, or Separated status = 1 and otherwise=0
pub	Working in the public sector=1 and otherwise=0
sten	Work in the state-enterprise sector=1 and otherwise=0
occ	Occupation of workers Occupation=0 is a basic worker

Table Appendix B.1 (continued)

variable	Description
occ (continued)	Occupation=1 is legislators Occupation=2 is professionals Occupation=3 is technician Occupation=4 is clerk Occupation=5 is service worker Occupation=6 is skill agricultural Occupation=7 is craft Occupation=8 is machine operator
ind	Type of industry Industry= 1 is agricultural sector (and fishery sector) Industry=2 is mining sector Industry=3 is utility sector Industry=4 is low-skill manufacturing Industry=5 is high-skill manufacturing Industry=6 is construction Industry=7 is low -skill service Industry=8 is high-skill service
cwt	Provincial code Cwt=10 is Bangkok Metropolis Cwt=11 is Samut Prakan Cwt=12 is Nonthaburi Cwt=13 is Pathum Thani Cwt=14 is Phra Nakhon Si Ayutthaya Cwt=15 is Ang Thong Cwt=16 is Lop Buri Cwt=17 is Sing Buri Cwt=18 is Chai Nat Cwt=19 is Saraburi Cwt=20 is Chon Buri

Table Appendix B.1 (continued)

variable	Description
cwt	Cwt=21 is Rayong
(continued)	Cwt=22 is Chanthaburi
	Cwt=23 is Trat
	Cwt=24 is Chachoengsao
	Cwt=25 is Prachin Buri
	Cwt=26 is Nakhon Nayok
	Cwt=27 is Sa Kaeo
	Cwt=30 is Nakhon Ratchasima
	Cwt=31 is Buri Ram
	Cwt=32 is Surin
	Cwt=33 is Si Sa Ket
	Cwt=34 is Ubon Ratchathani
	Cwt=35 is Yasothon
	Cwt=36 is Chaiyaphum
	Cwt=37 is Am Nat Charoen
	Cwt=39 is Nong Bua Lam Phu
	Cwt=40 is Khon Kaen
	Cwt=41 is Udon Thani
	Cwt=42 is Loei
	Cwt=43 is Nong Khai
	Cwt=44 is Maha Sarakham
	Cwt=45 is Roi Et
	Cwt=46 is Kalasin
	Cwt=47 is Sakon Nakhon
	Cwt=48 is Nakhon Phanom
	Cwt=49 is Mukdahan
	Cwt=50 is Chiang Mai
	Cwt=51 is Lamphun
	Cwt=52 is Lampang

Table Appendix B.1 (continued)

variable	Description
cwt (continued)	Cwt=53 is Uttaradit
	Cwt=54 is Phrae
	Cwt=55 is Nan
	Cwt=56 is Phayao
	Cwt=57 is Chiang Rai
	Cwt=58 is Mae Hong Son
	Cwt=60 is Nakhon Sawan
	Cwt=61 is Uthai Thani
	Cwt=62 is Kamphaeng Phet
	Cwt=63 is Tak
	Cwt=64 is Sukhothai
	Cwt=65 is Phitsanulok
	Cwt=66 is Phichit
	Cwt=67 is Phetchabun
	Cwt=70 is Ratchaburi
	Cwt=71 is Kanchanaburi
	Cwt=72 is Suphan Buri
	Cwt=73 is Nakhon Pathom
	Cwt=74 is Samut Sakhon
	Cwt=75 is Samut Songkhram
	Cwt=76 is Phetchaburi
	Cwt=77 is Prachuap Khiri Khan
	Cwt=80 is Nakhon Si Thammarat
	Cwt=81 is Krabi
	Cwt=82 is Phangnga
	Cwt=83 is Phuket
	Cwt=84 is Surat Thani
	Cwt=85 is Ranong
	Cwt=86 is Chumphon

Table Appendix B.1 (continued)

variable	Description
cwt (continued)	Cwt=90 is Songkhla
	Cwt=91 is Satun
	Cwt=92 is Trang
	Cwt=93 is Phatthalung
	Cwt=94 is Pattani
	Cwt=95 is Yala
	Cwt=96 is Narathiwat
yr	Year of survey
	Year=44 is 2001
	Year=45 is 2002
	Year=46 is 2003
	Year=47 is 2004
	Year=48 is 2005
	Year=49 is 2006
	Year=50 is 2007
	Year=51 is 2008
	Year=52 is 2009
	Year=53 is 2010
	Year=54 is 2011
	Year=55 is 2012
	Year=56 is 2013
	Year=57 is 2014
	Year=58 is 2015

Source: Author's variable explanation

Table Appendix B.2

Variable description in the provincial model

Variable	Description
Lninc	Natural logarithm of the average monthly income of labors in the province
Sch	The average years of schooling of labors in the province
exp	The average years of working experience of worker in the province
expsq	The quadratic form of the average years of working experience of worker

Source: Author’s variable explanation

Table Appendix B.3

The summarize statistic of each variable in the individual level model

Variable	2001	2002	2003	2004	2005	2006	2007	2008
No. of observation	156,870	159,460	156,819	154,532	159,426	170,599	219,263	225,177
Population size	46,947,135	47,538,979	48,117,981	49,116,212	49,663,091	50,193,826	65,676,629	66,439,864
lninc mean (s.d.)	8.44 (0.12)	8.45 (0.11)	8.49 (0.11)	8.50 (0.10)	8.56 (0.10)	8.64 (0.09)	8.69 (0.09)	8.75 (0.09)
sch mean (s.d.)	5.42 (0.41)	5.52 (0.37)	5.71 (0.38)	5.91 (0.36)	5.92 (0.35)	6.07 (0.34)	4.86 (0.29)	4.99 (0.26)
exp mean (s.d.)	26.79 (0.61)	26.86 (0.56)	27.01 (0.57)	27.08 (0.54)	27.82 (0.49)	27.87 (0.48)	22.32 (0.23)	22.72 (0.20)
expsq mean (s.d.)	1,111.13 (39.44)	1,118.36 (36.97)	1,134.06 (37.84)	1,142.54 (36.06)	1,182.89 (34.15)	1,188.85 (33.85)	928.83 (19.10)	952.91 (17.29)
hr mean (s.d.)	46.77 (0.60)	46.84 (0.58)	46.58 (0.72)	46.68 (0.76)	45.52 (0.80)	45.32 (0.75)	45.31 (0.64)	45.65 (0.54)
male=1 proportion (s.d.)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.49 (0.00)	0.49 (0.00)	0.49 (0.00)	0.49 (0.00)
munc=1 proportion (s.d.)	0.32 (0.08)	0.33 (0.08)	0.33 (0.08)	0.33 (0.08)	0.30 (0.07)	0.30 (0.07)	0.30 (0.07)	0.31 (0.07)
marit=1 proportion (s.d.)	0.47 (0.00)	0.47 (0.00)	0.48 (0.00)	0.48 (0.00)	0.49 (0.00)	0.50 (0.01)	0.50 (0.01)	0.50 (0.00)
divorced=1 proportion (s.d.)	0.07 (0.00)	0.07 (0.00)	0.07 (0.00)	0.07 (0.00)	0.08 (0.00)	0.08 (0.00)	0.08 (0.00)	0.08 (0.00)
pub=1 proportion (s.d.)	0.04 (0.00)	0.04 (0.00)	0.04 (0.00)	0.04 (0.00)	0.04 (0.00)	0.04 (0.00)	0.04 (0.00)	0.05 (0.00)

Table Appendix B.3 (continued)

Variable	2001	2002	2003	2004	2005	2006	2007	2008
sten=1 proportion (s.d.)	0.01 (0.00)	0.01 (0.00)	0.00 (0.00)	0.01 (0.00)	0.01 (0.00)	0.00 (0.00)	0.01 (0.00)	0.01 (0.00)
occ=0 proportion (s.d.)	0.04 (0.00)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)	0.07 (0.00)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)
occ=1 proportion (s.d.)	0.03 (0.00)	0.04 (0.00)	0.00 (0.00)	0.04 (0.00)	0.04 (0.00)	0.04 (0.00)	0.04 (0.00)	0.01 (0.00)
occ=2 proportion (s.d.)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)
occ=3 proportion (s.d.)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)
occ=4 proportion (s.d.)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)
occ=5 proportion (s.d.)	0.07 (0.01)	0.07 (0.01)	0.07 (0.01)	0.08 (0.01)	0.08 (0.01)	0.08 (0.01)	0.08 (0.01)	0.09 (0.01)
occ=6 proportion (s.d.)	0.24 (0.03)	0.23 (0.03)	0.22 (0.03)	0.21 (0.03)	0.22 (0.02)	0.21 (0.02)	0.22 (0.02)	0.22 (0.02)
occ=7 proportion (s.d.)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)
occ=8 proportion (s.d.)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.05 (0.01)	0.05 (0.01)	0.04 (0.00)	0.05 (0.00)	0.04 (0.00)
ind=1 proportion (s.d.)	0.25 (0.03)	0.25 (0.03)	0.24 (0.03)	0.23 (0.03)	0.24 (0.03)	0.24 (0.03)	0.24 (0.03)	0.24 (0.03)

Table Appendix B.3 (continued)

Variable	2001	2002	2003	2004	2005	2006	2007	2008
ind=2 proportion (s.d.)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
ind=3 proportion (s.d.)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
ind=4 proportion (s.d.)	0.06 (0.01)	0.07 (0.01)	0.07 (0.01)	0.07 (0.01)	0.07 (0.01)	0.06 (0.01)	0.07 (0.01)	0.06 (0.01)
ind=5 proportion (s.d.)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)
ind=6 proportion (s.d.)	0.02 (0.00)	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)
ind=7 proportion (s.d.)	0.14 (0.02)	0.14 (0.02)	0.14 (0.02)	0.15 (0.02)	0.15 (0.01)	0.15 (0.02)	0.15 (0.02)	0.15 (0.01)
ind=8 proportion (s.d.)	0.05 (0.01)	0.05 (0.01)	0.05 (0.01)	0.05 (0.01)	0.06 (0.01)	0.06 (0.01)	0.06 (0.01)	0.06 (0.01)

Table Appendix B.3 (continued)

Variable	2009	2010	2011	2012	2013	2014	2015
No. of observation	220,525	147,571	216,636	239,118	233,801	232,145	228,994
Population size	66,870,762	67,248,208	67,589,020	67,885,603	67,987,829	66,741,447	67,090,445
lninc mean (s.d.)	8.76 (0.09)	8.83 (0.08)	8.92 (0.07)	9.04 (0.08)	9.15 (0.08)	9.26 (0.07)	9.29 (0.08)
sch mean (s.d.)	5.12 (0.23)	5.31 (0.23)	5.42 (0.24)	5.53 (0.25)	5.73 (0.26)	5.95 (0.40)	6.07 (0.41)
exp mean (s.d.)	22.99 (0.18)	23.21 (0.16)	23.44 (0.16)	23.61 (0.17)	23.79 (0.17)	24.79 (0.48)	25.04 (0.48)
expsq mean (s.d.)	970.38 (15.16)	984.59 (14.08)	997.87 (14.22)	1,006.23 (14.99)	1,011.30 (14.18)	1,081.12 (39.36)	1,100.12 (40.27)
hr mean (s.d.)	45.71 (0.58)	45.85 (0.61)	45.67 (0.61)	45.08 (0.57)	43.88 (0.59)	44.34 (0.71)	43.06 (0.63)
male=1 proportion (s.d.)	0.49 (0.00)	0.49 (0.00)	0.49 (0.00)	0.49 (0.00)	0.49 (0.00)	0.49 (0.00)	0.49 (0.00)
munc=1 proportion (s.d.)	0.32 (0.07)	0.34 (0.07)	0.34 (0.07)	0.34 (0.07)	0.34 (0.07)	0.44 (0.07)	0.45 (0.07)
marit=1 proportion (s.d.)	0.51 (0.00)	0.51 (0.01)	0.50 (0.50)	0.52 (0.01)	0.52 (0.01)	0.52 (0.01)	0.52 (0.01)
divorced=1 proportion (s.d.)	0.08 (0.00)	0.09 (0.00)	0.09 (0.00)	0.09 (0.00)	0.09 (0.00)	0.09 (0.00)	0.09 (0.00)
pub=1 proportion (s.d.)	0.05 (0.00)	0.05 (0.00)	0.05 (0.00)	0.05 (0.00)	0.05 (0.00)	0.05 (0.00)	0.05 (0.00)

Table Appendix B.3 (continued)

Variable	2009	2010	2011	2012	2013	2014	2015
sten=1 proportion (s.d.)	0.01 (0.00)	0.00 (0.00)	0.01 (0.00)	0.00 (0.00)	0.00 (0.00)	0.01 (0.00)	0.01 (0.00)
occ=0 proportion (s.d.)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)	0.05 (0.00)	0.06 (0.00)	0.06 (0.00)
occ=1 proportion (s.d.)	0.02 (0.00)	0.02 (0.00)	0.01 (0.00)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)	0.02 (0.00)
occ=2 proportion (s.d.)	0.02 (0.00)	0.02 (0.00)	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)	0.03 (0.01)	0.03 (0.01)
occ=3 proportion (s.d.)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.03 (0.01)	0.03 (0.00)
occ=4 proportion (s.d.)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)
occ=5 proportion (s.d.)	0.10 (0.01)	0.10 (0.01)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.11 (0.01)	0.11 (0.01)
occ=6 proportion (s.d.)	0.22 (0.02)	0.22 (0.02)	0.23 (0.02)	0.23 (0.02)	0.23 (0.02)	0.19 (0.03)	0.18 (0.02)
occ=7 proportion (s.d.)	0.07 (0.00)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)	0.07 (0.00)	0.06 (0.00)
occ=8 proportion (s.d.)	0.04 (0.00)	0.04 (0.00)	0.04 (0.00)	0.05 (0.00)	0.04 (0.00)	0.05 (0.01)	0.06 (0.01)
ind=1 proportion (s.d.)	0.24 (0.03)	0.23 (0.03)	0.24 (0.03)	0.25 (0.03)	0.24 (0.03)	0.20 (0.03)	0.19 (0.03)

Table Appendix B.3 (continued)

Variable	2009	2010	2011	2012	2013	2014	2015
ind=2 proportion (s.d.)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
ind=3 proportion (s.d.)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
ind=4 proportion (s.d.)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)	0.06 (0.00)	0.07 (0.01)	0.07 (0.01)
ind=5 proportion (s.d.)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)
ind=6 proportion (s.d.)	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)
ind=7 proportion (s.d.)	0.16 (0.01)	0.16 (0.01)	0.16 (0.01)	0.15 (0.01)	0.15 (0.01)	0.17 (0.02)	0.17 (0.02)
ind=8 proportion (s.d.)	0.07 (0.01)	0.07 (0.01)	0.07 (0.01)	0.07 (0.01)	0.07 (0.01)	0.07 (0.01)	0.08 (0.01)

Source: Author’s calculation from Labor Force Survey

Table Appendix B.4

The summarize statistic of each variable in the provincial level model

Variable	2001	2002	2003	2004	2005	2006	2007	2008
No. of observation	76	76	76	76	76	76	76	76
lninc mean(s.d.)	8.59 (0.03)	8.59 (0.02)	8.61 (0.02)	8.67 (0.02)	8.74 (0.02)	8.82 (0.02)	8.84 (0.02)	8.96 (0.03)
sch mean(s.d.)	4.99 (0.11)	5.13 (0.11)	5.33 (0.10)	5.54 (0.10)	5.60 (0.12)	5.68 (0.12)	4.56 (0.10)	4.72 (0.08)
exp mean(s.d.)	27.50 (0.25)	27.54 (0.25)	27.67 (0.245)	27.72 (0.24)	28.24 (0.28)	28.41 (0.30)	22.57 (0.25)	23.03 (0.23)
expsq mean(s.d.)	761.05 (13.74)	762.75 (13.67)	769.87 (13.61)	772.51 (13.32)	803.28 (15.57)	813.75 (16.50)	514.35 (11.37)	534.19 (10.74)

Table Appendix B.4 (continued)

Variable	2009	2010	2011	2012	2013	2014	2015
No. of observation	76	76	76	76	76	76	76
lninc mean(s.d.)	8.93 (0.02)	9.00 (0.02)	9.09 (0.02)	9.15 (0.02)	9.24 (0.02)	9.33 (0.02)	9.35 (0.02)
sch mean(s.d.)	4.89 (0.08)	5.08 (0.07)	5.18 (0.07)	5.28 (0.09)	5.48 (0.09)	5.36 (0.13)	5.47 (0.12)
exp mean(s.d.)	23.24 (0.23)	23.44 (0.23)	23.66 (0.22)	23.83 (0.23)	24.02 (0.23)	25.52 (0.32)	25.75 (0.32)
expsq mean(s.d.)	544.15 (10.96)	553.41 (10.70)	563.39 (10.46)	571.87 (10.87)	580.97 (11.24)	659.02 (16.20)	670.33 (16.10)

Source: Author’s calculation from Labor Force Survey

APPENDIX C
Table of the estimated results

Table appendix C.1
The result of the individual models

	Indiv_model1	Indiv_model2	Indiv_model3	Indiv_model4
sch	0.1159***	0.0730***	0.0704***	0.0577***
	(0.0034)	(0.0020)	(0.0020)	(0.0025)
exp	0.0491***	0.0404***	0.0368***	0.0324***
	(0.0008)	(0.0007)	(0.0006)	(0.0005)
expsq	-0.0005***	-0.0004***	-0.0004***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
hr		0.0062***	0.0062***	0.0069***
		(0.0009)	(0.0009)	(0.0009)
male		0.1447***	0.1462***	0.1482***
		(0.0098)	(0.0076)	(0.0078)
munc		0.2085***	0.0756***	0.0595***
		(0.0506)	(0.0091)	(0.0079)
marit		-0.0040	0.0204***	0.0340***
		(0.0068)	(0.0057)	(0.0056)
divorced		-0.0776***	-0.0470***	-0.0453***
		(0.0102)	(0.0095)	(0.0094)
pub		-0.0543	0.0763	0.1388**
		(0.0560)	(0.0523)	(0.0508)
sten		0.3483***	0.3574***	0.4385***
		(0.0428)	(0.0454)	(0.0506)
0.occ		0.0000	0.0000	0.0000
		(.)	(.)	(.)
1.occ		0.6065***	0.5992***	0.6269***
		(0.1269)	(0.1258)	(0.1332)
2.occ		0.6695***	0.6603***	0.7245***
		(0.0083)	(0.0076)	(0.0111)
3.occ		0.4427***	0.4082***	0.4707***
		(0.0094)	(0.0117)	(0.0069)
4.occ		0.2964***	0.2675***	0.3188***
		(0.0155)	(0.0110)	(0.0129)
5.occ		0.1463***	0.1496***	0.1359***
		(0.0121)	(0.0094)	(0.0080)

6.occ	-0.0164	-0.0314	0.0209
	(0.0416)	(0.0310)	(0.0282)
7.occ	0.0622***	0.0811***	0.1033***
	(0.0091)	(0.0058)	(0.0044)
8.occ	0.1506***	0.1340***	0.1424***
	(0.0136)	(0.0115)	(0.0111)
1.ind	0.0000	0.0000	0.0000
	(.)	(.)	(.)
2.ind	0.3230***	0.3309***	0.3120***
	(0.0375)	(0.0390)	(0.0326)
3.ind	0.4178***	0.3986***	0.3267***
	(0.0357)	(0.0384)	(0.0363)
4.ind	0.3084***	0.2183***	0.2148***
	(0.0287)	(0.0211)	(0.0190)
5.ind	0.4333***	0.3131***	0.2991***
	(0.0299)	(0.0214)	(0.0187)
6.ind	0.3123***	0.3044***	0.2833***
	(0.0256)	(0.0209)	(0.0183)
7.ind	0.2962***	0.2422***	0.2373***
	(0.0339)	(0.0253)	(0.0225)
8.ind	0.3651***	0.2968***	0.2719***
	(0.0377)	(0.0277)	(0.0243)
10.cwt		0.0000	0.0000
		(.)	(.)
11.cwt		-0.0519	-0.1051***
		(0.0363)	(0.0303)
12.cwt		0.0080	-0.0213*
		(0.0235)	(0.0094)
13.cwt		-0.0668**	-0.1365***
		(0.0209)	(0.0254)
14.cwt		-0.1993***	-0.2418***
		(0.0263)	(0.0247)
15.cwt		-0.2642***	-0.2908***
		(0.0147)	(0.0182)
16.cwt		-0.2396***	-0.2611***
		(0.0150)	(0.0195)
17.cwt		-0.3358***	-0.3538***
		(0.0149)	(0.0146)
18.cwt		-0.3885***	-0.4205***
		(0.0314)	(0.0165)
19.cwt		-0.1469***	-0.1862***
		(0.0281)	(0.0182)
20.cwt		-0.0596***	-0.1194***

	(0.0163)	(0.0277)
21.cwt	-0.0054	-0.0762**
	(0.0170)	(0.0252)
22.cwt	-0.1558**	-0.1998***
	(0.0480)	(0.0517)
23.cwt	-0.1403***	-0.2177***
	(0.0265)	(0.0309)
24.cwt	-0.1169***	-0.1695***
	(0.0177)	(0.0191)
25.cwt	-0.2297***	-0.2903***
	(0.0204)	(0.0186)
26.cwt	-0.1818***	-0.2331***
	(0.0194)	(0.0178)
27.cwt	-0.3869***	-0.4383***
	(0.0232)	(0.0233)
30.cwt	-0.2987***	-0.3400***
	(0.0134)	(0.0158)
31.cwt	-0.4975***	-0.5667***
	(0.0297)	(0.0327)
32.cwt	-0.4848***	-0.5396***
	(0.0217)	(0.0268)
33.cwt	-0.6592***	-0.7004***
	(0.0630)	(0.0733)
34.cwt	-0.4130***	-0.4755***
	(0.0328)	(0.0446)
35.cwt	-0.4752***	-0.5172***
	(0.0216)	(0.0291)
36.cwt	-0.5152***	-0.5440***
	(0.0248)	(0.0315)
37.cwt	-0.6371***	-0.6635***
	(0.0618)	(0.0724)
39.cwt	-0.5624***	-0.5950***
	(0.0387)	(0.0398)
40.cwt	-0.4072***	-0.4368***
	(0.0441)	(0.0429)
41.cwt	-0.3774***	-0.4121***
	(0.0167)	(0.0221)
42.cwt	-0.4685***	-0.4739***
	(0.0439)	(0.0352)
43.cwt	-0.4293***	-0.4664***
	(0.0186)	(0.0211)
44.cwt	-0.5464***	-0.5944***
	(0.0311)	(0.0405)

45.cwt	-0.5552***	-0.5887***
	(0.0419)	(0.0379)
46.cwt	-0.5916***	-0.6248***
	(0.0583)	(0.0475)
47.cwt	-0.5509***	-0.5854***
	(0.0558)	(0.0599)
48.cwt	-0.5655***	-0.5978***
	(0.0550)	(0.0619)
49.cwt	-0.5442***	-0.5747***
	(0.0811)	(0.0678)
50.cwt	-0.3665***	-0.3988***
	(0.0462)	(0.0271)
51.cwt	-0.3851***	-0.4012***
	(0.0257)	(0.0203)
52.cwt	-0.4319***	-0.4525***
	(0.0350)	(0.0309)
53.cwt	-0.4050***	-0.4394***
	(0.0227)	(0.0184)
54.cwt	-0.5205***	-0.5392***
	(0.0155)	(0.0147)
55.cwt	-0.5422***	-0.5642***
	(0.0290)	(0.0309)
56.cwt	-0.5709***	-0.5955***
	(0.0338)	(0.0181)
57.cwt	-0.4784***	-0.5123***
	(0.0299)	(0.0251)
58.cwt	-0.4298***	-0.5060***
	(0.0216)	(0.0277)
60.cwt	-0.3218***	-0.3497***
	(0.0199)	(0.0167)
61.cwt	-0.3339***	-0.3740***
	(0.0189)	(0.0175)
62.cwt	-0.3713***	-0.4221***
	(0.0273)	(0.0244)
63.cwt	-0.5394***	-0.6018***
	(0.0246)	(0.0223)
64.cwt	-0.4562***	-0.5048***
	(0.0162)	(0.0192)
65.cwt	-0.3444***	-0.3846***
	(0.0149)	(0.0172)
66.cwt	-0.4083***	-0.4448***
	(0.0217)	(0.0158)
67.cwt	-0.4208***	-0.4563***

	(0.0170)	(0.0179)
70.cwt	-0.2361***	-0.2729***
	(0.0245)	(0.0260)
71.cwt	-0.2270***	-0.2751***
	(0.0223)	(0.0257)
72.cwt	-0.2549***	-0.2864***
	(0.0158)	(0.0162)
73.cwt	-0.1084***	-0.1557***
	(0.0233)	(0.0265)
74.cwt	-0.0358	-0.1372***
	(0.0212)	(0.0312)
75.cwt	-0.2012***	-0.2444***
	(0.0230)	(0.0176)
76.cwt	-0.2672***	-0.3134***
	(0.0310)	(0.0219)
77.cwt	-0.2138***	-0.2624***
	(0.0307)	(0.0207)
80.cwt	-0.2801***	-0.3218***
	(0.0236)	(0.0205)
81.cwt	-0.1091**	-0.1818***
	(0.0404)	(0.0302)
82.cwt	-0.1156**	-0.1749***
	(0.0374)	(0.0322)
83.cwt	-0.0140	-0.0942***
	(0.0126)	(0.0234)
84.cwt	-0.0149	-0.0633
	(0.0800)	(0.0725)
85.cwt	-0.1505***	-0.2571***
	(0.0302)	(0.0232)
86.cwt	-0.0438	-0.1167**
	(0.0401)	(0.0398)
90.cwt	-0.1442***	-0.1913***
	(0.0198)	(0.0267)
91.cwt	-0.2006***	-0.2610***
	(0.0230)	(0.0225)
92.cwt	-0.1764***	-0.2204***
	(0.0365)	(0.0300)
93.cwt	-0.2533***	-0.2884***
	(0.0166)	(0.0227)
94.cwt	-0.3751***	-0.4456***
	(0.0206)	(0.0165)
95.cwt	-0.2947***	-0.3022***
	(0.0276)	(0.0225)

96.cwt			-0.4358***	-0.4870***
			(0.0209)	(0.0265)
44.yr				0.0000
				(.)
45.yr				0.0243**
				(0.0090)
46.yr				0.0416***
				(0.0109)
47.yr				0.0722***
				(0.0126)
48.yr				0.1056***
				(0.0130)
49.yr				0.1677***
				(0.0164)
50.yr				0.2002***
				(0.0152)
51.yr				0.2648***
				(0.0173)
52.yr				0.2631***
				(0.0160)
53.yr				0.3117***
				(0.0184)
54.yr				0.3761***
				(0.0287)
55.yr				0.4825***
				(0.0234)
56.yr				0.5676***
				(0.0307)
57.yr				0.6093***
				(0.0340)
58.yr				0.6326***
				(0.0307)
_cons	7.0649***	6.6545***	7.0368***	6.9152***
	(0.0408)	(0.0569)	(0.0450)	(0.0462)
Prob>F	0.000	0.000	0.000	0.000
R-sq	0.449	0.566	0.607	0.669
Observations	745099.000	745098.000	745098.000	745098.000

Note: ***, ** and * are significant at 1%, 5% and 10% respectively; Standard Error is in parenthesis, and dependent variable is Lninc

Source: Author's calculation from Labor Force Survey

BIOGRAPHY

Name	Mr. Jetson Chitdecha
Date of Birth	May 19, 1992
Educational Attainment	2014: Bachelor of Economics, Prince of Songkla University, Thailand (First-Class Honors)
Scholarship	Year 2015-2016: Mitsubishi UFJ Foundation Scholarship