PIER

PUEY UNGPHAKORN INSTITUTE
FOR ECONOMIC RESEARCH

Labor Income Inequality in Thailand: the Roles of Education, Occupation and Employment History

by

Nada Wasi, Sasiwimon Warunsiri Paweenawat,
Chinnawat Devahastin Na Ayudhya, Pucktada Treeratpituk and
Chommanart Nittayo

December 2019

Discussion Paper

No. 117

Labor Income Inequality in Thailand:
the Roles of Education, Occupation and Employment History

Nada Wasi[*]

Sasiwimon Warunsiri Paweenawat[+]

Chinnawat Devahastin Na Ayudhya[±]

Pucktada Treeratpituk[±]

Chommanart Nittayo[±]

November 2019

Abstract

Thailand's income inequality has reportedly declined since the mid-1990s. This paper examines possible mechanisms underlying the dynamic patterns of the country's labor income inequality. Using the Thai labor force survey between 1988 and 2017, we document that the country's reduction in income inequality is likely driven by the fact the earnings at the bottom part of the distribution have become more similar. The median wage gap between college and non-college workers, however, still gets larger over time. Our key explanation is the changes in education-occupation composition. Recently college graduates are no longer concentrated in high skill jobs. A larger share of secondary educated workers works in low-skill jobs instead of the middle-skill ones. Using panel administrative data from the Thai Social Security Office, we find that wage disparity can also be explained by employment history. The high wage earners earn more since they enter the market, and the gap gets wider as the workers age. Additionally, the top of the group can command higher wages by working at a large firm or switching to a new job. These findings highlight the fact that to tackle the income inequality issue, the country needs to understand the underlying mechanisms behinds its dynamics.

JEL: J21; J24; J31;O15

[*] Puey Ungphakorn Institute for Economic Research, Bank of Thailand; Corresponding author, e-mail: nadaw@bot.or.th

[+] School of Economics, University of the Thai Chamber of Commerce

[±] Data Analytics Unit, Bank of Thailand

We are grateful to the Thai Social Security Office and National Statistical Office for data access. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Bank of Thailand or Puey Ungphakorn Institute for Economic Research.

1. Introduction

Several existing studies document that Thailand's income inequality has declined over the past two decades (World Bank, 2016; Kilenthong, 2016; UN-ESCAP, 2018). However, Thailand's income inequality level is still high compared to the countries in East Asia (World Bank, 2016). Macroeconomic factors such as economic growth and financial deepening as well as microeconomic factors such as education, occupation, the number of earners reportedly contribute to the changes in Thailand's income inequality (Jeong, 2008; Paweenawat and McNown, 2014). Although some studies, *e.g.*, Pootrakul (2013), have looked at how the income inequality within and between subpopulation changed over time, there is not much discussion on mechanisms underlying the declining inequality.

In this paper, we focus on how education, occupation and employment history interact to affect individual labor income over time. Changes in individual earnings imply the change in the disparity of the aggregate labor income, and disparity within and across subgroups of population. Even though household income consists of both labor income (earned income) and unearned income, we choose to focus on individual's labor income since it is the main source of income for most households and is the income component that best reflects individual's human capital. For Thailand, the evolution of labor income inequality appears mimicking the decreasing trend of the total income inequality being documented.

Our contribution is two-fold. First, using the Thai labor force cross-sectional survey between 1988 and 2017, we show that occupation and education are intertwined in determining both the level and dispersion of labor income. Second, using the panel data from the Thai Social Security records, we explore how employment history could explain the labor income dispersions across different groups of employees and over their working lives. We are not aware of any study exploiting the panel data to examine how such factors are related to income inequality in Thailand.

The individual-level analysis reveals a complex picture behind the decline in the country's income inequality. We find that this declining trend is likely driven by the convergence of labor income among low-skill workers both within- and across- subgroups. The wage differentials between the high-skill and low-skill workers, however, have risen over time. Our key explanation is the changes in education-occupation composition. The number of secondary educated workers have vastly increased but more of them work in low-skill jobs rather than the middle-skill jobs like they used to do in the past. College workers, while some no longer work in high-skill jobs, the talented ones earn much more. We also find that wage disparity can be partly explained by employment history. The high wage earners tend to be those who always work in the formal sector, work in larger firms, and switch jobs for a higher wage over time. The age-earning profiles of low earners are rather flat throughout their working lives and uncorrelated with neither firm size nor job switching.

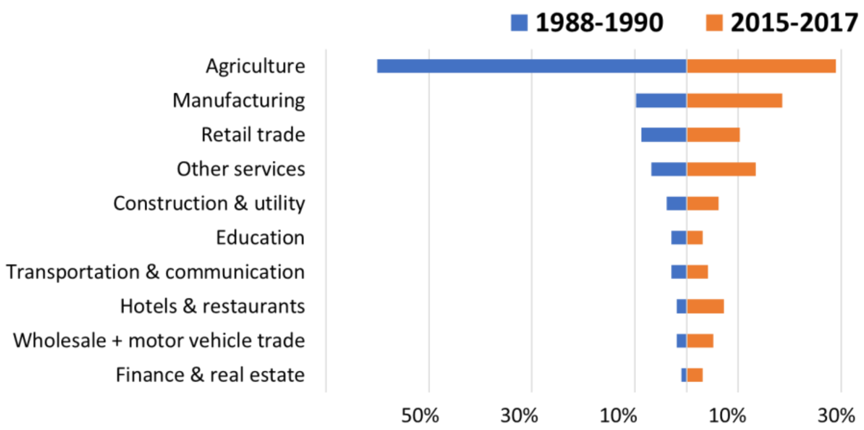
We should note that while we use the word "inequality" or "disparity" to describe the dispersion of the labor income, we take the microeconomic notion that wage differentials likely reflect the combination of workers' and firms' productivity; which are influenced by education, experience, observed and unobserved ability, technology, and institutions. Given heterogeneous abilities, a certain degree of disparity is to be expected in a market economy.

The paper is organized as followed. The next section provides the background about Thailand's economy and related literature. Section 3 describes the two data sets used in the analysis. Section 4 documents how the changing roles of education and occupation explain the changes in inequality. Section 5 explores how individuals' work history is related to their earned income and income paths. The last section provides conclusions and discussion.

2. Background and previous studies

Thailand has been through significant transformations over the past three decades. The economy grew rapidly during 1990s, faced the Asian crisis in 1997, and started to recover since 2000s (Paweenawat and McNown, 2014). The Thai workforce composition has also been changed over time. The share of workforce in the agricultural sector, once the country's backbone, has largely declined from 60% in 1990 to 30% in 2017; while the shares in the manufacturing, trade and service sectors have all risen (see Figure 1).

Figure 1: The shares of workers by industry sector



Source: Authors' calculation from the Thai Labor Force Survey

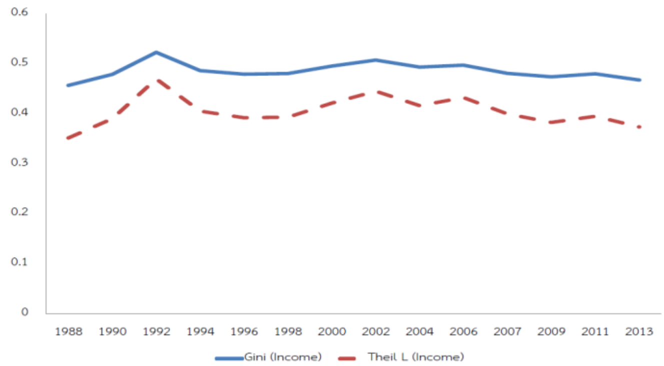
Types of employment, which vary by age and gender, have also changed over time (See Figure A1 in Appendix A). In 1988-1990, a large proportion of prime age men were self-employed in the agricultural sector. Three decades later, the private sector accounts for more than 40% of those aged 25-44 years old. Women has changed their working status from being unpaid family workers to being either the private sector employees or self-employed workers. The number of employees in the Thai Social Security record reached 11 million in 2018, almost doubled from 2002, confirming the increasingly importance of the country's formal sector.

Several studies have explored how Thailand's income inequality has evolved over time. Relying on the aggregate measures of inequality such as the Gini coefficient and Theil index, most studies find that Thailand's income inequality is inversely related to the country's economic growth. Between 1970s and the early 1990s, Thailand's income inequality increased (Kakwani and Krongkaew, 2003; Jeong, 2008). During the mid-1990s to the mid-2010s, Thailand's inequality has reportedly declined (Paweenawat and McNown, 2014; World Bank, 2016; Kilenthong, 2016; UN-ESCAP, 2018; See Figure 2). However, Vanitcharoentham (2017) suspects that the estimates from the earlier studies using household survey data could be biased because of its poor coverage of the rich. By using tax return data to re-estimate the Gini coefficient, he finds that the income inequality increased between 2007 and 2009.

Some of these studies also explore whether certain factors are related to the shape of the aggregate income distribution. Jeong (2008) reports that the joint composition change of education, occupation and financial deepening explain 53-58% of the increase in inequality during 1976-1996. Paweenawat and McNown (2014) find that the variance of education and number of earners are positively associated with the variance of log of income per capita. Pootrakul (2013), on the other hand, finds that while the inequality between groups increased, the intra-inequality decreased for those with education less than college degree and increased among the college graduates.

A few studies look at other measures of inequality. Kilenthong (2016) reports that both consumption inequality and wealth inequality (proxied by vehicle and mobile phone possession) has declined over time. Laovakul (2016) yet reports a high degree of landownership inequality. Approximately 60% of the land in Thailand is owned by the top decile of the country's landowners.

Figure 2 : Thailand's income inequality over time



Source : Kilenthong (2016)

While the Thai literature provides an overview of how the aggregate inequality has changed over time, there is not much discussion about the underlying mechanism. Most studies in the developed countries insist that education, labor market and progressive tax policies play key roles in explaining differences in income inequality across countries. Higher inequality is also associated with lower social mobility (Krueger, 2012). In the US, inequal opportunities when young such as inequal access to education and good neighborhood have a far-reaching effect over people's lives (Heckman 2006; Chetty and Hendren, 2018). In addition, technological progress plays a role in driving earnings disparity. Computers complement the non-routine cognitive tasks of high-skill jobs but substitute the routine tasks of middle-skill jobs. Empirical studies find that wages of college graduates rose; while wages of non-colleges fell after the computer revolution (Autor *et al.*, 2006; Autor 2019).

The economic life-cycle model predicts that individual wages and wage growth also depend on employment history. Wages are expected to grow over time as workers accumulate more human capital through learning-by-doing (*e.g.*, Keane and Wasi, 2016). Wages are also found to be related to job tenure, industry experience and firm sizes. Large firms are often argued to operate more efficiently and benefit from their size as they can purchase inputs at lower costs and invest more on efficient production technology. Empirically, larger firms in the US are found to pay observationally equivalent workers higher wages (Brown and Medoff, 1989; Abowd et al., 2019).

Overall, there are many possible mechanisms which would drive earnings' disparity. Some are institutional and macroeconomic factors, but many of them are individual-level factors. We now turn to describe our data in the next section.

3. Data

Our analysis relies on two data sources. The first data set is the Thai Labor Force Survey (LFS), which is administered on a quarterly basis by the National Statistical Office (NSO) of Thailand. The third quarter (July–September) rounds covering the years 1988-2017 are used. The LFS is a national representative sample, collecting detailed information about individuals' education, work status, employment sector, industry, and earnings. The number of observations in each year is approximately 100,000. The weighted sample represents about 19 million of prime age population (age 25-54) in 1988 and that number increases to 30 million in 2017.

An individual working status is reported as private sector employee, government or state enterprise employee, self-employed (with and without employees), unpaid family worker, and not working. However, wages and earnings for self-employed workers are not observed.[1] Wages are reported in term of daily, weekly or monthly basis; while working hours are reported on a weekly basis. We converted all types of wages into hourly wage, which is later deflated by the Thailand Consumer Price Index (CPI) using 2015 as the base year.[2]

The second data set is the Social Security Office (SSO) employment data, which is a panel administrative data providing additional dynamic aspects. No previous studies has utilized this type of data on examining income inequality in Thailand. The Thai Social Security Act was enacted in 1990. The Act originally required employers in non-agricultural sectors with 20 or more employees to register for and contribute to the Social Security (SS) fund.[3] It was later extended to cover employers with 10 or more employees in 1993 and employers with at least one employee in 2002. This compulsory mandate, also known as Article 33, requires contribution from 3 parties (5% of monthly wage from employee, 5% of monthly wage from employer, and 2.75% of monthly wage from the government).

The minimum monthly wage base has been 1,650 and the maximum SS taxable wage has been 15,000 since the fund started. The number of employees earned more than this cap was 10% in 2002, but was about 33% in 2018. Similar to LFS, wages are deflated by the 2015 CPI. We create two sets of data from two cohorts of workers and follow them for 96 months. The first cohort consists of 5.2 million workers aged 15-44 years in April 2002. The second cohort consists of 6.8 million workers aged 15-44 years in April 2010. The two cohorts are

[1] Although another commonly used Thai Socio-Economic Survey includes all types of income, its household sampling frame makes individual-level analysis inappropriate. While there often are multiple earners per household, using that survey would limit the analysis to use information of household head only.
[2] The constant of 4.3 is used to convert weeks to months.
[3] Government, state enterprise and school staff are excluded as they are covered by other social insurance schemes.

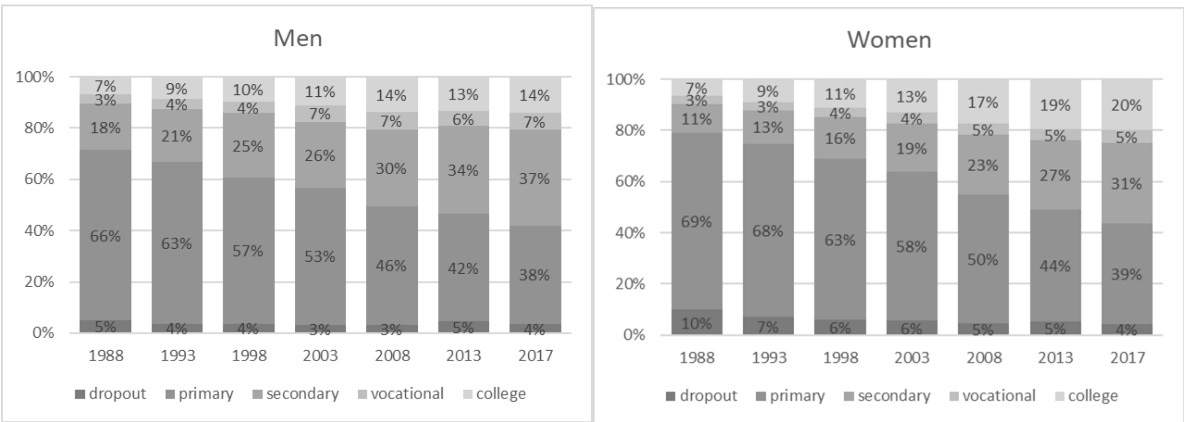
accounted for 77% and 82% of SSO employees in their corresponding year, respectively. This paper presents only analysis of the first cohort. The results of the second cohort are similar in most dimensions and are available upon request.

While the SSO data lacks information on education, socioeconomic characteristics and actual wages for top earners, its advantages are that (i) we can track individual employees and firms over a long period of time; (ii) administrative data escape the measurement error problem; and (iii) monthly data are better for understanding job entry and exit. These two datasets, hence, complement each other and give us an advantage to explore the relationship of dynamic in earnings and the roles of education, occupation and work history.

4. The intertwining roles of education, occupation and wages

Thailand has heavily invested in education over the past four decades, hoping that reducing inequal opportunities to access to education would enhance human capital and thus improving labor market and other long-term outcomes. The first educational reform was the 1978 six-year primary education compulsory reform (see Hawley 2004; Liao and Paweenawat 2019). Next, the 1999 National Education Act extended the compulsory education to 9 years and provided 12 years of free education. The free education was then stretched to include 3 years of pre-school in 2009.

Figure 3: Education of Thai men and women aged 25-54 years



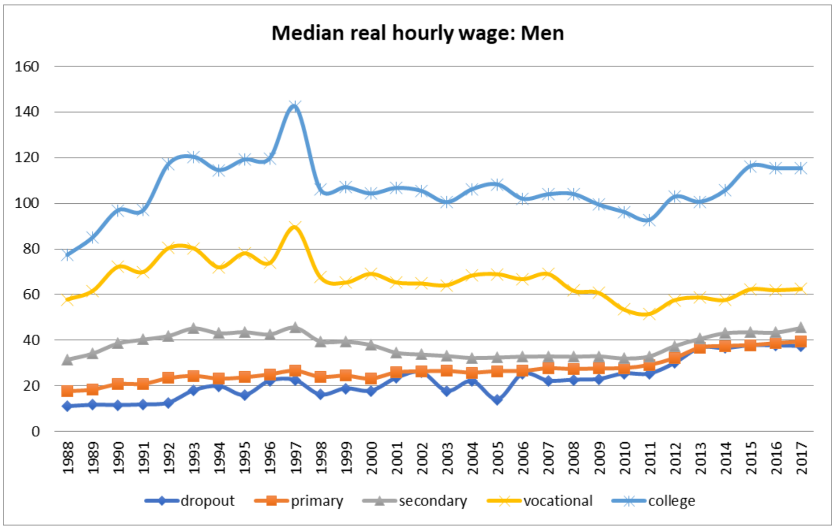
Source: Authors' calculation from LFS for population aged 25-54 years old

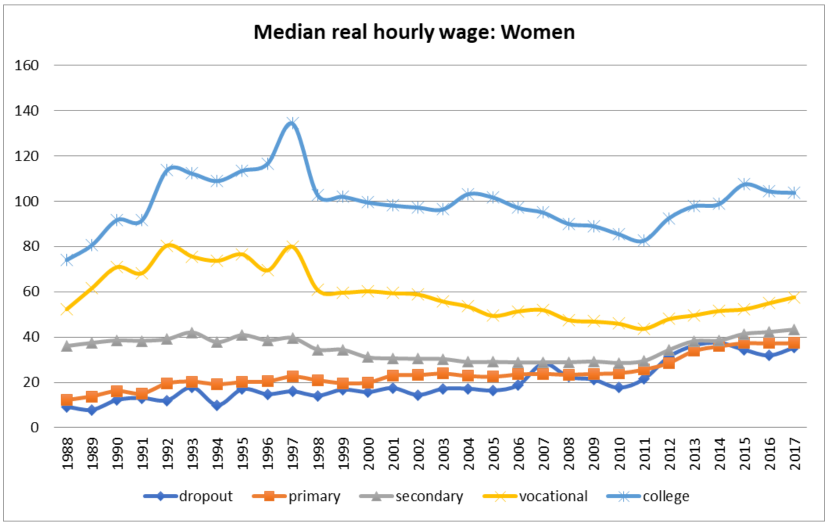
With the education promotion policies and parents' higher investment in education, the average education levels for Thai men and women have risen over time. Figure 3 shows that in

1988, almost 70% of our prime age men and women were primary school graduates. Over time, the shares of college and secondary educated have significantly increased; while the shares of primary educated men and women remain similar.[4]

We observe that while average years in school continuously rose over time, income inequality went up and slightly declined. Figure 4 further explores the relationship between educational wages and education over time at the individual level. Each line plots real median wages for male and female employees with potential experience of 10-20 years.

Figure 4: Median real hourly wage by Education





Note: The real wages are adjusted by the headline CPI (base = 2015).
Source: Authors' calculation from LFS

[4] See Table B1 for education classification.

The pattern of wage gap between the highest and lowest education groups observed in Figure 4 are consistent with the aggregate income inequality presented in Figure 2. While the gap was widest during 1996-1997, it slightly declined over time for both men and women in all education groups. Two observations emerge from this figure. First, the relationship between inequality and education are not stable. While Figure 3 suggests that the average years of education rose monotonically over time, the labor income inequality across education groups went up during the boom and then declined.

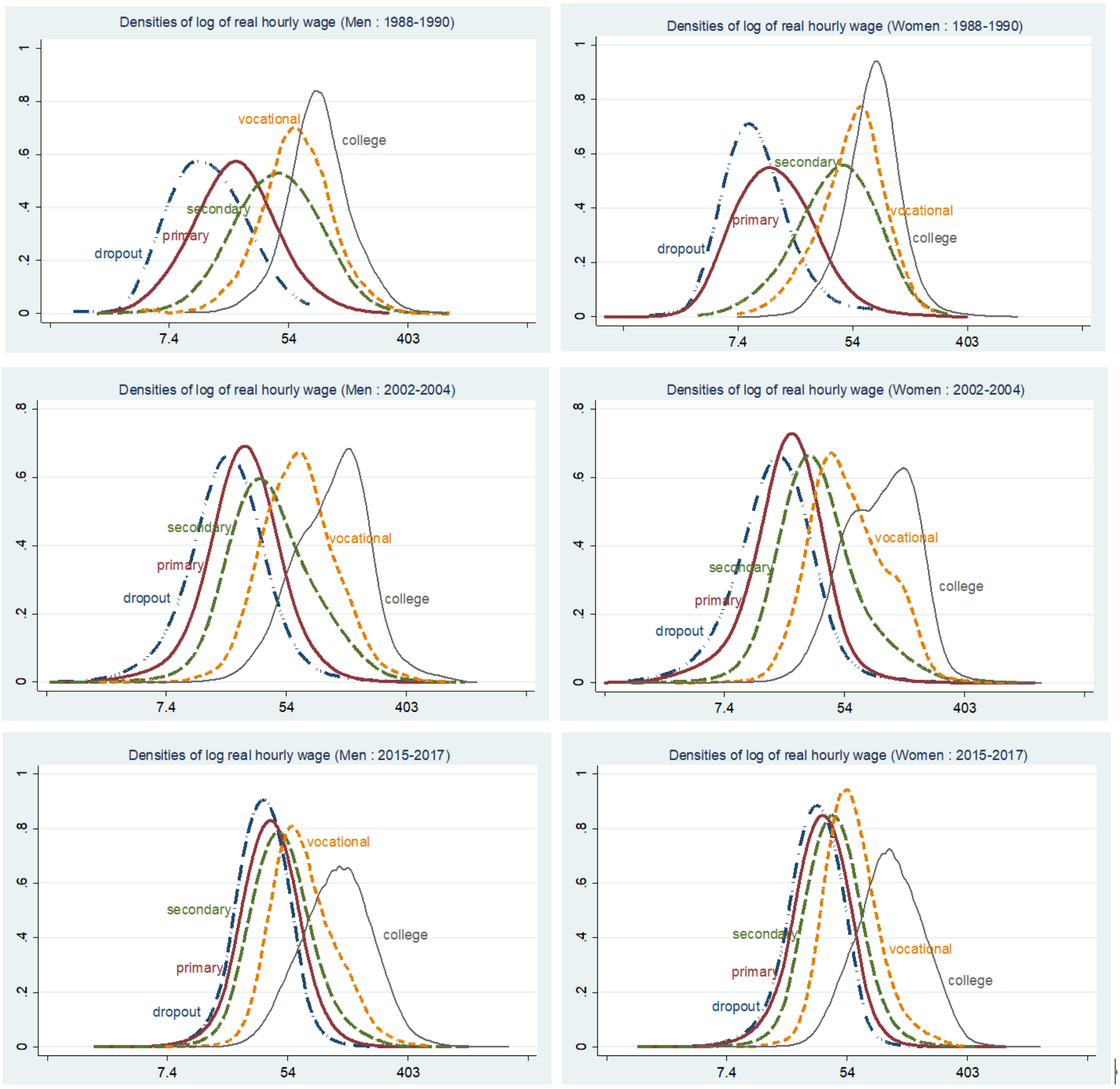
Second, while earlier studies often use “years of schooling” as a measure of education, earnings associated with each additional school year has changed over time. In the late 1980s, median wages of higher educated groups are clearly higher than the lower educated group. Typical college graduates earned more than those with a vocational degree and typical secondary graduates earned more than those completing only primary school. However, over time, the wage gap between secondary and primary groups has diminished and almost disappeared in the late 2010s. College graduates, on the other hand, have experienced a relatively higher increase in their wage, leading to a larger wage gap between college and vocational educated workers.

The data thus far suggest that the decrease in income inequality post-crisis is likely driven by the decrease in inequality from the bottom part of the income distribution. While the share of secondary workers has risen over time, their real wage declined between 1998 and 2003 and only started to rise in 2012. In fact, the median real wages of the bottom three education groups rose after 2012, which is the period where the minimum wages for all provinces were increased by at least 40%.

To explore the within group inequality, Figure 5 plots kernel densities of log real hourly wages for each education level in three periods. During 1988-1990, wage densities for college graduates were least dispersed for both men and women; while the densities of vocational group located close to that of college graduates, especially for women. During 2002-2004, while the wage densities of the dropout, primary and secondary groups were less dispersed than the previous period; the density of the college group was flatter. The density of vocational group also moved closer to that of the secondary group for both men and women.

Finally, in 2015-2017, the wage density of the college graduates had become the most dispersed density. This outcome is consistent with the explanations that the proportion of post-graduates also increased over time; and that technical progress led to higher wages of high skilled workers. However, its larger mass at the left side could reflect heterogeneous abilities or excess supply. The wage distributions of the dropout, primary and secondary graduates had become similar. While Figures 4 and 5 suggest that the wage differentials between primary and secondary educated has diminished, below we will show that their employment probabilities still differ.

Figure 5: Densities of log of real hourly wage by education level



Source : Authors' kernel density estimates from LFS

Next, we turn to occupation, another factor found to be related to inequality. We categorize occupations into 13 categories where they are further clustered into 3 broad occupation groups: (1) high-skill occupations including managers, professionals, technicians and associated professionals; (2) middle-skill occupations including clerical and sale/service workers; and (3) low-skill occupations including agricultural workers, craft and manual workers, machine operators, assemblers, drivers and laborers. Table 1 provides descriptions for these occupations.[5]

Table 2 presents the occupational shares of workers in 1988-1990, and 2015-2017. The shares of high-skill and middle-skill groups moderately increased (+6.1 and +9.6 percentage points, respectively) and these increases were offset by the fall of low-skill occupations (-15.6 percentage points). Among high-skill jobs, the shares of technicians and associated profession rose the most. For the middle-skill jobs, the increase is largely driven by doubling of the shares of service and sale workers. Regarding the low-skill category where its share is still largest at 63%, the share of agricultural jobs largely fell while the shares of other low-skill jobs rose, especially the machine operator group. These changes are likely driven by the growth of manufacturing, trade and service sectors and/or the changing labor demand for different types of skills led by the computer revolution.

Figure 6 present the real hourly wage at 10th , 25th , 50th , 75th and 90th percentiles by occupations in 1988-1990 and 2015-2017 for men and women. Three key facts can be seen in this figure. First, our occupation classification does reflect skills. Most jobs categorized as high-skill jobs do pay higher wages than jobs in the middle-skill category.[6] Likewise, jobs in the middle-skill category pay higher wages than those in low-skill category. Second, the higher skill the jobs require, the wider their wage dispersion, implying that a tiny fraction of workers receive wage much higher than the rest. Finally, the median occupational wages are quite stable over time and across gender. The median wages of the same occupations for men and women are similar, suggesting that gender wage differentials are partly driven by different occupations held by men and women.

We have shown that overtime the Thai workforce were more educated, the shares of high- and middle-skill jobs moderately increased, the occupational wage differentials have not changed much, but how do these explain the changing inter- and intra- inequality across education groups in Figures 3 and 4.

[5] See Appendix B for occupation classification.
[6] The exceptions are artists and journalists. This group probably contains workers with more heterogeneous abilities than other groups.

Table 1: Occupation classification

Occupation		Job examples	Likely skills and education required
High skill			
1	Managers, legislators, senior officials	managers, school principals	
	Professionals		
2	Sciences, doctors, engineers, college professors	physicians, engineers, architects, biologists	Complex problem solving and decision making based on knowledge in a specialized field. Most occupations here require college degree. Vocational degree may be substituted by experience.
3	Business/finance related professionals	business analysts, economists,	
4	Lawyers and other social science professionals	lawyers, HR-related professionals	
5	Artists and journalists	movie directors, journalists, composers	
6	School teachers and associates		
7	Technician & associate professionals	medical technicians, health-safety inspectors, associated nurses	
Middle skill			Making written record of work completed; performing simple arithmetic; good interpersonal communication
8	Clerical support workers	clerks, secretaries, accountants	
9	Service and sales workers	shop sale assistance, hairdressers	
Low skill			Need skills can acquired by on-the-job training
10	Agricultural, forestry, and fishery workers	farmers, fishermen	Agricultural work, not related to research
11	Craft and related trade workers	plant sewing workers, craft workers	Manual tasks, sorting, sewing
12	Plant and machine operators, assemblers, drivers	A/C repairers, bus drivers	Operating or repairing machinery and electronic equipment; assemblers; driving vehicles
13	Laborers in non-agricultural sector	cleaner, gardener, construction workers	Simple & routine physical tasks

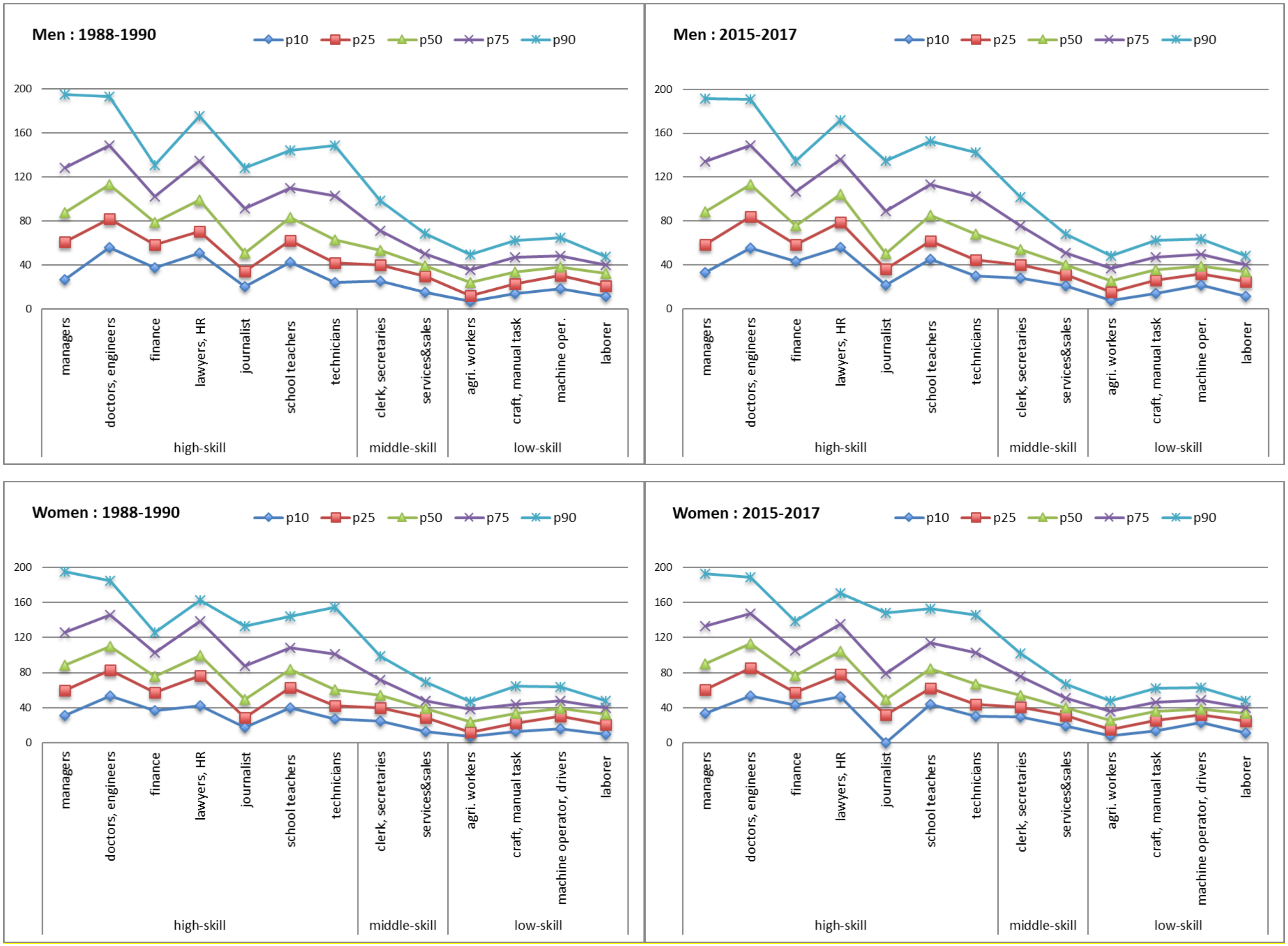
Source: Authors' classification based on task similarities and compatibility of classifications across years

Table 2: Occupational shares among workers in 1988-1990 and 2015-2017

Occupation	1988-1990 (%)	2015-2017 (%)	Change in share (percentage point)
High-skill occupations	**8.3**	**14.4**	**6.1**
Managers, legislators, senior officials	2.8	3.9	1.1
Sciences, doctors, engineers, college professors	0.6	1.7	1.1
Business/finance related professionals	0.5	1.6	1.1
Lawyers and other social science professionals	0.3	1.0	0.6
Artists and journalists	0.2	0.3	0.1
School teachers and associates	2.7	2.5	-0.2
Technician & associate professionals	1.2	3.4	2.2
Middle-skill occupations	**12.9**	**22.5**	**9.6**
Clerical support workers	4.1	4.9	0.8
Service and sales workers	8.8	17.6	8.7
Low-skill occupations	**78.8**	**63.1**	**-15.6**
Agricultural, forestry, and fishery workers	60.1	28.9	-31.2
Craft and related trade workers	9.7	14.2	4.5
Plant and machine operators, assemblers, drivers	3.8	11.3	7.5
Laborers in non-agricultural sector	5.2	8.8	3.6

Source: Authors' calculation from LFS. The sample consists of workers who were 25-54 years old.

Figure 5: Median real hourly wage by occupation (Unit: Baht, base in 2015)



Source: Authors' calculation from LFS

Tables 3-4 present the within-group occupation composition change by education, the missing puzzle. In 1988-1990, approximately 80% of college graduates worked in the high-skill occupation; while the vocational group split between high-skill and middle-skill occupation. For secondary educated workers, men had more chance to work in high-skill jobs than women and had a higher share of low-skill jobs. Most women with secondary education who worked tended to work in a middle-skill job. The pattern of men taking low-skill jobs where women taking middle-skill jobs exemplifies when looking at the primary or lower education group. This is not too surprising because many low-skill jobs are more physically demanding, and some middle-skill jobs like secretary or clerk traditionally belong to women.

The pattern dramatically changed in 2015-17. The shares of college men and women employed in high-skill jobs were greatly reduced by 28 and 26 percentage points, respectively. For men, the fall was compensated by the equal rise of the shares of middle-skill jobs and low-skill jobs by 14 percentage points each. For women, the share of middle-skill jobs rose more (+18 percentage points), but also the fraction of women who chose not to work (+4 percentage points). The fact that college graduated was concentrated in high-skilled jobs in the past, but now more of them work in lower skill jobs could explain why a flatter distribution of the college wage density in recent years.

For vocational degree workers, the chance of being employed in a high-skill job has gone down from approximately 40% to 15% for both men and women. The share of middle-skill jobs rose for women but fell for men. The patterns of moving down the occupation-skill ladder for secondary graduates are similar to the vocational group except that the share of middle-skill occupations fell more substantially. Regarding to the primary and low-educated groups, while Figures 3-4 suggest that these groups are slightly better off as their real wages have caught up with the secondary graduates, the last rows of Tables 3 and 4 reveal a different picture. A larger fraction of primary or lower educated men and women aged 25-54 did not work compared to their counterparts three decades ago.[7]

[7] LFS also asked individuals who do not work on why they were not ready to work. For the primary or lower education, reports of disability as the reason increased over time. Since LFS is a cross-sectional survey, we cannot distinguish whether disability led them to exit the workforce or not-working led them to state the reason.

Table 3: Occupational shares among men workers by education

Occupation	1988-1990					2015-2017				
	dropout	**primary**	**secondary**	**vocational**	**college**	**dropout**	**primary**	**secondary**	**vocational**	**college**
High-skill	**3%**	**3%**	**22%**	**40%**	**80%**	**1%**	**2%**	**6%**	**15%**	**52%**
Managers, legislators, senior officials	2.1%	2.7%	7.5%	13.0%	26.2%	0.5%	1.9%	3.9%	5.1%	13.0%
Sciences, doctors, engineers, college professors			0.1%	0.9%	9.2%			0.2%	1.0%	8.0%
Business/finance related professionals			0.2%	1.6%	6.7%			0.1%	0.4%	1.5%
Lawyers and other social science professionals			0.2%	0.7%	6.7%			0.1%	0.3%	4.9%
Artists and journalists		0.2%	0.9%	1.3%	0.8%	0.1%	0.1%	0.4%	0.3%	0.7%
School teachers and associates	0.1%	0.04%	1.7%	17.0%	25.8%	0.02%	0.01%	0.1%	0.2%	10.1%
Technician & associate professionals	0.3%	0.4%	11.5%	5.9%	4.4%	0.2%	0.3%	1.9%	7.7%	13.3%
Middle-skill	**11%**	**11%**	**35%**	**31%**	**13%**	**7%**	**10%**	**21%**	**26%**	**27%**
Clerical support workers	0.4%	1.1%	15.0%	19.4%	4.8%	0.1%	0.2%	2.2%	5.6%	6.6%
Service and sales workers	10.6%	10.2%	19.8%	11.7%	7.9%	6.7%	9.4%	18.8%	20.9%	20.0%
Low-skill	**78.3%**	**82.9%**	**38.5%**	**25.4%**	**4.2%**	**75.9%**	**82.4%**	**67.5%**	**55.0%**	**18.1%**
Agricultural, forestry, and fishery workers	57.6%	53.8%	10.5%	5.0%	1.1%	40.1%	50.0%	28.4%	14.3%	6.7%
Craft and related trade workers	8.6%	13.3%	15.3%	16.4%	1.8%	12.4%	14.9%	16.8%	24.0%	6.3%
Plant and machine operators, assemblers, drivers	3.7%	9.0%	8.5%	2.9%	0.5%	8.6%	9.3%	15.7%	14.1%	4.2%
Laborers in non-agricultural sector	8.6%	6.9%	4.2%	1.0%	0.7%	14.8%	8.2%	6.5%	2.6%	0.9%
Not work	**8.1%**	**2.4%**	**4.5%**	**3.1%**	**3.3%**	**16.6%**	**5.5%**	**5.1%**	**3.5%**	**3.7%**

Source: Authors' calculation from LFS for men who were 25-54 years old.

Table 4: Occupational shares among women workers by education

Women	1988-1990					2015-2017				
	dropout	**primary**	**secondary**	**vocational**	**college**	**dropout**	**primary**	**secondary**	**vocational**	**college**
High skill	**1%**	**1%**	**14%**	**40%**	**78%**	**0%**	**1%**	**4%**	**15%**	**52%**
Managers, legislators, senior officials	0.4%	0.7%	2.3%	3.8%	11.0%	0.1%	0.5%	1.5%	2.6%	6.3%
Sciences, doctors, engineers, college professors	0.0%	0.0%	1.9%	5.5%	9.9%		0.0%	0.1%	0.5%	8.5%
Business/finance related professionals		0.0%	0.2%	0.5%	8.5%		0.0%	0.6%	5.9%	7.6%
Lawyers and other social science professionals		0.0%	0.2%	0.4%	4.3%	0.0%	0.0%	0.1%	0.7%	4.1%
Artists and journalists	0.1%	0.1%	0.3%	0.1%	0.3%	0.0%	0.1%	0.1%	0.1%	0.4%
School teachers and associates	0.0%	0.1%	4.5%	27.3%	42.1%	0.0%	0.0%	0.3%	0.9%	17.9%
Technician & associate professionals	0.1%	0.2%	4.3%	2.7%	1.9%	0.1%	0.3%	1.4%	4.4%	7.4%
Middle skill	**14.6%**	**18.0%**	**45.9%**	**39.4%**	**14.1%**	**11.9%**	**18.3%**	**33.0%**	**46.0%**	**31.6%**
Clerical support workers	0.3%	0.7%	23.1%	27.6%	7.6%	0.1%	0.4%	3.9%	16.3%	14.5%
Service and sales workers	14.3%	17.3%	22.8%	11.8%	6.5%	11.7%	17.9%	29.1%	29.7%	17.2%
Low skill	**59.5%**	**62.2%**	**16.1%**	**5.8%**	**1.8%**	**56.1%**	**59.8%**	**41.0%**	**20.2%**	**5.8%**
Agricultural, forestry, and fishery workers	45.7%	45.7%	5.0%	2.8%	0.4%	34.6%	41.2%	19.6%	7.8%	2.7%
Craft and related trade workers	5.5%	8.7%	6.8%	1.7%	0.6%	5.0%	5.9%	6.1%	4.2%	1.6%
Plant and machine operators, assemblers, drivers	0.7%	1.0%	0.9%	0.5%	0.1%	4.0%	3.5%	7.7%	4.5%	0.8%
Laborers in non-agricultural sector	7.7%	6.8%	3.4%	0.7%	0.7%	12.5%	9.2%	7.6%	3.7%	0.7%
Not work	**25.2%**	**18.7%**	**24.3%**	**14.5%**	**6.0%**	**31.7%**	**21.0%**	**21.7%**	**18.7%**	**10.4%**

Source: Authors' calculation from LFS for women who were 25-54 years old.

In this section, we demonstrate that the relationship of education, occupation and labor income over time are intertwined and depend on many factors. The fact that the college graduates are no longer concentrated in the high-skill jobs is also consistent with the earlier study by Paweenawat and Vechbanyongratana (2015) who examine types of jobs performed by recent college graduated men. In the next section, we will look at the role of dynamic work pattern, the feature we do not observe from the repeated cross-section survey.

5. Employment history and labor income growth

Besides education, labor market experience is known to contribute to wage differentials across individuals and over an individual's working life. In this section, we employ the panel data from SSO to explore the role of employment history in driving wage disparity. The information about individual work pattern is very rich. Some workers switched jobs very often while others held only one job. Some workers also exit the formal labor market and return. It is impossible to analyze wage or wage growth by every possible path of work history. To reduce the set of work patterns, we use the k-means clustering technique to group workers with similar employment profiles together. This technique allows us to find natural segmentations of workers based on the given profiles. Appendix C provides the details about k-means clustering. Six employment-related profiles, which capture different sets of information, are chosen to characterize each employee's work history over 96 months. The profiles are explained below:

1) the number of months since 2002 that a worker is observed in the Social Security record including absent months (the last month observed – April 2002)
2) the total number of jobs over 96 months[8]
3) the number of unemployment spell (number of times exiting the formal sector)
4) the median job tenure (number of months that a worker works consecutively for the same employer)
5) the median length of unemployment spell (number of consecutive months that a worker has no record in the formal sector)
6) Same firm repetition (the number of times a worker works for the same employer in multiple job spells)

[8] We allow for a maximum of one job per month. Approximately 22% of employees have at least one record showing more than one employer in the same month. Most of these cases are when workers switching from one job to another job.

Using the above employment-related profiles, the k-mean clustering technique suggests that there exist four distinct clusters of Thai employee work patterns. Table 5 presents the summary statistics (25^{th} 50^{th} and 75^{th} of variables in each cluster). The first cluster consists of 38% of employees. We label them *Stayers* because they tend to stay with one job for a long period of time (the median job tenure is 96 months). Most of them have 1-2 jobs within the span of 8 years (90%) but some also have 3 or more jobs. This group does not have an unemployment period when switching jobs.

The second cluster consists of 33% of employees. This group typically have 2-4 jobs within the span of 8 years and generally have at least some periods of unemployment between jobs. The median length of their unemployment is 5 months. Because they often move jobs and move in-and-out of the formal sector, we call them *Movers*. The third cluster accounting for 14% of employees represents seasonal behaviors, hence being labeled *Seasonal*. A typical *Seasonal* worker works 3-7 months followed by 2-8 months of unemployment. The last cluster consists of the remaining 15% of employees are those we only observed for a short period of time, typically less than 2 years.

Figure 7 illustrates their typical employment patterns for each cluster. The four clusters reveal that while all of these workers are counted as *employees in the formal sector*, we can confidently call only 38% (the *Stayers*) the formal sector employees over the subsequent 96 months. The other groups feature a hybrid pattern, switching between being in the formal and informal sectors (or inactive). Table 6 presents wage distributions for these four clusters by age groups. The top panel are the observed monthly wage deflated by CPI. Three patterns emerge.

First, the median wage of *Stayers* is always higher than the other three clusters, suggesting that the disparity of labor income inequality is partly driven by the workers work in the formal sector all the time. Second, both labor income inequality between groups (see Figure 8) and the labor income inequality within group (see the interquartile range in Table 6) are larger as the workers age. Note that the interquartile within group for *Stayers* age 31 or above could be underestimated because observed nominal monthly wages in the record have been capped at the maximum contribution base of 15,000 Baht. Lastly, although the median wage increases with age as expected for all groups, the median real wage only increases with ages until about 32 years old for the three hybrid formal-informal groups.

Table 5: The summary statistics of the employment profiles in each cluster

Percentile	cluster 1 (38%) "*Stayers*" 25th	50th	75th	cluster 2 (33%) "*Movers*" 25th	50th	75th	cluster 3 (14%) "*Seasonal*" 25th	50th	75th	cluster 4 (15%) "*Shortly Observed*" 25th	50th	75th
number of months observed in SS from 2002-2010	96	96	96	77	96	96	77	96	96	7	16	28
total number of jobs	1	1	2	2	3	4	5	6	8	1	1	1
number of unemployment spell	0	0	0	1	1	2	2	3	4	0	0	0
job tenure	48	96	96	9	16	30.5	3	4	7	5	12	23
length of unemployment spell	0	0	0	2	5	15.5	2	4	8	0	0	1
number of times working for the same employers	1	1	1	1	1	1	2	2	3	1	1	1

Source: Authors' calculation from Social Security records (2002-2010) for population aged 15-44 years old

Figure 7: Examples of employment patterns for different types of workers

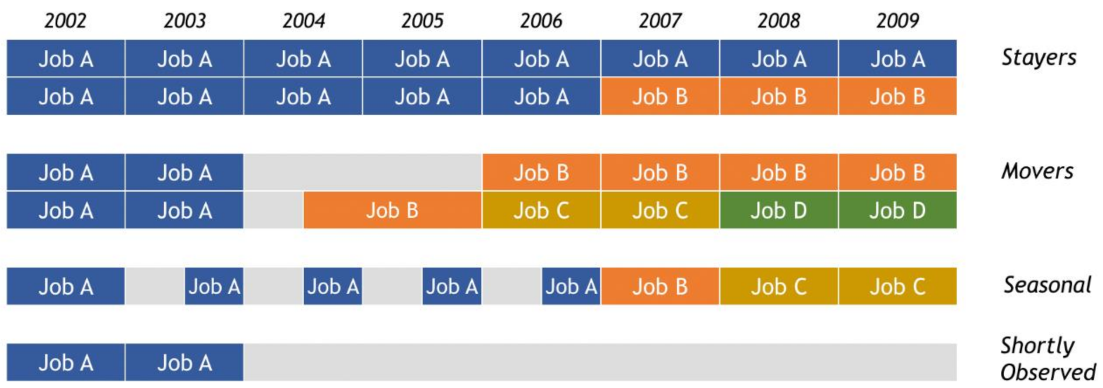


Table 6: The wage distributions of workers in each cluster

Real wage	cluster 1 (38%) "*Stayers*"				cluster 2 (33%) "*Movers*"				cluster 3 (14%) "*Seasonal*"				cluster 4 (15%) "*Shortly Observed*"			
monthly	percentile			IQR	percentile			IQR	percentile			IQR	percentile			IQR
	25th	50th	75th		25th	50th	75th		25th	50th	75th		25th	50th	75th	
age 21-25	5818	6973	9056	3238	5292	6224	7812	2520	4779	5787	6945	2166	5102	6086	7534	2432
age 26-30	6553	8842	12800	6247	5691	7151	10204	4513	5107	6250	8626	3519	5479	6632	9123	3644
age 31-35	6993	10466	16949*	9956*	5750	7653	12375	6625	5235	6657	10667	5432	5479	6849	10535	5056
age 36-40	7008	11301	17103*	10095*	5632	7489	13220	7588	5165	6712	12514	7349	5342	6696	10958	5616
hourly**	percentile			IQR	percentile			IQR	percentile			IQR	percentile			IQR
	25th	50th	75th		25th	50th	75th		25th	50th	75th		25th	50th	75th	
age 21-25	34	41	53	19	31	36	45	15	28	34	40	13	30	35	44	14
age 26-30	38	51	74	36	33	42	59	26	30	36	50	20	32	39	53	21
age 31-35	41	61	99*	58*	33	44	72	39	30	39	62	32	32	40	61	29
age 36-40	41	66	99*	59*	33	44	77	44	30	39	73	43	31	39	64	33

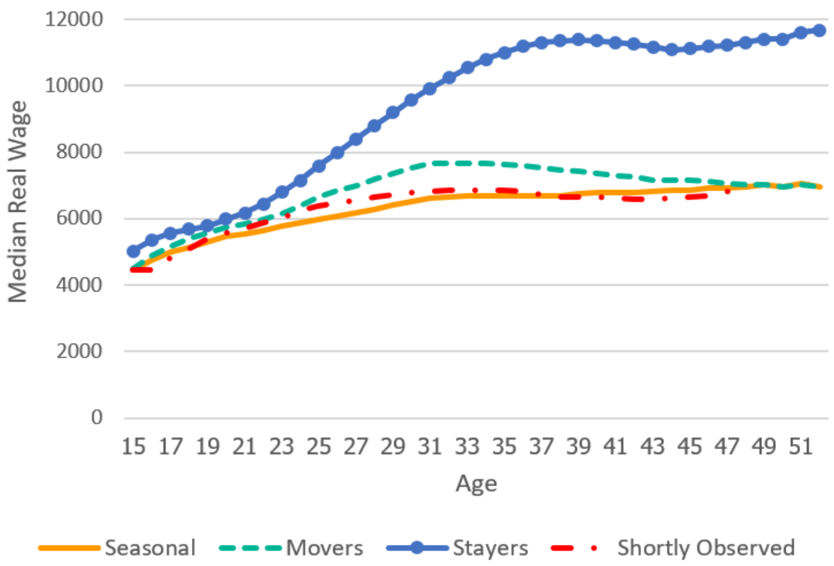
Source: Authors' calculation from Social Security records for employees aged 15-44 years in 2002

Notes *the numbers could be underestimates because wages of some observations are censored at 15000 Baht per month.

** the hourly wages are computed from the reported monthly wage in SSO assuming that workers work 40 hours a week and 4.3 weeks a month

To check whether this distribution is similar to what we obtained from LFS, the bottom panel of Table 6 reports the converted hourly wages, assuming that workers work 40 hours a week and 4.3 weeks a month. The distributions appear to be in reasonable ranges where the wage distribution of *Stayer*s likely corresponds to vocational or college graduate workers. The wage distributions of other groups likely correspond to those of secondary or primary graduates. The fact that wage-age gradient for low-skill workers is flat after age of 32 is consistent with the explanation that wages reflect productivities and that the initial wage growth is due to higher skills obtained by on-the-job training. The marginal return to experience for low-skill jobs likely diminish after 10 years of experience.

Figure 8: Median real wages for different types of workers at various ages

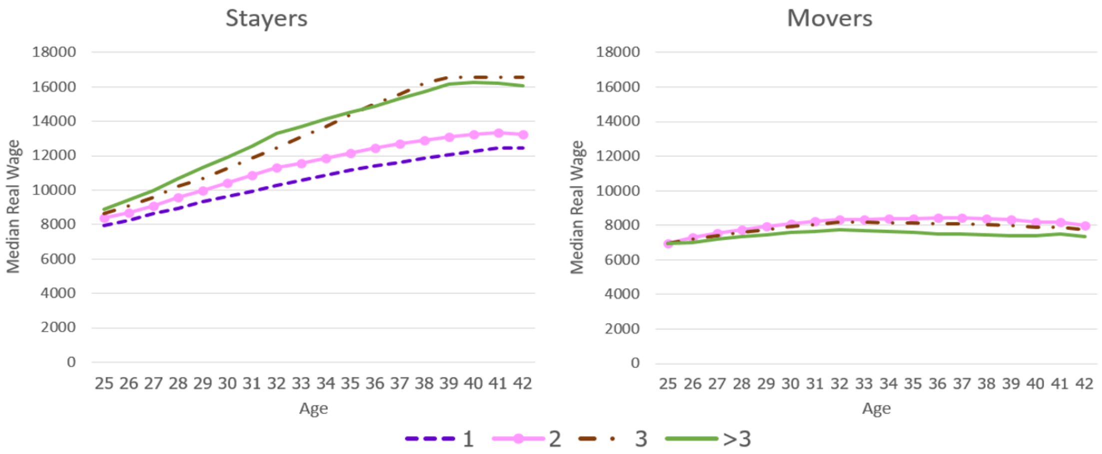


Source: Authors' calculation from Social Security records for employees aged 15-44 years in 2002

We further explore wage disparities among the two largest clusters (*Stayers* and *Movers*) who were observed for the whole 96 months in two dimensions: number of jobs held over the 8-year period and their employer sizes. While we expect large firms to pay higher wages (except the monopsony case), the relationship between number of jobs and wage is ambiguous. On the one hand, firms compete for talent and offer higher wages for skilled workers. On the other hand, non-productive workers may have more chances to be fired and lead them to have more jobs and lower wage relative to productive ones.

Figure 9 shows that some of the *Stayers* are likely to be the talented ones as who switch more jobs have a steeper wage path as they get older. In contrast, there is no evidence that switching to a new job help *Movers* reach a higher wage level. These two pictures suggest that either there exists a high wage penalty for exiting the formal labor market or *Movers* are low-skill workers whose wage would not grow much as low-skill are easy to be replaced.

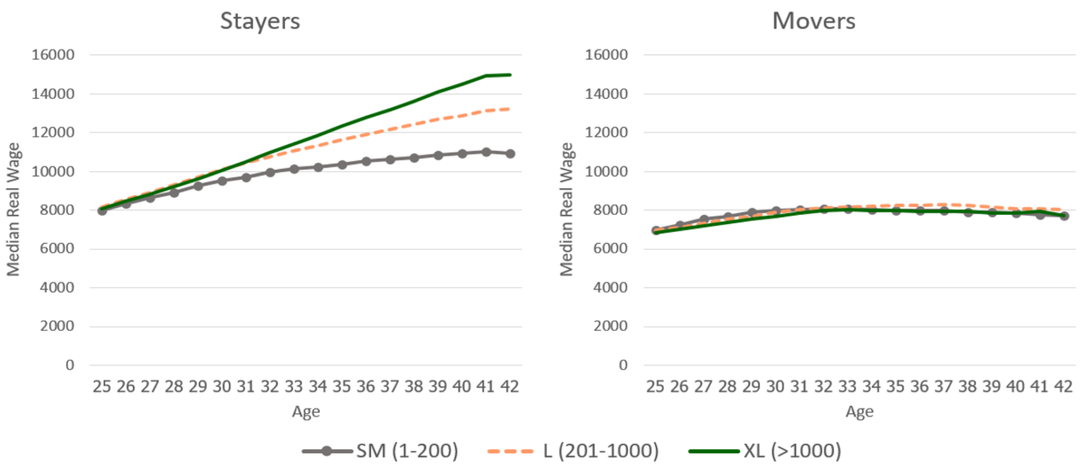
Figure 9: Median real wages of *Stayers* and *Movers* separated by the number of jobs



Source: Authors' calculation from Social Security records for employees aged 25-34 years old who were employed every month during the 96-month being studied

Regarding firm sizes, Figure 10 illustrates that, among *Stayers* of the same age, those work in a large firm earn more. However, for *Movers,* who presumably possess lower skill than *Stayers,* their wages are not correlated with firm sizes. This suggests that even though large firms benefit from their sizes, these benefits are not distributed equally among their workers. They only pay higher wages for the top skill workers where the supply is likely more scarce.

Figure 10: Median real wages of *Stayers* and *Movers* separated by employers' firm size



Source: Authors' calculation from Social Security records for employees aged 25-34 years old who were employed every month during the 96-month being studied

Overall, the SSO data paints a picture of two different groups of employees. On one hand, the *Stayers*, who stay in the formal sector for most of their work history, have their wages grow as they age. Within this group, there is also a substantial degree of earnings dispersion. On the other hand, he hybrid workers, the *Movers*, the *Seasonal* and the *Shortly Observed* – altogether accounting for 62% of employees, have similar earnings. Their wages do not grow much over time and are uncorrelated with neither job switching nor firm sizes.

6. Conclusions and discussion

Several studies suggest that Thailand's income inequality has declined since the mid-1990s. However, income inequality is a broad and complicated concept. This paper takes a closer look at the evolution of labor income and possible mechanisms underlying its dynamics. At the aggregate level, the trend of labor income inequality still mimics the patterns of the total income inequality being documented. However, our analysis based on individual-level cross-sectional and panel data reveals complicated forces behind the observed trend.

First, both datasets suggested that the force underlies the disappearing disparity in the labor income is the falls of the within- and between- inequality among low-skill workers. In other words, those in the bottom part of the distribution had become more equal. The college graduated workers, accounting only for 17% of the total workforce, have left the rest far behind. This implies that while Thailand's economic pie has been growing, albeit slowly, the majority of the country's workforce has not yet reaped the benefits.

Second, we documented that Thai workers have been moving down the occupational-skill ladder over the last thirty years. Recently college graduates are no longer concentrated in the high-skill jobs, occupying some of the middle-skill jobs. More secondary graduates have then been pushed from the middle-skill to low-skill jobs, compare to their counterparts three decades ago. The primary educated group, while they are still concentrated in low-skill job and receive somewhat higher pay, their employment probability is slightly lower. These results suggest that there exist heterogeneous skill sets among workers completing the same level of the education, and the mismatch between the labor demand and labor supply in the country.

Third, when clustering employees by their employment patterns, the age-earnings profiles also suggest the disparities between those always working in the formal sector and the hybrid-workers (those moving between the formal and the informal sectors). Those entering the market with the needed skills, not only earn more when they are young, but also have higher wage growth throughout their working lives. This larger gap potentially leads to greater inequality in consumption, saving and wealth over their lifetime.

Finally, several market forces also enable top-skill workers to earn much more. Among the *high skill*, those on the top can also switch to a new job for a higher salary, and these jobs tend to be in a large firm. On the other hand, wages of *low-skill workers*, accounting for more than half of workers in formal or semi-formal sector, remain low even when these workers' age and are uncorrelated with neither job switching nor firm sizes.

Over the past three decades, the Thai government has put a huge effort to address the country's inequality problem – both reducing unequal opportunities which generate income disparity and alleviating existing unequal income. Series of education and healthcare reforms have been implemented to allow more equal access to education and healthcare. Labor protection policies such as minimum wage and lay-off compensation also have been put in places to help the low-skill workers. The country, however, is still struggling to combat the income inequality. The number of people registered for low-income living assistance last year is more than triple the number of the country's taxpayers.

As the world is getting richer and more connected, a greater level of inequality is expected as from market economics (Piketty, 2015; Jones 2015). Firms compete for talents. If highly talented workers are in short supply, these workers can command high earnings. In contrast, those with skills easily substituted by machines or other workers likely have their wages suppressed unless they can acquire some demanded skill sets.

The promising way to reduce inequality in the long run is to enhance human capital since early childhood and throughout people's lives. Once born in poor families, chances of moving up in the income distributions are tiny in most countries (Clark, 2015). While more Thai children go to school, the school quality is vastly diverse, especially between urban and rural areas (Kilenthong, 2017; World Bank, 2018). Furthermore, although predicting the rapid changing labor demand is not easy, the country can prepare its children for a lifelong learning, allow college students to adjust their major more flexibly, and support workers to upgrade their skills throughout their working lives.

Finally, our study takes a first step to highlight the fact that in tackling the income inequality issue, understanding mechanisms behind its dynamics is crucial. Future research should take a closer look at the elites (the top 1%) and the extremely poor in the agricultural sector as these two groups are likely missing from our analysis. Furthermore, analyses of the distributive consequences of taxes, income transfers, and labor protection policies in the short-run and long-run are fruitful research directions.

References

Abowd, J., Abramowitz, J., Levenstein, M. C., McCue, K., Patki, D., Raghunathan, T., Rodgers, A.M., Shapiro, M.D. & Wasi, N. (2019). Optimal Probabilistic Record Linkage: Best Practice for Linking Employers in Survey and Administrative Data. *US Census Bureau, Center for Economic Studies*.

Autor, D. (2019). Work of the past, work of the future. Working Paper 25588. *National Bureau of Economic Research*.

Autor, D.H., Katz, L. F., & Kearney, M. S. (2006). The polarization of the US labor market. *American economic review*, *96*(2), 189-194.

Brown, C., & Medoff, J. L. (1989). The Employer Size-Wage Effect. *Journal of Political Economy, 97,* 1027–1059.

Chetty, R. & Hendren, N. (2018). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects*. *The Quarterly Journal of Economics*, *133*(3), 1107-1162.

Clark, G. (2015). *The son also rises: Surnames and the history of social mobility* (Vol. 49). Princeton University Press.

Hawley, Joshua D. (2004). Changing returns to education in times of prosperity and crisis, Thailand 1985-1998. *Economics of Education Review*, *23*(3), 273-286. Elsevier.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*(5782), 1900-1902.

Jeong, H. (2008). Assessment of Relationship between Growth and Inequality: Micro Evidence from Thailand. *Macroeconomic Dynamics*, *12*(S2), 155-197.

Jones, C. I. (2015). Pareto and Piketty: The macroeconomics of top income and wealth inequality. *Journal of Economic Perspectives*, *29*(1), 29-46.

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.

Keane, M. P., & Wasi, N. (2016). Labour supply: the roles of human capital and the extensive margin. *The Economic Journal*, *126*(592), 578-617.

Kilenthong, W. (2016). Finance and Inequality in Thailand. *Thailand and the World Economy*, *34*(3), September-December 2016.

Kilenthong, W. (2017). The situation of Education in Thailand 2015/2016: the need of competition and decentralization in the Thai educational system.

Krueger, A. (2012). The Rise and Consequences of Inequality in the United States. Speech at the Center for American Progress, Washington D.C.

Laovakul, D. (2016). Concentration of Land and Other Wealth in Thailand. In P.Pauk & C. Baker (Eds.), Unequal Thailand (pp.32-42). NUS Press.

Liao, L., & Paweenawat, S. W. (2019). “Gold Miss" or "Earthy Mom"? Evidence from Thailand. *PIER Discussion Papers* 110, Puey Ungphakorn Institute for Economic Research, revised Jul 2019.

Paweenawat, S. W. & McNown, R. (2014). The Determinants of Income Inequality in Thailand: A Synthetic Cohort Analysis. *Journal of Asian Economics, Elsevier, 31*, 10-21. DOI: 10.1016/j.asieco.2014.02.001

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

Piketty, Thomas. (2015). "About *Capital in the Twenty-First Century*." *American Economic Review*, 105 (5): 48-53.

Pootrakul, K. (2013). Quality of Economic Growth from a Dimension of Income Distribution. In Proceedings in the bank of Thailand annual symposium 2013. [in Thai]

UN-ESCAP. (2018). Inequality in Asia and the Pacific in the Era of the 2030 Agenda for Sustainable Development. https://www.unescap.org/commission/74/files /Inequality-L.pdf

Vanitcharoentham, V. (2017). Top Income Shares and Inequality: Evidences from Thailand. *Kasetsart Journal of Social Sciences, [S.l.], 40*(1), 40-46, Aug. 2017. ISSN 2452-3151. Available at: <http://kuojs.lib.ku.ac.th/index.php/kjss/ article/view/2729>.

World Bank (2016). Getting Back on Track: Reviving Growth and Securing Prosperity for All, Bangkok: World Bank Thailand. <http://www.worldbank.org/thailand>

Appendices to accompany

Labor Income Inequality in Thailand: the Roles of Education, Occupation and Employment History

Nada Wasi

Sasiwimon Warunsiri Paweenawat

Chinnawat Devahastin Na Ayudhya

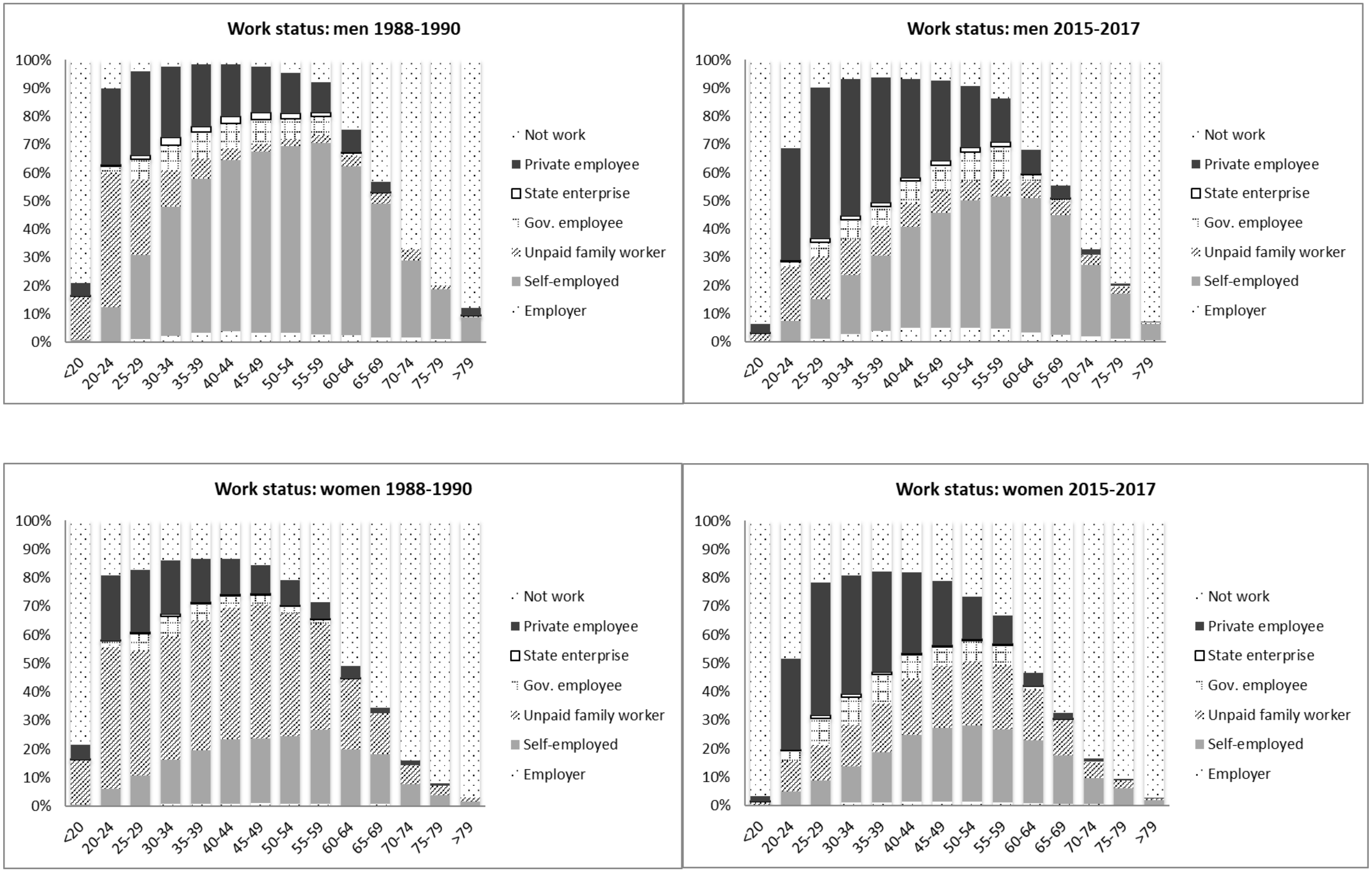
Pucktada Treeratpituk

Chommanart Nittayo

November 2019

Appendix A : Additional background information about Thailand's economy

Figure A1: Type of work for Thai men and women in 1988-1990 and 2015-2017 by age and gender



Source: Authors' calculation from LFS

Appendix B: Education and occupation codes in LFS

Because education codes changed over time, we categorize educations into five levels based on years of schooling as showed in Table B1.

Table B1: Education Level Classification used for LFS

Education level	Years of schooling
Dropout	3 or lower
Primary	4-8
Secondary	9-13
Vocational	14-15
College	16 or more

For occupation, we categorize detailed occupation codes into 13 groups, ensuring the compatibility among the three versions of codes, and that occupations within the same group still reflect similar complexity of tasks and skills required. This is similar to Autor (2019), ISCO (1988), ISCO (2008), Lathapipat (2008).[9] The Thai LFS surveys have two major changes in the classifications of occupation in 2001 and 2011. Since 2011, the classification is based on the International Standard Classification of Occupations (ISCO)-2008. Between 2001 and 2010, the occupational codes are based on the ISCO-1988. The classification is based on the Thai NSO standard prior to 2001. Thirteen categories are further clustered into 3 broad occupation groups: (1) high-skill occupations including managers, professionals, technicians and associated professionals; (2) middle-skill occupations including clerical and sale/service workers; and (3) low-skill occupations including agricultural workers, craft and manual workers, machine operators, assemblers, drivers and laborers. Table B2 provides the codes.

[9]International Standard Classification of Occupations (ISCO). (1988). ISCO-88. Retrieved from: https://www.ilo.org/.
International Standard Classification of Occupations (ISCO). (2008). ISCO-08. Retrieved from: https://www.ilo.org/.
Lathapipat, D. (2008). The changing educational distribution and its impact on the evolution of wages in Thailand, 1987-2006.

Table B2: Occupation Codes from LFS

Occupations:	Codes:						
1988-2000							
Manager, legislator, senior officials	0680/0682	1010/1019	1100/1199	4016/4019			
Science, doctors, engineer, college professors	0010/0039	0110/0199	0210/0399	0410/0419	0510/0519	0610/0619	
Business/finance related professionals	0Y10/0Y49	0Y95					
Lawyers and other social science professionals	0800/0899	0Y20/0Y39	0Y90/0Y94	0Y96/0Y99			
Artists & journalists	0937	0900/0929	0930/0936				
School teacher and associates	0620/0629	0683/0689	0710/0719				
Technician & associate professionals	0420/0429	0490/0499	0520	0530/0599	0939	9010/9090	0X10/0X99
	6200/6299	6710/6729	6930/6939	7414/7416	3110/3119	9610/9669	9711/9719
Clerical support workers	2010/2019	2100/2999	9910				
Services and sale workers	3010/3090	3210/3299	3320/3319	3320	3390/3399	4417/4418	6510/6519
	6610/6629	6810/6829	6910/6911	6920/6949	9010	9019	9091/9099
	9110/9119	9129/9129	9194/9198	9210/9219	9410/9419	9810/9819	9911/9919
	9196	9811					
Agricultural workers	4010/4015	4110/4113	4119	4210/4219	4310/4319	4410/4415	4419
Craft and related trade workers	5014	5994/5999	7010/7099	7100/7299	7320/7329	7410/7413	7419
	7420/7499	7530/7599	7600/7659	7710/7729	7790/7799	7810/7829	7990/7992
	8010/8099	8110/8149	8220/8299	8351/8359	8390/8399	8410/8499	8510/8519
	9533/8539	8540/8599	7910/7999	6110			
Plant and machine operator, assemblers, drivers	5010/5013	5019	5110/5199	5200/5219	5990/5993	6010/6029	6111/6115
	6120/6129	6300/6339	6410/6419	7310/7319	7330/7359	7418	7500/7529
	7690	7730/7739	8210/8219	8310/8349	8350/8359	8610/8699	8700/8729
	8730/8759	9511/9514	8350				
Laborer (non agriculture)	3321/3329	4114/4115	4416	5999	6119	6420/6439	6912/6913
	7993/7999	8190/8199	8760/8769	8810/8899	8900/8999	9122	9190/9193
	9195/9199	9310/9329	9510/9519	X200/X300			
2001-2010							
Manager, legislator, senior officials	1000/1999						
Science, doctors, engineer, college professors	2100/2299	2230/2239	2310/2319				

Business/finance related professionals	2411/2419	2441	3411	3434			
Lawyers and other social science professionals	2412	2420/2429	2431/2432	2442/2446			
Artists & journalists	2450/2459	3472					
School teacher and associates	2320/2359	2460	3310/3340				
Technician & associate professionals	3100/3199	3200/3299	3412/3419	3421/3429	3431/3433	3439	3441/3449
	3450	3460	3471/3475	3480			
Clerical support workers	4100/4299	2999					
Services and sale workers	5100/5469	9970	9998	9999			
Agricultural workers	6100/6299	9210/9219					
Craft and related trade workers	7113/7114	7120/7149	7200/7499				
Plant and machine operator, assemblers, drivers	7111/7112	8110/8179	8210/8299	8310/8349			
Laborer (non agriculture)	9111/9112	912/916					
	930/939						
2011-2017							
Manager, legislator, senior officials	1000/1999						
Science, doctors, engineer, college professors	2100/2169	2200/2259	2260/2263	2310/2319	2510/2529		
Business/finance related professionals	2411/2413	2631	3311/3314				
Lawyers and other social science professionals	2421/2424	2431/2432	2610/2629	2632/2635			
Artists & journalists	2640/2659						
School teacher and associates	2320/2359	2636					
Technician & associate professionals	2163	2264/2269	2433/2434	3110/3119	314/315	3210/3259	3312/3315
	3320/3339	3341	3351/3359	3410/3429	3431/3435	3511/3524	0100/0399
Clerical support workers	3341/3344	4100/4499					
Services and sale workers	3434	5100/5499	9411/9412	9970			
Agricultural workers	6100/6399	9211/9216					
Craft and related trade workers	7100/7599						
Plant and machine operator, assemblers, drivers	3121/3123	3131/3139	810/839				
Laborer (non agriculture)	9111/9129	9310/9339	9510/9520	9610/9629			

Appendix C: Clustering analysis

The k-means clustering technique works as follows: to cluster m employment profiles into C clusters, k-means clustering first randomly picks C cluster centroids and then assigns each employment profile to the cluster with the closest centroid. Once all employment profiles are assigned to clusters, each centroid is recomputed to the mean of the employment profiles within that cluster. Then each employment profile is reassigned to a cluster that it is closest to, based on the newly computed centroids. The process is repeated until all the centroids are stabilized. The Euclidean distance is used to measure the similarity between two different employment profiles. For any two profiles $X = (x_1, x_2, x_3, x_4, x_5, x_6)$ and $Y = (y_1, y_2, y_3, y_4, y_5, y_6)$, the Euclidean distance $d_E(X, Y)$ between X and Y is defined as Equation 1.

$$d_E\big(X, Y\big) = \sqrt{\sum_{k=1}^{6}(x_k - y_k)^2} \qquad (1)$$

All six characteristics in the employment profile are first standardized to Normal(0,1), in order to make the difference in each characteristic contributes equally to the distance calculation. [10]

To determine the appropriate number of clusters, we use the “elbow analysis” to find an optimal point that balances the trade-off between the within group homogeneity and the simplicity of the model. By increasing the number of clusters, the model can always increase the degree of similarity within group because the complex model has more freedom in grouping similar data points together. However, after a certain point, the improvement in within group homogeneity became marginal, resulting in an “elbow” shape plot. The typical practice is to select the “elbow” point to be the number of clusters. [11]

The within group homogeneity is typically measured by the “within-clusters distance” where a lower number indicates a higher degree of similarity. For a given assignment of m data points to C clusters, the “within-clusters distance” is shown in Equation 2.

[1] Except for the length of observation, other characteristics are also first log-transformed before the standardization.

[2] The concept is analogous to selecting the number of classes for the semi-parametric latent class model. The model selection criteria such as BIC balances the in-sample fit with the model complexity (the number of parameters).

$$\sum_{c=1}^{C} \sum_{x_i \in C} \left\| x_i - \mu_c \right\|^2 \tag{2}$$

where μ_c is the centroid of a cluster C_c and $\|x_i - \mu_c\|$ denotes the Euclidean distance between x_i and its cluster's centroid μ_c.

Figure C1 shows the plot of "within-cluster distance" against the number of clusters for the employment profiles of the cohort presenting in the paper, suggesting that the appropriate number of clusters is 4 (at the kink).

Figure C1: The Total Within-Cluster Distance by the Number of Clusters