Essays in Labor and Development Economics: (1) Gender Inequality in Thailand: Analysis of Mean Earnings Differences; (2) Gender Inequality in Thailand: Analysis of Wage Distribution Using DFL Decomposition; and (3) The Determinants and Effect of Training on U.S. Immigrant Workers

A Dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Economics

by

Charles Saharuk Mutsalklisana

August 2011

Dissertation Committee:
Dr. David Fairris, Co-Chairperson
Dr. Aman Ullah, Co-Chairperson
Dr. Todd Sorensen
The Dissertation of Charles Saharuk Mutsalklisana is approved:

___________________________________________________

___________________________________________________

___________________________________________________

Committee Co-Chairperson

Committee Co-Chairperson

University of California, Riverside
Essays in Labor and Development Economics: (1) Gender Inequality in Thailand: Analysis of Mean Earnings Differences; (2) Gender Inequality in Thailand: Analysis of Wage Distribution Using DFL Decomposition; and (3) The Determinants and Effect of Training on U.S. Immigrant Workers

by

Charles Saharuk Mutsalklisana

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, August 2011
Professor David Fairris and Professor Aman Ullah, Co-Chairpersons

This dissertation is composed of three essays that study gender inequality in Thailand and the effect of job training on immigrant workers in the United States. Essay 1 identifies the factors that account for mean earning differences in men and women, such as occupational sorting, demographic differences, human capital differences and the unexplained.

The outcome from OB decomposition indicates that the reduction over time in the mean wage gap is mostly due to an increase in female human capital accumulation and the improvement in female occupation outlook relative to men. One of the reasons that a sizable mean wage gap still exists in Thailand, despite the increase in education level of women, is because the increase in female human capital accumulation over the past decade is overshadowed by an increase in the return to observables characteristics of men.
Essay 2 finds that the result from DFL decomposition is consistent with the OB decomposition. We find that if Thai women possess similar observable characteristics as men, the gender wage inequality will be greater for the majority of the wage distribution, particularly, for middle to high income workers.

Chapter 3 studies the effects of job training on immigrant workers found in the U.S., using the data from the Survey of Income and Program Participation (SIPP). This essay is one of the first empirical papers that look into the effects of job training on immigrants in the United States, using Random Effect (RE) model, a Propensity Score Matching, a Quantile regression (QREQ) model and a semi-parametric reweighting method. The Random Effects model indicates that the conditional effect of job training on the average earnings of immigrants (at 3.9 percent) is less than that of natives (at 7.6 percent).

From our distribution study, we found that job training had a positive effect on the wages of immigrant workers over most of the wage distribution. The results from QREQ also show that immigrants enjoy the largest conditional job training premium at the lower and middle part of worker earning quantiles. Examining counterfactual study, DFL reweighting technique shows that similar to Abadie, Angrist and Imbens (AAI) (2002) we found that the largest proportional impact of job training is at the upper part of the wage distribution for both natives and immigrants. Nevertheless, we still found that job training increases the wage premium of lower and middle income workers.
# Table of Contents

Chapter 1 Gender Inequality in Thailand: Analysis of Mean Earnings Differences .................................................................1

1.1 Introduction ..................................................................................................................1

1.2 Literature Review: Gender Inequality in Thailand ..........................................................3

1.3 Data Source ..................................................................................................................5

1.3.1 Data .......................................................................................................................5

1.3.2 Descriptive Statistics ............................................................................................7

1.4 Methodology ................................................................................................................7

1.4.1 Ordinary Least Square (OLS) ...............................................................................7

1.4.2 Oaxaca-Blinder Decomposition (OB) ...................................................................9

1.5 Analysis Mean Earnings Differences ...........................................................................10

1.5.1 Ordinary Least Square (OLS) ...............................................................................10

1.5.2 Oaxaca-Blinder Decomposition (OB) ...................................................................12

1.6 Conclusion ..................................................................................................................16

Chapter 2 Gender Inequality in Thailand: Analysis of Wage Distribution Using DFL Decomposition .........................................................................................................20

2.1 Introduction ................................................................................................................20

2.2 Methodology .............................................................................................................22

2.2.1 DiNardo, Fortin and Lemieux Decomposition (DFL) .............................................22

2.3 Analysis Distribution .................................................................................................26

2.3.1 DiNardo, Fortin and Lemieux Decomposition (DFL) .............................................26

2.4 Conclusion ..................................................................................................................30

2.4.1 Policy Implication .................................................................................................32

Bibliography ..................................................................................................................33

Appendix ..........................................................................................................................37

Chapter 3 The Determinants and Effects of Job Training on U.S. Immigrant Workers ............................................................................................................................70

3.1 Introduction ................................................................................................................70

3.1.1 Issues on Training .................................................................................................73

3.2 Literature Review .....................................................................................................75

3.2.1 Study of Average Treatment Effect on Treated ......................................................77

3.2.2 Study of the Distribution Effect ............................................................................78

3.3 Empirical Model .......................................................................................................79

3.3.1 A Model of Quantile Regression ..........................................................................79

3.3.2 Counterfactual Study .............................................................................................80

3.3.3 Estimating with Propensity Score Matching .........................................................81

3.3.4 Sensitivity Analysis for Average treatment Effects on the Treated .................85
List of Tables

Table 1.1. Government Interventions Related to Gender Inequality ......................................35
Table 1.2. Five categorical regions .......................................................................................35
Table 1.3. Education categories ..........................................................................................35
Table 1.4. Kernel density .....................................................................................................36
Table 1.5. Dependent and Independent Variables and Definitions .....................................37
Table 1.6. Descriptive Statistics of Dependent and Independent Variable in 1997 and 2006 ..........................................................................................................................................................38
Table 1.7. Descriptive Statistics of Men and Women in 1997 .............................................39
Table 1.8. Descriptive Statistics of Men and Women in 2006 .............................................40
Table 1.9. The results of unconditional OLS pooled model using 2006 and 1997 data ..........41
Table 1.10. The results of conditional OLS pooled model using 2006 and 1997 data ........41
Table 1.11. The results of OLS pooled model with different specifications ......................42
Table 1.12. The results of OLS models 1997 and 2006 ......................................................42
Table 1.13. The results of OLS male and female models using 1997 and 2006 .................43
Table 1.14. Oaxaca decomposition 1997 and 2006 ............................................................44
Table 1.15. Oaxaca decomposition 1997 and 2006 (Occupation) .......................................45
Table 3.1. Mean Values of Natives .....................................................................................105
Table 3.2. Mean Values of Immigrants ...............................................................................106
Table 3.3. Ordinary Least Squares Regression Results ......................................................107
Table 3.4. OLS Model. Estimate Effect of Training on Earnings. Dependent Variable: Log of Hourly Earnings .................................................................107
Table 3.5. OLS Model. Estimate Effect of Training on Earnings. Dependent Variable: Log of Hourly Earnings .................................................................108
Table 3.6. OLS Models. Estimate Effect of Training on Earnings. Dependent Variable: Log of Hourly Earnings .................................................................109
Table 3.7. Sensitivity Analysis. OLS Estimate Effect of Training on Earnings. Dependent Variable: Log of Hourly Earning .................................................................110
Table 3.8. OLS/Matching Estimate Effect of Training on Earnings ......................................111
List of Figures

Figure 1.1. 1997 and 2006 Actual Gender Log Wage Gap by Percentile ..................46

Figure 1.2. 2006 Counterfactual Wage Gap DiNardo, Fortin, Lemieux Decomposition
Results .......................................................................................................................47

Figure 1.3. 1997 Counterfactual Wage Gap DiNardo, Fortin, Lemieux Decomposition
Results .......................................................................................................................48

Figure 1.4. Probability Distribution of Log Hourly Earnings. Panel (A) 1997 Actual
Wage Density for Thai Female and Male Workers ..................................................49

Figure 1.5. Panel (A-G). Counterfactual Probability Distribution of Log Hourly Earnings.  
1997 Female Counterfactual Wage Density ............................................................50

Figure 1.6. Panel (A-G). Counterfactual Probability Distribution of Log Hourly Earnings.  
2006 Female Counterfactual Wage Density ............................................................53

Figure 3.1 Unconditional Effect of Job Training on Earnings for Natives and Immigrants.  
Dependent Variable: Log of Hourly Earnings .......................................................112

Figure 3.2 Unconditional Effect of Job Training on Earnings for Natives and Immigrants.  
Dependent Variable: Log of Hourly Earnings .......................................................113

Figure 3.3 Conditional Effect of Job Training on Earnings for Natives and Immigrants (Quantiles Regression).  
Dependent Variable: Log of Hourly Earnings .......................................................114

Figure 3.4 Conditional Effect of Job Training on Earnings for Natives and Immigrants (Quantiles Regression).  
Dependent Variable: Log of Hourly Earnings .......................................................115

Figure 3.5 Unconditional and Counterfactual Effect of Job Training on Earnings for  
Natives and Immigrants (DFL Model). Dependent Variable: Log of Hourly Earnings ..116

Figure 3.6 Unconditional and Counterfactual Effect of Job Training on Earnings for  
Natives and Immigrants (DFL Model). Dependent Variable: Log of Hourly Earnings ..117
Chapter 1

Gender Inequality in Thailand:

Analysis of Mean Earnings Differences

1.1 Introduction

The objective of this paper is to empirically examine gender inequality in Thailand. The study of gender inequality is important because gender inequality has an extensive effect on welfare of Thai people. Gender inequality holds back the growth of individuals, the development of the country and the evolution of society, to the disadvantage of both women and men (United Nations, 2000). In recent decades, Thailand experienced a remarkable decline in the gender wage gaps. This paper will seek to explain the root causes of the decrease of the gender wage gaps by looking at the income disparity between men and women for year 1997 and 2006 individually and by exploring how Thailand’s gender wage gaps changed from 1997 to 2006.

There are number of reasons why we should address the topic of gender wage differences in the developing countries. The wage inequality might discourage some female workers from entering into the labor market or reduce the incentive for them to remain in the labor force\(^1\). Since gender inequality decreases female earning, it worsens

\(^1\) Suppose certain groups of women decide not to enter in to the workforce, the absence of this specific group of women may cause the bias in the result of wage gap computation. We acknowledge that selection in to the labor force is a problem in the study of discrimination. The female choice in labor force participation is essential to the structure issue that affects the gender wage inequality. However, labor force participant problem is beyond the scope of this paper, so it will be treated as exogenous.
the situation for already poverty stricken women. Also, the increase in the number of poor mothers directly contributes to children suffer, since there are a large number of single mothers in Thailand\(^2\). Therefore, persistence of gender inequality lead to more poverty and exacerbation of other related social issues such as acting as an impediment to educational advancement of poor women and their children, female and child prostitution, and malnutrition of children.

In chapter 1, we use Oaxaca Blinder (OB) decomposition to quantify how much of the mean gender wage gap is attributed to observable characteristics differences and the “price” differences associated with these characteristics for year 1997 and 2006 individually. We explore the importance of each Thai attribute such as educational attainment, age group, occupation, establishment size, urban area and region.

The paper is structured as follows. Section 2 discusses the literature on gender difference, focusing on the past econometric studies and issues relating to Thailand. Section 3 describes the data source. Section 4 discusses the methodology that will be applied in the study of gender inequality. Section 5 summarizes OB decomposition findings on gender inequality, examining the disparity between men and women with respect to wages within a single period (for 1997 and 2006 individually). Also, we comment on the transformation of the overall wage structures from year 1997 to 2006. Lastly, we comment on policy implications.

\(^2\)Approximately 24 percent of the Thai households are head by single women, where over half of the women households heads are widowed, NSO (2006).
1.2 Literature Review: Gender Inequality in Thailand

This section focuses on the gender inequality studies that have been conducted on Thailand. There are very few literatures on Thai wage inequality and only a handful on gender wage gap. Moreover, most of the existing studies are often the offshoot of other studies such as return on education, the Asian Crisis and program evaluation. Nearly all of past studies of Thai gender inequality either utilized the descriptive statistic tables and graphs. Recent studies employed the parametric regression methodologies such as Ordinary Least Square (OLS) and mean decomposition. As a result, this dissertation will improve upon the existing studies by being the first paper to implement Antecol, Jong, and Steinberger (2007) wage gap density methodology of DiNardo, Fortin and Lemieux (DFL) semi-parametric decomposition to explain the Thai distribution gender wage gap.

The majority of past studies were about return on education such as Schultz (1993) who suggested that for the year 1980 and 1981, average Thai women have a higher return on education at 20.1%, while the return on education for average men is approximately 11.3%. There were several studies on the Asian Crisis such as Deolalikar et.al (2000) which briefly addressed that the aggregate male wage earnings decline approximately four times as much as female during the Crisis of 1997, causing the wage differential to reduce from about 18% to 14% during the crisis. Our studies moved beyond the initial analysis by showing that the Thai gender wage gap continues to reduce beyond Asian Crisis and giving explanation of why Thai gender wage gap continued to decline.
Two key papers that used Thai mean decomposition were Jeraputtiruk (2004) and Blunch (2004). Jeraputtiruk (2004), a paper on program evaluation, used two-step Heckman method and OB decomposition to examine the effect of the 1998 Labor Protection Act on wage earning of covered and uncovered sectors. Using 1996 and 1999 data, he suggested that the gender gap in earnings, which was approximately 12.27 percent before the Act, reduced to 8.32 percent in the post-Act period.

Blunch (2004) studied the return on Thailand education using OLS, OB decomposition, Reimers, Oxaca-Blinder and Cotton decomposition (1988), and Neumark (1988) methods. In addition, the paper had a section on gender wage inequality where he concluded that the return on education after Asian Crisis reached its highest point in 1999, return on education in terms of wage favored women, and mean gender wage gap and mean wage gap between municipal and non-municipal have been decreasing overtime. In chapter 1, our paper extends these mean decomposition studies by using more recent data of year 1997 and 2006, and we explore the wage differential for the entire wage distribution. Instead of using age variable as a proxy for experience similar to Jeraputtiruk (2004), Blunch (2004) and Nakavachara (2007), we calculate and use potential experience variable in the analysis.

Lastly, a wage distribution study that was written concurrently with our paper was a working paper by Nakavachara (2010). Nakavachara (2010) uses Juhn, Murphy and Pierce (JMP) and DFL decomposition method. The study shows the result of counterfactual income distribution. It found that the changes in unexplained and the improvement of female education attainment were the most important factors in
explaining the recent decline in mean wage gap. Similar to our results, while women’s gain in education helps the reduction in mean wage gap, the unexplained portion has the opposite effect.

In Chapter 2, our study moved beyond Nakavachara (2007) in quantify and identify the factors that account for differences in the male and female wage distributions in Thailand such as occupational sorting, demographic differences, human capital differences and unexplained. Beyond the typical DFL wage density study implemented in the above paper, our study constructs the counterfactual distribution gender wage gap using reweighing method of DFL. We decompose the counterfactual distribution wage gap to isolate the differences due to each of the observable characteristics to illustrate the relative importance of particular observables.

1.3. Data source

1.3.1 Data

The analysis of this paper uses third quarter 1997 and third quarter 2006 Labor Force Survey (LFS), collected by the National Statistical Office (NSO) of Thailand, Office of The Prime Minister. LFS is a cross sectional data of the household and individuals living in Thailand. LFS is ideal because it has a sufficiently large sample size for this type of analysis, and it contains essential variables needed in wage discrimination analysis such as demographics (region, metropolis status), human capital endowments (schooling, age), establishment (size of establishment) and labor wages (number of hours
worked per week, monthly income). It is the most inclusive nationwide data that have been used in several studies of gender inequality in Thailand. The reason that we choose 1997 and 2006 rather than other years is because ten years time span is a good time duration to detect significant changes in the gender wage inequality that is often employed in analysis for many countries.

The NSO began conducting LFS in 1963. It increased the data collection frequency to two times beginning 1971 and to three times beginning 1984. Starting from 1998, the NSO interviews approximately 60,000 households, which amount to 165,000 individuals aged 13 years and older. The present LFS are conducted four times yearly, specifically, in February, May, August and November, interviewing 220,000 plus individuals aged 15 years and older from 75 provinces and five regions of Thailand. Each province is subdivided according to municipal areas, sanitary districts, and non-municipal areas.

In this paper, the sample includes workers who are ages of 16 to 65 and are employed for wages and salaries. Hence, we exclude self-employed individuals, since the earnings pattern of these individuals may be erratic, varying drastically with respect to each season and throughout the year. Furthermore, we exclude unpaid household workers, seasonally inactive workers, workers who had not been working for the past 30

---

3 In this paper, we define discrimination as differences in wages associated with observably or “prices” paid to equivalent inputs.
4 The data information is from the Thai Workers and the Crisis Report.
5 Further description of the Thai Labor Force Survey data can be found in Juntavich (2000) paper.
6 We exclude employed workers with estimated hourly wage less than 2 baht/hour to eliminate potentially miscoded data. LFS classifies individuals 15 years old or older, who in a week, worked at least one hour for compensation, worked for business or farms in support of household with or without pay, and did not work due to illness, injury, vacation, holiday, strike, lockout, and off-season as employed.
day and workers who in a week work less than one hour. Also, we exclude individuals who were unable to be placed in categories or did not report their information such as occupation, industry and education. To isolate the retired workers and persons who are still in schools, our sample excludes workers that are students.

1.3.2 Descriptive Statistics

[Table 5] displays the variables and definition of variables used in our wage equation. The dependent variable used in our analysis is natural logarithm of the real hourly wage rate (baht)\(^7\). [Table 6] to [Table 8] display the descriptive statistics for dependent and independent variables. [Table 6] shows the statistics for 1997 and 2006, while [Table 7] and [Table 8] show the statistics for men and women.

For both years, average age of female workers is relatively younger than average age of male workers. Also, education levels of female workers are higher than that of male workers, and the advantage in female education increases from 1997 to 2006. Women have approximately 4 years less potential experience relative to men in 1997 and approximately 2 years less potential experience in 2006\(^8\). The differences in choices of hours work and choices of firm size are relatively small with men working relatively longer hours, and more women are now working in larger firms.

\(^7\) To eliminate the combining choice of hours work decision problem, estimated hourly wage is calculated using monthly wages, bonus and other incomes divided by [usual weekly hours times weeks worked in past month]. Other income includes income that provided by the employee in form food, uniforms, tips, room and board, and transportation etc. Real wage is calculated using inflation rates from Bank of Thailand.

\(^8\) Estimated potential experience is calculated as age minus education minus 6. We note that it is possible to overestimate the discrimination (underestimate wage gap) due to women having less experience.
1.4 Methodology

1.4.1 Ordinary Least Square

This section describes the tools that will be used to analyze gender inequality. First, using basic OLS, without controlling for male and female differences, we can obtain the raw mean wage gap using one equation for both men and women as following.9

\[ y_{it} = \delta_t s_{it} + \epsilon_{it}, \]

where \( y_{it} \) refers to the natural logarithm of the real hourly wage rate (BAHT), \( s_{it} \) is the binary variable taking value 0 for men and 1 for women. Subscript \( i \) and \( t \) represents the observation of each individual and for each time period. Hence, \( \delta_t \) is the raw mean wage gap between men and women for each time period, not accounted for individuals’ characteristics differences.

From the 1974 Mincer’s human capital earnings function, we know that wages depend on individual characteristics such as accumulated education and experience.10 In addition to human capital, we control for demographic characteristics such as living in different regions and living in urban rather than rural areas. Furthermore, we control for other job related characteristics such as occupation and size of the establishment. Consequently, the basic model can represented as follows:

\[ y_{it} = \beta x_{it} + \delta_t s_{it} + \epsilon_{it}, \]

---

9 Using Chow test, we reject pooling 1997 and 2006 data into one regression model, and we also reject pooling men and women into one regression model. Detail discussion of the results is in the mean earnings differences analysis section.

10 We use consumer price index (CPI) from Bank of Thailand (BOT) with 2002 as base year.

11 In our study, we use potential experience.
where \( x_{it} \) is the vector of human capital, demographic and job related variables, and \( \delta_t \) is now the wage gap between men and women for each time period, controlling for the above characteristics.

### 1.4.2 Oaxaca-Blinder Decomposition

OB decomposition is a counterfactual study, quantifying the average wage of female workers that would have prevailed suppose women are paid with their wage function (face the same wage structure) and women have the observable characteristics of men. OB decomposition method decomposes mean wage gap into the explained and unexplained parts. For our study, we decompose mean wage gap of separated male and female equations selecting one period \( t \) at a time. Consider the separate structure male and female equations:

\[
\begin{align*}
  y^s_t &= \beta_t^s x^s_t + \epsilon^s_t, \\
  \bar{y}_t^m &= \beta_t^m x^m_t \quad \text{and} \quad \bar{y}_t^f = \beta_t^f x^f_t, \\
\end{align*}
\]

Oaxaca-Blinder (1973) decomposition can be written by follow equation:

\[
\bar{y}_t^m - \bar{y}_t^f = \beta_t^m (\bar{x}_t^m - \bar{x}_t^f) + (\beta_t^m - \beta_t^f) \bar{x}_t^f  \tag{4}
\]

where bar indicates the average values.

---

12 We can rearrange the decomposition equation using women as the point of reference; however, the empirical result would deviate only slightly.

\[
\bar{y}_t^m - \bar{y}_t^f = \beta_t^f (\bar{x}_t^m - \bar{x}_t^f) + (\beta_t^m - \beta_t^f) \bar{x}_t^m,
\]
The explained gap is the differences in wages attributable to differences in observed characteristics, such as human capital and job characteristics. It is captured by the term $eta_m (x^m - x^f)$. The unexplained differences between the two groups are captured by the term $(\beta_m - \beta^f)x^f$. The unexplained gap measures the disparity in the return on observed characteristics of the two groups$^{13}$.

1.5 Analysis Mean Earnings Differences

1.5.1 Ordinary Least Square

In this section, we discuss the main findings of the empirical analysis. [Table 9] shows the results of estimates of raw mean wage gap using 1997 and 2006 pooled model that we consider in equation (1). [Table 9] displays the expected positive wage gap for pooled data. [Table 10] shows the results of estimates controlling for characteristics differences of pooled model we consider in equation (2) and the results of the Chow statistic test.

We utilize the Chow statistic test to evaluate whether it is suitable to use pooled model with 1997 and 2006 data. Chow statistic test indicates that we reject the null hypothesis, indicating structural differences for 1997 and 2006 wage equation. Hence, we reject pooling 1997 and 2006 data into one regression model and confirm that the sample split is valid.

[Table 10] reveals a striking result that the magnitude of Thai mean wage gap becomes higher when we control for observable characteristics differences. This is

---

13 The unexplained gap is often referred as gender discrimination. However, we note that a portion of the log hourly wage gap may be incorrectly attributed to discrimination if unobserved variables that are correlated with being male or female are excluded from the analysis.
contrary to the mean wage gap of the US and most countries, since we often observe narrowing of the mean wage gap when controlling for explanatory variables. We further explore the result in more detail by running a number of specifications which each add several explanatory variables. When we control for educational attainment, region, occupation and firm size, we find an increase in the magnitude of the male coefficient, indicating that the mean gender wage gap conditional upon these variables become larger (See [Table 11]).

It is noteworthy that the wage differences between men and women increase most when we control for education. Typically, in most countries, men have better endowment than women do in observable characteristics that are positively rewarded in the labor market, so one would expect that adding these control variables would decrease the measure of the gender wage gap. The unexpected result here is an artifact of the unique composition of the characteristics of women in Thailand. Particularly, women in Thailand have higher average rates of college education than men.

Furthermore, we evaluate whether it is suitable to force equivalent returns on men and women. Since human capital and job related characteristics do not affect men and women in the same way, we may observe the differences between men and women using separate regression equations. [Table 12] and [Table 13] show the results that we consider in equation (3). The tables display the estimate of the earning equations for men and women separately and the results of Chow statistic test. The result from Chow statistic test confirms the structural break in favor of splitting the model into two equations. Hence, the sample split for men and women is valid.
[Table 12] and [Table 13] show that the conditional mean gender wage gap remain relatively steady even though the unconditional mean gender wage gap has decreased. The total unconditional mean wage gap reduced from 0.123 log point or 5.6 baht per hour in 1997 to 0.072 log point or 2.5 baht per hour in 2006 (See Table 14).

According to the 1997 results in [Table 4], with respect to the returns to education, we observe that men have higher returns to education than women do for each educational category even though the percentage of women completed college and post college categories are larger than men. In the case of human capital, we observe that men have higher returns to work experience than women. [Table 13] shows similar results of estimates when we control for occupation and firm size.

For 2006, we observe somewhat different results. With respect to the returns to education, we observe that women have similar or higher returns to education than men do for secondary education categories and higher, while the shares of women in completed college and post college categories continue to be larger than men and the percentage of men in rest of the categories are larger than women\textsuperscript{14}. In the case of human capital, we observe that women have higher returns to work experience than men. Considering that men ought to have higher returns to working experience than women, this result is rather unexpected. Next, we further explore the problem using the decomposition methods.

\textsuperscript{14} Testing education attainment as a single continuous variable, women still have a higher return to education than men.
1.5.2 Oaxaca-Blinder Decomposition

By using OB decomposition, we can find the aggregate effect of different endowments and different returns on these endowments of gender wage gap between men and women. Before discussing OB result tables, let's consider what we expected to find. In our pooled regression, we found that the coefficient on the male dummy variable became more positive after we conditioned on a number of factors (See Table 11). For example, suppose that women earn 5% less than men. Conditional upon their observables, they earn 15% less than men. This would indicate that the explained portion of the gap is actually negative 10%; i.e. conditioning on observables reveals a bigger gap than not conditioning on them. Hence the result may first appear somewhat counter intuitive.

The result of the OB decomposition is displayed in [Table 14] and [Table 15]. For [Table 14], columns 1 to 4 report the explained and unexplained portions of the wage decomposition for 1997 and 2006, respectively. The explained portion represents the difference in the wage gap that can be attributed to differences in the mean characteristics of men and women. The unexplained portion is attributed to a combination of differences in returns to the observable characteristics and error terms. The results of the OB decomposition in [Table 6] are evaluated with respect to male equation. Rows 1 and 2 report the mean log hourly wage of men and women respectively, while row 3 displays the unadjusted mean wage gap. A positive mean wage gap value indicates that men enjoyed wage advantage relative to women.

In the bottom half of the table, row 10 displays the portion of the mean wage gap decomposed into the explained and unexplained portions, while rows 4 to 9 further
decompose the share of the explained part of the wage gap that can be attributed to different observable characteristics.\textsuperscript{15} A positive explained portion indicates that men enjoyed an advantage in the labor market that can be explained in terms of their observable characteristics relative to women; hence, observable covariates partially explain the gender wage gap. The negative value indicates that women enjoyed wage advantage respect to observable attributes relative to men.

The same logic holds for the unexplained portions, except that they are related to differences in returns to characteristics rather than differences in characteristics. Hence the explained difference is generally not thought of as evidence of discrimination (at least in the labor market) while the unexplained portion is. Whereas positive unexplained value indicates that men enjoyed advantage in terms of price or return on observable characteristics relative to women.

[Table 6] shows that on the whole, conditioning the gender wage gap on observable characteristics does not help explain the wage disparity: rather it reveals that the unexplained gender wage gap is in fact larger than the raw unconditional gender wage gap. For both years, the total value of explained portion is negative. Furthermore, the total value of unexplained portion is positive and greater than the unconditional gender wage gap. Thus, negative value of explained portion indicates that in general, Thai women have observable characteristics advantage over men, particularly in terms of education in 1997. The advantage increases in 2006. Experience is the only observable

\textsuperscript{15} Similar to the explained portion, the unexplained portion can be further decomposed into the share attributed to different characteristics. Though, the result may vary slightly according to decomposition order if the price vectors that weight the differences in coefficient vary between men and women. See Oaxaca and Ransom (1999). Yet, we find the result informative.
attribute that help explain the wage advantage benefited men relative to women, explaining 212 and 114 percent of the wage gap for 1997 and 2006 respectively. Yet, for both years, the total unexplained portion accounts for more than 100 percent of the wage differences.

The explained portion from the education covariate is negative for both 1997 and 2006. Hence, instead of helping to explain the wage disparity, removing the education differences will not lead to a smaller mean wage gap. This unexpected result is likely due to the large number of employed Thai women who have higher education than men as indicated in the mean statistic table. Suppose these Thai women have identical characteristics as men, the wage gap will actually be even larger than when unconditioned raw wage gap. Specifically, since the average woman possesses higher education relative to the average man, average woman would receive even lesser pay suppose they have the same level of education as men.

Looking at the unexplained portion, we find that the return to experience improves for women, changing from positive to negative return. In 2006, as women gain more experience, they are also compensated more relative to men. Secondly, the return to education improves for women, reducing from 115 percents to 18 percents. Yet, men still have higher return to education than women in 2006. Hence, even though women are more educated, they are still receiving lesser pay than men.

[Table 7] displays OB results including occupation and firm size covariates. We observe that women have endowment advantage in terms of occupation relative to men, -62 percent and -10 percent of the wage gap for 1997 and 2006 respectively. Considering
that the total unexplained is large favoring men; thus, one possible explanation is that it is plausible that some of the Thai women have overcompensated for lowering pay, by sorting themselves into higher paying occupations such as government jobs. Furthermore, although it is difficult to quantify occupation discrimination, another possible explanation is that there might exist a problem of employment barrier. Thai women refrain from upward mobility in the labor market due to social barrier such as glass ceiling. Thai women involuntarily opt to have low paying job because discrimination impedes their ability to enter the labor market. In other words, for Thai women to attain the same job and same pay as the men, they must be over qualified for the job and willing to settle in terms of a pay cut and limited advancement.

1.6 Conclusion

To summarize, the outcome from OB decomposition indicates that the unexplainable accounts for most of wage differential favored men. The advantage in the return to observable favoring the men is the main cause for the wage differences between men and women in Thailand. This is true particularly for the return to experience, return to occupation and return to education in 1997 and return to occupation in 2006. The reduction over time in the mean wage gap is mostly due to an increase in female human capital accumulation (Education and Experience) and the improvement in female occupation outlook relative to men.

Also, the reduction of wage gap is due to the decrease in return to experience, return to occupation and return to education favoring men. One reason that a sizable mean wage gap still exists in Thailand (despite the increase in human capital and
occupation outlook) is because the increase in endowment favoring women over the past decade are negated or overshadowed by an increase in the return to observables characteristics in men. The return to occupation and return to experience favoring men are the two main reasons for the sustained wage differences in the past decade. Lastly, human capital accumulation differences, demographic differences and occupation differences appear to have minimal affect on explaining men’s wage advantage relative to women.

Hence, we see that the increase in human capital endowment favoring women is overshadowed by the return to observable characteristics favoring men (a finding similar to “Swim Upstream” of Blau and Khan (1997)). We observe that an average Thai woman have higher human capital endowment (such as education) than an average Thai man. Since the return to observable characteristics still favored men, we find that women who have the same or higher endowment of observable characteristics than that of men still receive lower wages.

There are several possible explanations on what appears to be reverse discrimination on endowment. First, the wage differential can be attributed in that there are a higher percentage of Thai women who are highly endowed then those of the past while the demand for these qualified position remain constant. Hence, more highly endowed women are competing for the same jobs with other qualified women, particularly in high female concentrated occupations. Perhaps the economy needs time to adjust for the additional supply of these highly endowed women. Lastly, the Thai government has been actively seeking to promote growth and welfare through the National Education Act, the
Labor Protection Act and the National Economic Development Plan. Women benefited greatly from these interventions, resulting in the observed falling of gender wage gap in the past decade. Yet, in terms of the workplace, the Thai government left decisions in the hands of the private sector hence, the wage differential is still active and strong.

Despite OB decomposition being a very useful tool in general wage decomposition analysis, it has several disadvantages. For the most precise result, OB method requires wage density to conform to normality distribution condition. According to [figure 1], the wage distribution deviates considerably from the normal distribution with the negative skew. DiNardo and Tobias (2001) and Barsky, Bound, Charles and Lupton (2001) assert that when the distribution is not normal, applying parametric based type approach (profoundly uses in gender inequality studies) may not yield the best results and analysis can be misleading.

Secondly, OB method only summarizes the wage difference at the mean. Thus, this disregards the possibilities that wage differences may vary throughout the distribution. Unlike OLS or OB decomposition which compares groups at the mean aggregate, DFL (semi-parametric) describes the entire wage distribution.

[Figure 1] shows the wage advantage for men relative to women is lowest near third quantile and considerably increased for the other locations in the distribution. As a result, wage differential is not at all uniform throughout the entire wage distribution. Hence, it is interesting and beneficial for welfare study to observe the wage differential at different points on the distribution, since it is important for the policymakers to be able to target and understand the needs of different group of workers in the wage distribution.
We will thoroughly discuss the distribution wage gap in DFL result section. Thus, we now shift our focus from OB decomposition to the alternative method of Dinardo, Fortin, and Lemieux (1996) decomposition. In the next chapter, we discuss the method of DFL decomposition.
Chapter 2

Gender Inequality in Thailand:

Analysis of Wage Distribution Using DFL Decomposition

2.1 Introduction

The central questions of this chapter are the following: Which groups of Thai women in the wage distribution are benefiting from the recent decline in the raw wage gap, to what extent are women better off, what are the main attributes to help explain the distribution wage gap and to what degree do these characteristics help explain the distribution wage gap?

In this chapter, we investigate the effect of recent decline in gender wage gap on the wage distribution. Distributional analysis shows the effect of wage decline on the wage distribution that is not captured by the average impact, and the analysis has important implication on welfare of different income groups. We analyze the explanatory power of occupation sorting, demographic differences, human capital accumulation differences and unexplained portion otherwise known as discrimination as
defined in terms of the observable skills, on the change of the distribution of the gender wage gap in Thailand.\(^\text{16}\)

Kernel density and the DiNardo, Fortin and Lemieux (DFL) decomposition are used in the counterfactual analysis by enabling us to illustrate the actual female density, the actual male density and the actual wage gap distribution, as well as, the female counterfactual density and the female counterfactual wage gap distribution in graphical representation. The DFL method assigns weight to each point of the earnings distribution, while kernel density depicts and smoothes the graphical density.

The counterfactual distribution wage gap controlling for all observable characteristics is the resulting wage gap after removing all characteristics differences between women and men, assuming conditional distribution of female wage does not depend on the distribution of observable covariates. For both the counterfactual density and the counterfactual wage gap distribution, we sequentially decompose the wage distribution in the order of occupation, education, potential experience, metropolitan area region and unexplained portion respectively. The result of the reverse order decomposition is relatively robust and available upon request.

We find that gender wage gap decreases for the majority of the wage distribution from 1997 to 2006, yet the differences were not shared equally for all Thai women. We find that through education Thai women were able to lower the wage gap in middle and high earning jobs, but only because they were, on average, more educated than their male

\(^\text{16}\) For the rest of the paper, we define discrimination as the differences in pay for observably similar covariates. By change of the distribution of the gender wage gap, we mean the counterfactual study observing the changes from actual distribution gender wage gap to hypothetical distribution gender wage gap, indicating the wage distribution that would have prevailed if the women had been paid according to women’s wage function and these women have the same observable characteristics as men.
counterparts. Interestingly, observable factors cannot explain the gender distribution wage gap faced by Thai women. The Antecol, Jong, and Steinberger (AJS) (2007) wage gap density methodology of the DFL suggests that if Thai women possess similar observable characteristics as men, gender wage inequality would be greater for the majority of the wage distribution, particularly, at the middle and high income distribution.

2.2 Methodology (DFL)

2.2.1 DiNardo, Fortin and Lemieux Decomposition (DFL)

The main improvement from the Oaxaca Blinder (OB) mean decomposition method is that the DFL approach decomposes the wage gap along the entire wage distribution instead of being limited to the mean analysis. The a priori assumption of normality in the wage distribution is not required. The DFL decomposition is the counterfactual study that quantifies the hypothetical wage distribution of female workers that would have prevailed assuming that women had been paid according to women’s wage function (face the women’s wage structure) and women have the same observable characteristics as men. As such, it is akin to the OB decomposition, but with distribution. The DFL approach utilizes a reweighing method by attaching weights to each observation and kernel density estimate in order to obtain counterfactual distribution\textsuperscript{17}.

Each worker can be represented as a vector \((w, z, s)\), where \(w\) indicates the log hourly earnings of the workers, \(z\) is a vector of worker observable attributes (e.g. occupation, firm estimated size, education, age, region and metropolitan area) and \(s\) is the

\textsuperscript{17} For a discussion of kernel density and their assumptions see Pagan and Ullah (1999) and Sun (2006)
gender of the workers (s=M or F). The actual density of female earnings can be written as follow:

\[
(7) \quad f^f(w) = \int_{z \in \Omega_z} f(w | z, s = F) dF(z | s = F) = \int_{z \in \Omega_z} f(w | z, s = F) f(z | s = F) \Psi(z) d\Psi(z) = \\
\int_{z \in \Omega_z} f(w, z | s = F) \Psi(z) d\Psi(z) = \int_{z \in \Omega_z} f(w, z | s = F) \Psi(z) dF(z | s = F)
\]

where \( f(w | z, s = F) \) is the distribution of earnings conditional on observable characteristics for women, \( dF(z | s = F) \) is the distribution of workers’ characteristics conditional on workers being women, and \( \Omega_z \) denotes these set of given attributes. The actual density of male earnings can be written similarly, replacing sex from female to male.

The counterfactual distribution of female earnings if the distribution of women’s attributes has the same distribution as men can be defined as follow\(^\text{18}\):

\[
(8) \quad f^c_f(w) = \int_{z \in \Omega_z} f(w | z, s = F) f(z | s = M) d\Psi(z) = \int_{z \in \Omega_z} f(w | z, s = F) f(z | s = F) \Psi(z) d\Psi(z) = \\
\int_{z \in \Omega_z} f(w, z | s = F) \Psi(z) d\Psi(z) = \int_{z \in \Omega_z} f(w, z | s = F) \Psi(z) dF(z | s = F)
\]

\[
(9) \quad \Psi(z) = \frac{dF(z | s = M)}{dF(z | s = F)} = \frac{f(z | s = M)}{f(z | s = F)}
\]

where \( \Psi(z) \) denotes “reweighing function.” The reweighting function is the product of hours sample weight and \([p/(1-p)]\), where \( p \) is the predicted probability of being either a man or a woman conditioned on their distinctive characteristics\(^\text{19}\). The predicted weight

\(^{18}\) Similar to AJS, we assume that \( f(w | z, S=F) \) does not change if the distribution of \( z \) changed. This assumption states that conditional distribution of female wage does not depend on the distribution of observable covariates.

\(^{19}\) The hours sample weight is calculated by multiplying the number of working hours per week and the observation weight.
can be estimated using either a probit or logit model. The reweighing function gives higher weight to the observations that belong in the minority categories.

The counterfactual distribution gender wage gap is defined as the counterfactual wage density of women minus the actual wage density of men. Specifically, the log hourly counterfactual distribution wage gap at each percentile is defined as the differences between the counterfactual female log hourly wage at that percentile and the actual male log hourly wage at the same percentile. The counterfactual female log hourly wage is the conditional distribution where observable covariates of females are changed so they are equivalent to their male counterparts.

The distribution gender wage gap can be decomposed into the counterfactual distribution wage gap that is attributed to if women possessing each of male observable characteristic and a residual factor. In other words, the counterfactual distribution wage gap can be viewed as the resulting wage gap after controlling for the differences between male and female occupation, firm size, education, age, region, and urban separately. Hence, after controlling for all the explainable characteristics differences between men and women, the remaining counterfactual distribution wage gap is the residual or the unexplainable portion of the distribution wage gap.

The DFL decomposition can be generalized by the following equation:

\[
f^f(w) - f^m(w) = [f^f(w) - f^f_{c1}(w)] + [f^f_{c1}(w) - f^f_{c2}(w)] + [f^f_{c2}(w) - f^f_{c3}(w)] + [f^f_{c3}(w) - f^f_{c4}(w)] + [f^f_{c4}(w) - f^f_{c5}(w)] + \ldots + [f^f_{cn-1}(w) - f^f_{cn}(w)] + [f^f_{cn}(w) - f^m(w)]
\]
The general equation for the female actual distribution and the counterfactual female distribution can be expressed as:

\[
\begin{align*}
 f^f(w) &= \int_{z_1} \cdots \int_{z_n} dF(w, z_1, ..., z_n \mid s = F) = \int_{z_1} \cdots \int_{z_n} f(w, z_1, ..., z_n \mid s = F)dz_1 ... dz_n = \\
(11) \\
\int_{z_1} \cdots \int_{z_n} f(w \mid z_1, ..., z_n, s = F) f(z_1 \mid z_2, ..., z_n, s = F) ... f(z_n \mid s = F)dz_1 ... dz_n
\end{align*}
\]

\[
\begin{align*}
 f^{f}_{ck}(w) &= \int_{z_1} \cdots \int_{z_n} f(w \mid z_1, ..., z_n, s = F) f(z_1 \mid z_2, ..., z_n, s = M) ... \\
&\quad ... f(z_k \mid z_{k+1}, ..., z_n, s = M) f(z_{k+1} \mid z_{k+2}, ..., z_n, s = F) ... f(z_n \mid s = F)dz_1 ... dz_n \\
(12) \\
&= \int_{z_1} \cdots \int_{z_n} f(w \mid z_1, ..., z_n, s = F) \Psi(z_1) ... \Psi(z_k) f(z_{k+1} \mid z_{k+2}, ..., z_n, s = F) ... f(z_n \mid s = F)dz_1 ... dz_n
\end{align*}
\]

where

\[
\Psi(z_i) = \frac{dF(z_i \mid z_2, ..., z_n, s = M)}{dF(z_i \mid z_2, ..., z_n, s = F)} = \frac{f(s = M \mid z_1, ..., z_i, z_{i+1}, ..., z_n)}{f(s = F \mid z_1, ..., z_i, z_{i+1}, ..., z_n)}
\]

(13) \[
\Psi(z_k) = \frac{dF(z_k \mid z_{k+1}, ..., z_n, s = M)}{dF(z_k \mid z_{k+1}, ..., z_n, s = F)} = \frac{f(s = M \mid z_{k+1}, ..., z_n)}{f(s = F \mid z_{k+1}, ..., z_n)}
\]

Since DFL uses conditional probability, the decomposition is order sensitive. The order in the sequential DFL procedure is important because we are conditioning on later covariates. Typically, covariates that are accounted later in the process typically carry more explanatory power. Hence, the later decomposition covariates are given greater weight in comparison to the earlier decomposition covariates. This is because later decomposition covariates are conditioned on lesser variables, and conditioning on fewer variables implies that covariate usually has more explanatory power. Nevertheless, when conditioning for all observable characteristics, the total effect is not order sensitive.
For our study, the sequential decomposition of 5 observable covariates will be in order of occupational sorting, human capital (education, experience), metropolitan area and region. As a robust check, the DFL decomposition in reverse covariates order is also examined. The result from the decomposition will be used to isolate the contribution from occupation sorting, human capital and demographic factors. In the next section, we will summarize and discuss the result of the DFL decomposition.

2.3 Analysis (DFL)

2.3.1 DiNardo, Fortin and Lemieux Decomposition (DFL)

This section examines the explanatory power of the factors that account for the differences in the male and female wage distributions in Thailand, such as occupational sorting, demographic differences, human capital differences and unexplained. Using the DFL technique and the AJS wage gap density methodology, we construct the counterfactual distribution of gender wage gap (male actual wage minus female’s counterfactual wage that would have prevailed if women had been paid according to female wage function and these women have similar distribution of observable characteristics as men).

Panels (A) to (F) of figure 2 (2006) and 3 (1997) exhibit our sequential DFL decomposition of the counterfactual distribution wage gap. We isolate the contribution of occupation sorting, demographic differences, human capital differences and unexplained portion play in changing counterfactual distribution wage gap. In panels (A), a solid line displays the actual raw log hourly wage gap (female actual wage minus male actual wage). Positive values on the distribution indicate wage advantage for women.
Hence, negative values show that men earn higher wage than women at the same wage percentile within each respect wage distribution.

A dashed line displays the counterfactual log hourly distribution wage gap (male actual wage minus female’s counterfactual wage that would have prevailed if conditional upon all other observable characteristics, the distribution of occupational sorting of women is changed such that it is similar to that of men). If occupational sorting is the main attribute that explain the wage gap, the gap between the dashed line and X-axis would decrease where dashed line would approximate the X-axis.

Panel (A) show that the dashed line does not overlap with the X-axis (See [Figure 2]). In fact, the graph shows that the space between the dashed line and X-axis widen. This is true especially at the top of the distribution. The increase in the distance between the dashed line and X-axis confirms our OB analysis that occupational sorting does not explain the wage gap. Also, we observe that the relative role of occupational sorting vary considerably across the distribution of wages.

The effect is greater at the top of the distribution than the bottom of the distribution. We find that suppose women have the similar occupation as men, the wage advantage enjoyed by men, especially for middle to high income, would greatly increase. In fact, except for very rich and very poor women, women are much worse off having the same occupation as men. A possible explanation is that women are sorting themselves into the high paying occupation (government jobs etc.) in comparison to men. For 1997, we find similar result except that the effect is greatest at middle of the distribution (See
In addition, we observe that the effect is approximately two times larger than the effect in 2006.

Panel (B) and (C) look at the effect of human capital factors (education and potential experience) on the distribution wage gap (See [Figure 2]). Dashed line that displays occupation sorting in Panel (A) is now the solid line in Panel (B). Panel (B) explores the role education play in explaining the distribution wage gap. We observe that the distribution wage gap would increase extensively for the top of the distribution.

After controlling for occupation sorting, we find that differences in education attainment cannot explain the wage different across the wage distribution. After accounting for differences in occupation and education between men and women, we would actually expect a much larger wage penalty than we find empirically. This is true especially between the 60th to 90th percentiles of the distribution where wage advantage enjoy by men would double. This may be a result of a large portion of Thai women have attained high education, especially those high earning workers.

Panel (C) exhibits the role potential experience plays in explaining the distribution wage gap. We find that the differences in potential experience have a very limited role in explaining the wage different across the wage distribution. Also, we observe that potential experience has more of an effect for explaining the wage differential for high-earning workers, especially between the 80th to 100th percentiles. However, in general, the differences in potential experience do not play a significant role in explaining the distribution wage gap. For 1997, we find similar result except that the
effect is larger for Education in 2006 and the effect is larger for potential experience in 1997 (See [Figure 3-Panel (B and C)]).

For the next two sequential decompositions, panel (D) and (E) look at the effect of the differences in metropolitan area and region between men and women workers on the distribution wage gap (See [Figure 2]). We find that the differences in metropolitan area and region do not play a significant role in explaining the distribution wage gap. For 1997, we find similar result that that the differences in metropolitan area and region do not play a significant role in explaining the distribution wage gap (See [Figure 3-Panel (B and C)]).

Panel (F) explore the role discrimination plays in explaining the distribution wage gap (See [Figure 2]). A solid line exhibits the actual raw log hourly wage gap at each percentile. A dashed line displays the counterfactual log hourly distribution wage gap, if conditional upon all other observable characteristics (occupational sorting, human capital differences, and differences in metropolitan area and region). The distance between the dashed line and the X-axis quantifies the role discrimination plays in explaining distribution wage gap.

Discrimination plays the most important role in explaining distribution wage gap. Specifically, discrimination accounts for the entire distribution wage gap plus the wage premium that women should have earned from having accrued more human capital accumulation relative to men. Furthermore, discrimination contributes more in explaining wage differential above the median than below the median of the distribution, especially between the $70^{th}$ to $90^{th}$ percentiles of the wage distribution. Hence,
discrimination plays a larger role in explaining the wage differential enjoyed by high earning men relative to low earning men.

To summarize, we find that discrimination explains wage differential by men relative to women at all percentiles of the distribution. In fact, when we control for occupation differences and education differences, we observe a much larger distribution wage gap unexplained. Even though wage differential is smallest at the upper part of the distribution, high earning women face more discrimination relative to low earning women. Since women accumulated more human capital relative to men, especially for high earning workers, the counterfactual distribution wage gap is greater than the empirically observed distribution wage gap, particularly at the upper part of the wage distribution. Although our results appear contrary to our intuition, it is because Thai women have attained higher education relative to men and managed to over compensate for the majority of the differences in wage premium by sorting themselves into higher paying occupation such as government jobs. Yet, taken together, our main findings of the mean decomposition analysis continue to hold.

2.4 Conclusion

In the last decade, Thailand has experienced a sizeable decline in gender wage gap. Using the 1997 and 2006 Thai Labor Force Survey, we examine the changes of the wage gap using the OB and the DFL decomposition. We explore possible explanations of gender wage gap: occupational sorting, human capital differences, demographic differences and unexplainable. Using the OB decomposition, we find that the unexplainable accounts for most of wage differential enjoyed by men. Particularly,
return to human capital in 1997 and return to occupation and education in 2006 are the main factors that account for the observed wage differential enjoyed by men relative to women.

Since the wage gap is not uniform throughout the wage distribution, we further explore our analysis to the entire distribution using the Antecol, Jong, and Steinberger (AJS) (2007) wage gap density methodology of the DFL decomposition (See [Figure 1]). The findings from the DFL decomposition are consistent with the OB decomposition. We find that discrimination plays the most important role in explaining the observed distribution wage gap for both 1997 and 2006.

For the most part, we find that observable characteristics do not help explain the distribution wage gap. After controlling for all observable covariate differences, wage differential become much larger for all percentile of the distribution, especially for middle income and high income workers (See [Figure 3-E]). Even though we find that wage differential for low income workers is largest empirically, wage differential for low income workers is smallest, when we control for all observable characteristics differences.

Looking at the distribution wage gap changing from 1997 to 2006, the result of DFL decomposition indicates that when we control for all observable characteristics differences, the counterfactual distribution wage gap becomes larger for both years, and largest in 2006 (See [Figure 2-E] and [Figure 3-E]). Specifically, the increase in the differences between counterfactual distribution wage gap relative to the empirically observed distribution wage gap is mostly resulted from the increase in already extensive
advantage that high earning men have in the return to occupation and education. Similar to the OB result, the increases in return of the observable covariates are the main driving force that resulted in the sustained wage differences of past decade. Yet, the distribution wage gap has declined for middle and high income workers because middle and high earning women manage to over compensate for the differences in return to the observable characteristics by attaining higher education relative to men and sorting themselves into higher paying occupations.

Consequently, although the recent decline in mean wage gap helped certain income segments of Thai workers more than the others, Thai women as a whole have benefited. Thailand set a good example in terms of improving women’s human capital accumulation, reducing occupation exclusion and preventing occupation segregation problem. Since unlike many other developing countries, Thai women have attained human capital accumulation, especially education, that are comparable to men. Nevertheless, there are still opportunities for improvement.

2.4.1 Policy Implication

Hence, Thailand can focus on mitigating the problem of discrimination, particularly for the middle and high income wage earners. Thus, Thailand can focus its attention on glass ceiling problem. Furthermore, though poor women are facing less discrimination relative to the rich, they are facing largest observed wage differential. These women are likely the individuals from poor families that do not have resources, such as the money to attain higher education. Thus, they would likely be trapped perpetually in low wage and low mobility jobs. Hence, Thailand can focus its attention
on the sticky floor problem. Thailand has definitely started on the right path in easing wage inequality problems. Thailand should continue with the current effort of improving human capital and search for more ways to minimize wage discrimination.
Bibliography


Mincer, Jacob, 1974. “Schooling, Experience, and Earnings.” *NBER working paper*


Sun, Wei, 2005. “Semiparametric Analysis of Income Inequality: An Application to China,” Dissertation, University of Riverside


Appendix

Table 1. Government Interventions Related to Gender Inequality

- The Asian Crisis of 1997
- Labor Protection Act B.E in 1998,
- National Education Act enacted in 1999
- Ten different versions of National and Economic Development Plan (NDP)
- Military Coup of 2006

Table 2. Five categorical regions

Thai Labor Force Survey divides Thailand into five regions.
- Central
- Northern,
- North Eastern
- Southern
- Bangkok, Capital of Thailand.

Table 3. Education categories

Prior to 2000, LFS categorizes education into five levels. The five groupings are Less than Primary, Primary, Secondary, University, Vocational, and Other. However, current LFS recodes education into fifteen levels. The seven main groupings of education are None, Less than Elementary, Elementary, Lower Secondary, Upper Secondary, Vocational, and University education. Upper Secondary level further subdivides into Academic, Vocational, and Teacher training. Also, both Diploma level and University education are subdivided into Academic, Higher technical education, and Teacher training.
Table 4. Kernel density

The kernel density estimate of the earning distribution can be represented as follows:

\[
(5) \quad \hat{f}_h(w) = \frac{1}{nh} \sum_{i=1}^{n} \psi_i K\left(\frac{w - W_i}{h}\right) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{w - W_i}{h}\right)
\]

\[
(6) \quad K(w) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{w^2}{2\sigma^2}\right]
\]

where \(W_i\) is the independent and identically distributed random sample of \(i\) worker’s earning, and \(h\) is the selected bandwidth (smoothing parameter). \(\psi_i\) is the weight for each observation with \(\sum \psi_i = 1\). The normalized hours weight density can be calculated by multiplying the number of working hours per week and the observation weight, \(\psi_i = 1/n\). \(K(.)\) is the Gaussian kernel density function with mean zero and variance \(\sigma^2\) that smoothen the area over \(W_i\) and \(w\).
### Descriptive Statistics Summary for Estimation Samples

#### Table 5. Dependent and Independent Variables and Definitions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>lwage</td>
<td>Natural logarithm of the real hourly wage rate (BAHT) in 1997 and 2006.</td>
</tr>
<tr>
<td>rtwage</td>
<td>Real hourly wage rate (BAHT) in 1997 and 2006.</td>
</tr>
<tr>
<td>age</td>
<td>Age in years.</td>
</tr>
<tr>
<td>age2</td>
<td>Squared of age/100.</td>
</tr>
<tr>
<td>exp</td>
<td>Potential experience, calculated as age minus education minus 6.</td>
</tr>
<tr>
<td>exp2</td>
<td>Squared of potential experience/100.</td>
</tr>
<tr>
<td>total_hr</td>
<td>Hours work per week.</td>
</tr>
<tr>
<td>_leduc_1</td>
<td>Dummy variable, =1 if one had no schooling; =0 otherwise.</td>
</tr>
<tr>
<td>_leduc_2</td>
<td>Dummy variable, =1 if one had some or completed lower elementary school (grade 4 and below); =0 otherwise.</td>
</tr>
<tr>
<td>_leduc_3</td>
<td>Dummy variable, =1 if one had completed upper elementary school (grade 5-6); =0 otherwise.</td>
</tr>
<tr>
<td>_leduc_4</td>
<td>Dummy variable, =1 if one had completed lower secondary school (grade 7-9); =0 otherwise.</td>
</tr>
<tr>
<td>_leduc_5</td>
<td>Dummy variable, =1 if one had completed upper secondary school (grade 10-12); =0 otherwise.</td>
</tr>
<tr>
<td>_leduc_6</td>
<td>Dummy variable, =1 if one had some or completed vocational college school (4 years or less); =0 otherwise.</td>
</tr>
<tr>
<td>_leduc_7</td>
<td>Dummy variable, =1 if one had some or completed college or post college school; =0 otherwise.</td>
</tr>
<tr>
<td>married</td>
<td>Dummy variable, =1 if one is married or one had ever been married; =0 otherwise.</td>
</tr>
<tr>
<td>_Ireg_1</td>
<td>Dummy variable, = 1 if Bangkok Metropolis; =0 otherwise</td>
</tr>
<tr>
<td>_Ireg_2</td>
<td>Dummy variable, = 1 if Central area; =0 otherwise</td>
</tr>
<tr>
<td>_Ireg_3</td>
<td>Dummy variable, = 1 if North area; =0 otherwise</td>
</tr>
<tr>
<td>_Ireg_4</td>
<td>Dummy variable, = 1 if North-East area Metropolis; =0 otherwise</td>
</tr>
<tr>
<td>_Ireg_5</td>
<td>Dummy variable, = 1 if South area; =0 otherwise</td>
</tr>
<tr>
<td>urban</td>
<td>Dummy variable, = 1 if Metropolitan Area; =0 otherwise</td>
</tr>
<tr>
<td>_Iocc_1</td>
<td>Dummy variable, =1 if occupation is a Managerial job; =0 otherwise.</td>
</tr>
<tr>
<td>_Iocc_2</td>
<td>Dummy variable, =1 if occupation is a Professional job; =0 otherwise.</td>
</tr>
<tr>
<td>_Iocc_3</td>
<td>Dummy variable, =1 if occupation is a Technician job; =0 otherwise.</td>
</tr>
<tr>
<td>_Iocc_4</td>
<td>Dummy variable, =1 if occupation is a Clerk job; =0 otherwise.</td>
</tr>
<tr>
<td>_Iocc_5</td>
<td>Dummy variable, =1 if occupation is a Service job; =0 otherwise.</td>
</tr>
<tr>
<td>_Iocc_6</td>
<td>Dummy variable, =1 if occupation is a Agricultural job; =0 otherwise.</td>
</tr>
<tr>
<td>_Iocc_7</td>
<td>Dummy variable, =1 if occupation is a Craft and related trades job; =0 otherwise.</td>
</tr>
<tr>
<td>_Iocc_8</td>
<td>Dummy variable, =1 if occupation is a Plant and machine operators and assemblers job; =0 otherwise.</td>
</tr>
<tr>
<td>_Iocc_9</td>
<td>Dummy variable, =1 if occupation is a Elementary job; =0 otherwise.</td>
</tr>
<tr>
<td>_Iestsize_1</td>
<td>Dummy variable, =1 if company size is small (workers 1-10); =0 otherwise.</td>
</tr>
<tr>
<td>_Iestsize_2</td>
<td>Dummy variable, =1 if company size is medium (workers 11-99); =0 otherwise.</td>
</tr>
<tr>
<td>_Iestsize_3</td>
<td>Dummy variable, =1 if company size is large (workers 100+); =0 otherwise.</td>
</tr>
</tbody>
</table>
Table 6. Descriptive Statistics of Dependent and Independent Variable in 1997 and 2006

<table>
<thead>
<tr>
<th>Variable</th>
<th>1997</th>
<th></th>
<th></th>
<th>2006</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs</td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Min</td>
<td>Max</td>
<td>Obs</td>
<td>Mean</td>
</tr>
<tr>
<td>lwage</td>
<td>24139</td>
<td>3.75</td>
<td>1.05</td>
<td>0.69</td>
<td>7.02</td>
<td>51491</td>
<td>3.41</td>
</tr>
<tr>
<td>rtwage</td>
<td>24139</td>
<td>67.98</td>
<td>67.26</td>
<td>2.00</td>
<td>1113.33</td>
<td>51491</td>
<td>45.22</td>
</tr>
<tr>
<td>total_hr</td>
<td>24139</td>
<td>41.74</td>
<td>15.14</td>
<td>5.00</td>
<td>98.00</td>
<td>51491</td>
<td>46.70</td>
</tr>
<tr>
<td>age</td>
<td>24139</td>
<td>35.10</td>
<td>10.34</td>
<td>16.00</td>
<td>65.00</td>
<td>51491</td>
<td>37.01</td>
</tr>
<tr>
<td>age2</td>
<td>24139</td>
<td>13.39</td>
<td>7.79</td>
<td>2.56</td>
<td>42.25</td>
<td>51491</td>
<td>14.86</td>
</tr>
<tr>
<td>exp</td>
<td>24139</td>
<td>18.16</td>
<td>11.16</td>
<td>0.00</td>
<td>59.00</td>
<td>51491</td>
<td>21.42</td>
</tr>
<tr>
<td>exp2</td>
<td>24139</td>
<td>4.58</td>
<td>5.17</td>
<td>0.00</td>
<td>34.81</td>
<td>51491</td>
<td>6.15</td>
</tr>
<tr>
<td>_leduc_1</td>
<td>24139</td>
<td>0.01</td>
<td>0.09</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.03</td>
</tr>
<tr>
<td>_leduc_2</td>
<td>24139</td>
<td>0.17</td>
<td>0.38</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.20</td>
</tr>
<tr>
<td>_leduc_3</td>
<td>24139</td>
<td>0.12</td>
<td>0.32</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.18</td>
</tr>
<tr>
<td>_leduc_4</td>
<td>24139</td>
<td>0.16</td>
<td>0.37</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.14</td>
</tr>
<tr>
<td>_leduc_5</td>
<td>24139</td>
<td>0.08</td>
<td>0.27</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.15</td>
</tr>
<tr>
<td>_leduc_6</td>
<td>24139</td>
<td>0.08</td>
<td>0.27</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.15</td>
</tr>
<tr>
<td>_leduc_7</td>
<td>24139</td>
<td>0.37</td>
<td>0.49</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.23</td>
</tr>
<tr>
<td>married</td>
<td>24139</td>
<td>0.71</td>
<td>0.45</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.75</td>
</tr>
<tr>
<td>_lreg_1</td>
<td>24139</td>
<td>0.19</td>
<td>0.39</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.09</td>
</tr>
<tr>
<td>_lreg_2</td>
<td>24139</td>
<td>0.22</td>
<td>0.41</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.39</td>
</tr>
<tr>
<td>_lreg_3</td>
<td>24139</td>
<td>0.17</td>
<td>0.37</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.18</td>
</tr>
<tr>
<td>_lreg_4</td>
<td>24139</td>
<td>0.34</td>
<td>0.47</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.17</td>
</tr>
<tr>
<td>_lreg_5</td>
<td>24139</td>
<td>0.08</td>
<td>0.27</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.17</td>
</tr>
<tr>
<td>urban</td>
<td>19991</td>
<td>0.69</td>
<td>0.46</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.66</td>
</tr>
<tr>
<td>_locc_1</td>
<td>24139</td>
<td>0.05</td>
<td>0.22</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.03</td>
</tr>
<tr>
<td>_locc_2</td>
<td>24139</td>
<td>0.29</td>
<td>0.45</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.13</td>
</tr>
<tr>
<td>_locc_3</td>
<td>24139</td>
<td>0.08</td>
<td>0.27</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.10</td>
</tr>
<tr>
<td>_locc_4</td>
<td>24139</td>
<td>0.16</td>
<td>0.36</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.09</td>
</tr>
<tr>
<td>_locc_5</td>
<td>24139</td>
<td>0.14</td>
<td>0.35</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.11</td>
</tr>
<tr>
<td>_locc_6</td>
<td>24139</td>
<td>0.02</td>
<td>0.13</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.07</td>
</tr>
<tr>
<td>_locc_7</td>
<td>24139</td>
<td>0.15</td>
<td>0.36</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.16</td>
</tr>
<tr>
<td>_locc_8</td>
<td>24139</td>
<td>0.02</td>
<td>0.15</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.13</td>
</tr>
<tr>
<td>_locc_9</td>
<td>24139</td>
<td>0.10</td>
<td>0.30</td>
<td>0.00</td>
<td>1.00</td>
<td>51491</td>
<td>0.19</td>
</tr>
<tr>
<td>_estsize_1</td>
<td>12733</td>
<td>0.35</td>
<td>0.48</td>
<td>0.00</td>
<td>1.00</td>
<td>36742</td>
<td>0.42</td>
</tr>
<tr>
<td>_estsize_2</td>
<td>12733</td>
<td>0.32</td>
<td>0.47</td>
<td>0.00</td>
<td>1.00</td>
<td>36742</td>
<td>0.28</td>
</tr>
<tr>
<td>_estsize_3</td>
<td>12733</td>
<td>0.32</td>
<td>0.47</td>
<td>0.00</td>
<td>1.00</td>
<td>36742</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Notes: Sample includes all employees between 16 to 65 years of age, in a week worked at least one hour for wage, who participated in 1997 and 2006 Thai Labor Force Survey, excluding self employ workers and students. The total hourly real wage includes wage, bonus and overtime (1997 base year). Calculations apply sample weights.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs</td>
<td>Mean</td>
</tr>
<tr>
<td>lwage</td>
<td>13195</td>
<td>3.80</td>
</tr>
<tr>
<td>rtwage</td>
<td>13195</td>
<td>70.56</td>
</tr>
<tr>
<td>total_hr</td>
<td>13195</td>
<td>42.45</td>
</tr>
<tr>
<td>age</td>
<td>13195</td>
<td>36.46</td>
</tr>
<tr>
<td>age2</td>
<td>13195</td>
<td>14.43</td>
</tr>
<tr>
<td>exp</td>
<td>13195</td>
<td>20.08</td>
</tr>
<tr>
<td>exp2</td>
<td>13195</td>
<td>5.41</td>
</tr>
<tr>
<td>_Ieduc_1</td>
<td>13195</td>
<td>0.01</td>
</tr>
<tr>
<td>_Ieduc_2</td>
<td>13195</td>
<td>0.20</td>
</tr>
<tr>
<td>_Ieduc_3</td>
<td>13195</td>
<td>0.12</td>
</tr>
<tr>
<td>_Ieduc_4</td>
<td>13195</td>
<td>0.20</td>
</tr>
<tr>
<td>_Ieduc_5</td>
<td>13195</td>
<td>0.09</td>
</tr>
<tr>
<td>_Ieduc_6</td>
<td>13195</td>
<td>0.08</td>
</tr>
<tr>
<td>_Ieduc_7</td>
<td>13195</td>
<td>0.31</td>
</tr>
<tr>
<td>married</td>
<td>13195</td>
<td>0.77</td>
</tr>
<tr>
<td>_Ireg_1</td>
<td>13195</td>
<td>0.19</td>
</tr>
<tr>
<td>_Ireg_2</td>
<td>13195</td>
<td>0.24</td>
</tr>
<tr>
<td>_Ireg_3</td>
<td>13195</td>
<td>0.17</td>
</tr>
<tr>
<td>_Ireg_4</td>
<td>13195</td>
<td>0.33</td>
</tr>
<tr>
<td>_Ireg_5</td>
<td>13195</td>
<td>0.08</td>
</tr>
<tr>
<td>urban</td>
<td>10711</td>
<td>0.67</td>
</tr>
<tr>
<td>_Iocc_1</td>
<td>13195</td>
<td>0.07</td>
</tr>
<tr>
<td>_Iocc_2</td>
<td>13195</td>
<td>0.21</td>
</tr>
<tr>
<td>_Iocc_3</td>
<td>13195</td>
<td>0.13</td>
</tr>
<tr>
<td>_Iocc_4</td>
<td>13195</td>
<td>0.12</td>
</tr>
<tr>
<td>_Iocc_5</td>
<td>13195</td>
<td>0.17</td>
</tr>
<tr>
<td>_Iocc_6</td>
<td>13195</td>
<td>0.02</td>
</tr>
<tr>
<td>_Iocc_7</td>
<td>13195</td>
<td>0.18</td>
</tr>
<tr>
<td>_Iocc_8</td>
<td>13195</td>
<td>0.03</td>
</tr>
<tr>
<td>_Iocc_9</td>
<td>13195</td>
<td>0.07</td>
</tr>
<tr>
<td>_Iestsize_1</td>
<td>6719</td>
<td>0.36</td>
</tr>
<tr>
<td>_Iestsize_2</td>
<td>6719</td>
<td>0.33</td>
</tr>
<tr>
<td>_Iestsize_3</td>
<td>6719</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Notes: Sample includes all employees between 16 to 65 years of age, in a week worked at least one hour for wage, who participated in 1997 and 2006 Thai Labor Force Survey, excluding self employ workers and students. The total hourly real wage includes wage, bonus and overtime (1997 base year). Calculations apply sample weights.
Table 8. Descriptive Statistics of Men and Women in 2006

<table>
<thead>
<tr>
<th>Variable</th>
<th>Men Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
<th>Women Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>lwage</td>
<td>27459</td>
<td>3.44</td>
<td>0.84</td>
<td>0.71</td>
<td>6.81</td>
<td>24032</td>
<td>3.37</td>
<td>0.88</td>
<td>0.71</td>
<td>6.55</td>
</tr>
<tr>
<td>rtwage</td>
<td>27459</td>
<td>46.39</td>
<td>51.72</td>
<td>2.04</td>
<td>907.04</td>
<td>24032</td>
<td>43.89</td>
<td>46.34</td>
<td>2.04</td>
<td>698.44</td>
</tr>
<tr>
<td>total_hr</td>
<td>27459</td>
<td>47.25</td>
<td>12.68</td>
<td>2.00</td>
<td>98.00</td>
<td>24032</td>
<td>46.09</td>
<td>12.24</td>
<td>1.00</td>
<td>98.00</td>
</tr>
<tr>
<td>age</td>
<td>27459</td>
<td>37.36</td>
<td>11.04</td>
<td>16.00</td>
<td>65.00</td>
<td>24032</td>
<td>36.62</td>
<td>10.48</td>
<td>16.00</td>
<td>65.00</td>
</tr>
<tr>
<td>age2</td>
<td>27459</td>
<td>15.18</td>
<td>8.60</td>
<td>2.56</td>
<td>42.25</td>
<td>24032</td>
<td>14.51</td>
<td>8.09</td>
<td>2.56</td>
<td>42.25</td>
</tr>
<tr>
<td>exp</td>
<td>27459</td>
<td>22.13</td>
<td>12.44</td>
<td>0.00</td>
<td>59.00</td>
<td>24032</td>
<td>20.61</td>
<td>12.48</td>
<td>0.00</td>
<td>59.00</td>
</tr>
<tr>
<td>exp2</td>
<td>27459</td>
<td>6.44</td>
<td>6.30</td>
<td>0.00</td>
<td>34.81</td>
<td>24032</td>
<td>5.81</td>
<td>6.07</td>
<td>0.00</td>
<td>34.81</td>
</tr>
<tr>
<td>_Ieduc_1</td>
<td>27459</td>
<td>0.02</td>
<td>0.15</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.04</td>
<td>0.19</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_Ieduc_2</td>
<td>27459</td>
<td>0.21</td>
<td>0.40</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.20</td>
<td>0.40</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_Ieduc_3</td>
<td>27459</td>
<td>0.20</td>
<td>0.40</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.15</td>
<td>0.36</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_Ieduc_4</td>
<td>27459</td>
<td>0.17</td>
<td>0.37</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.12</td>
<td>0.32</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_Ieduc_5</td>
<td>27459</td>
<td>0.17</td>
<td>0.37</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.13</td>
<td>0.34</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_Ieduc_6</td>
<td>27459</td>
<td>0.07</td>
<td>0.25</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.06</td>
<td>0.24</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_Ieduc_7</td>
<td>27459</td>
<td>0.18</td>
<td>0.38</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.30</td>
<td>0.46</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>married</td>
<td>27459</td>
<td>0.75</td>
<td>0.43</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.74</td>
<td>0.44</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_Ireg_1</td>
<td>27459</td>
<td>0.08</td>
<td>0.27</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.10</td>
<td>0.30</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_Ireg_2</td>
<td>27459</td>
<td>0.38</td>
<td>0.49</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.40</td>
<td>0.49</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_Ireg_3</td>
<td>27459</td>
<td>0.18</td>
<td>0.39</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.18</td>
<td>0.39</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_Ireg_4</td>
<td>27459</td>
<td>0.18</td>
<td>0.38</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.17</td>
<td>0.37</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_Ireg_5</td>
<td>27459</td>
<td>0.18</td>
<td>0.38</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.16</td>
<td>0.37</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>urban</td>
<td>27459</td>
<td>0.65</td>
<td>0.48</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.68</td>
<td>0.47</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_locc_1</td>
<td>27459</td>
<td>0.05</td>
<td>0.21</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.02</td>
<td>0.13</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_locc_2</td>
<td>27459</td>
<td>0.09</td>
<td>0.28</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.18</td>
<td>0.38</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_locc_3</td>
<td>27459</td>
<td>0.09</td>
<td>0.29</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.10</td>
<td>0.29</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_locc_4</td>
<td>27459</td>
<td>0.05</td>
<td>0.23</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.12</td>
<td>0.33</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_locc_5</td>
<td>27459</td>
<td>0.09</td>
<td>0.28</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.13</td>
<td>0.34</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_locc_6</td>
<td>27459</td>
<td>0.07</td>
<td>0.26</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.06</td>
<td>0.24</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_locc_7</td>
<td>27459</td>
<td>0.22</td>
<td>0.42</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.09</td>
<td>0.29</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_locc_8</td>
<td>27459</td>
<td>0.16</td>
<td>0.36</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.10</td>
<td>0.30</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_locc_9</td>
<td>27459</td>
<td>0.18</td>
<td>0.38</td>
<td>0.00</td>
<td>1.00</td>
<td>24032</td>
<td>0.20</td>
<td>0.40</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_Iestsize_1</td>
<td>19631</td>
<td>0.45</td>
<td>0.50</td>
<td>0.00</td>
<td>1.00</td>
<td>17111</td>
<td>0.39</td>
<td>0.49</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_Iestsize_2</td>
<td>19631</td>
<td>0.30</td>
<td>0.46</td>
<td>0.00</td>
<td>1.00</td>
<td>17111</td>
<td>0.27</td>
<td>0.44</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>_Iestsize_3</td>
<td>19631</td>
<td>0.26</td>
<td>0.44</td>
<td>0.00</td>
<td>1.00</td>
<td>17111</td>
<td>0.34</td>
<td>0.47</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Notes: Sample includes all employees between 16 to 65 years of age, in a week worked at least one hour for wage, who participated in 1997 and 2006 Thai Labor Force Survey, excluding self employ workers and students. The total hourly real wage includes wage, bonus and overtime (1997 base year). Calculations apply sample weights.
Table 9. The results of unconditional OLS pooled model using 2006 and 1997 data

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Hourly Wage Gap</td>
<td>0.090 (0.007)</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>3.471 (0.005)</td>
<td></td>
</tr>
<tr>
<td>Observation</td>
<td>75630</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.002</td>
<td></td>
</tr>
</tbody>
</table>

Chow test

\[(2006, 1997)\]

\[\text{chi}^2(1) = 7.87\]

\[\text{Prob} > \text{chi}^2 = 0.0050\]

Note: Standard Errors are in the parentheses.

Table 10. The results of conditional OLS pooled model using 2006 and 1997 data

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>0.200 (0.004)</td>
<td></td>
</tr>
<tr>
<td>exp</td>
<td>0.051 (0.001)</td>
<td></td>
</tr>
<tr>
<td>exp2</td>
<td>-0.059 (0.002)</td>
<td></td>
</tr>
<tr>
<td>_leduc_2</td>
<td>0.180 (0.015)</td>
<td></td>
</tr>
<tr>
<td>_leduc_3</td>
<td>0.455 (0.016)</td>
<td></td>
</tr>
<tr>
<td>_leduc_4</td>
<td>0.777 (0.017)</td>
<td></td>
</tr>
<tr>
<td>_leduc_5</td>
<td>0.991 (0.017)</td>
<td></td>
</tr>
<tr>
<td>_leduc_6</td>
<td>1.285 (0.019)</td>
<td></td>
</tr>
<tr>
<td>_leduc_7</td>
<td>1.594 (0.019)</td>
<td></td>
</tr>
<tr>
<td>married</td>
<td>0.052 (0.005)</td>
<td></td>
</tr>
<tr>
<td>_lreg_1</td>
<td>0.259 (0.008)</td>
<td></td>
</tr>
<tr>
<td>_lreg_2</td>
<td>0.182 (0.006)</td>
<td></td>
</tr>
<tr>
<td>_lreg_4</td>
<td>0.042 (0.007)</td>
<td></td>
</tr>
<tr>
<td>_lreg_5</td>
<td>0.277 (0.008)</td>
<td></td>
</tr>
<tr>
<td>urban</td>
<td>0.082 (0.005)</td>
<td></td>
</tr>
<tr>
<td>_loc_1</td>
<td>0.688 (0.017)</td>
<td></td>
</tr>
<tr>
<td>_loc_2</td>
<td>0.714 (0.012)</td>
<td></td>
</tr>
<tr>
<td>_loc_3</td>
<td>0.523 (0.011)</td>
<td></td>
</tr>
<tr>
<td>_loc_4</td>
<td>0.431 (0.010)</td>
<td></td>
</tr>
<tr>
<td>_loc_5</td>
<td>0.172 (0.009)</td>
<td></td>
</tr>
<tr>
<td>_loc_6</td>
<td>0.041 (0.013)</td>
<td></td>
</tr>
<tr>
<td>_loc_7</td>
<td>0.098 (0.008)</td>
<td></td>
</tr>
<tr>
<td>_loc_8</td>
<td>0.248 (0.009)</td>
<td></td>
</tr>
<tr>
<td>_lesstsize_2</td>
<td>-0.084 (0.006)</td>
<td></td>
</tr>
<tr>
<td>_lesstsize_3</td>
<td>0.066 (0.006)</td>
<td></td>
</tr>
<tr>
<td>_lyear_2006</td>
<td>-0.162 (0.006)</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>1.396 (0.018)</td>
<td></td>
</tr>
<tr>
<td>Observation</td>
<td>46826</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.477</td>
<td></td>
</tr>
</tbody>
</table>

Chow test

\[(2006, 1997)\]

\[\text{chi}^2(25) = 1765.66\]

\[\text{Prob} > \text{chi}^2 = 0.0000\]

Note: Standard Errors are in the parentheses.
### Table 11. The results of OLS pooled model with different specifications

<table>
<thead>
<tr>
<th>Male coefficient</th>
<th>0.090 (0.007)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconditional log real hourly wage differential</td>
<td>0.090 (0.007)</td>
</tr>
</tbody>
</table>

Log wage differential controlling for:

<table>
<thead>
<tr>
<th>Specification</th>
<th>Male coefficient</th>
<th>0.087 (0.007)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. experience and experience square/100</td>
<td>0.087 (0.007)</td>
<td></td>
</tr>
<tr>
<td>1. Variables in (0) plus education</td>
<td>0.187 (0.005)</td>
<td></td>
</tr>
<tr>
<td>2. Variables in (1) plus married</td>
<td>0.187 (0.005)</td>
<td></td>
</tr>
<tr>
<td>3. Variables in (2) plus region</td>
<td>0.189 (0.005)</td>
<td></td>
</tr>
<tr>
<td>4. Variables in (3) plus metropolitan area</td>
<td>0.188 (0.005)</td>
<td></td>
</tr>
<tr>
<td>5. Variables in (4) plus occupation</td>
<td>0.203 (0.004)</td>
<td></td>
</tr>
<tr>
<td>6. Variables in (5) plus firm size</td>
<td>0.209 (0.005)</td>
<td></td>
</tr>
</tbody>
</table>


### Table 12. The results of OLS models 1997 and 2006

<table>
<thead>
<tr>
<th>Variable</th>
<th>1997</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Hourly Wage Gap</td>
<td>0.172 (0.010)</td>
<td>0.184 (0.005)</td>
</tr>
<tr>
<td>exp</td>
<td>0.065 (0.002)</td>
<td>0.055 (0.001)</td>
</tr>
<tr>
<td>exp2</td>
<td>-0.062 (0.003)</td>
<td>-0.055 (0.001)</td>
</tr>
<tr>
<td>_Ieduc_2</td>
<td>0.550 (0.054)</td>
<td>0.169 (0.015)</td>
</tr>
<tr>
<td>_Ieduc_3</td>
<td>0.959 (0.056)</td>
<td>0.493 (0.016)</td>
</tr>
<tr>
<td>_Ieduc_4</td>
<td>1.463 (0.055)</td>
<td>0.884 (0.016)</td>
</tr>
<tr>
<td>_Ieduc_5</td>
<td>1.810 (0.057)</td>
<td>1.191 (0.016)</td>
</tr>
<tr>
<td>_Ieduc_6</td>
<td>2.143 (0.056)</td>
<td>2.119 (0.016)</td>
</tr>
<tr>
<td>married</td>
<td>0.063 (0.012)</td>
<td>0.057 (0.006)</td>
</tr>
<tr>
<td>_Ireg_1</td>
<td>0.093 (0.016)</td>
<td>0.390 (0.010)</td>
</tr>
<tr>
<td>_Ireg_2</td>
<td>0.056 (0.015)</td>
<td>0.255 (0.007)</td>
</tr>
<tr>
<td>_Ireg_4</td>
<td>0.055 (0.015)</td>
<td>0.045 (0.008)</td>
</tr>
<tr>
<td>_Ireg_5</td>
<td>0.428 (0.020)</td>
<td>0.250 (0.008)</td>
</tr>
<tr>
<td>urban</td>
<td>0.040 (0.011)</td>
<td>0.091 (0.005)</td>
</tr>
<tr>
<td>_locc</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>_Iestsiz</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.851 (0.057)</td>
<td>1.233 (0.018)</td>
</tr>
</tbody>
</table>

Observation | 19991 | 10711 | 9280 | 51491 | 27459 | 24032 |
R-squared | 0.577 | 0.526 | 0.636 | 0.605 | 0.567 | 0.647 |

Table 13. The results of OLS male and female models using 1997 and 2006

<table>
<thead>
<tr>
<th></th>
<th>1997 All</th>
<th>1997 Male</th>
<th>1997 Female</th>
<th>2006 All</th>
<th>2006 Male</th>
<th>2006 Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Hourly Wage Gap</td>
<td>0.217 (0.010)</td>
<td>0.198 (0.005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>exp</td>
<td>0.055 (0.002)</td>
<td>0.061 (0.002)</td>
<td>0.051 (0.002)</td>
<td>0.047 (0.001)</td>
<td>0.047 (0.001)</td>
<td>0.048 (0.001)</td>
</tr>
<tr>
<td>exp2</td>
<td>-0.053 (0.003)</td>
<td>-0.063 (0.004)</td>
<td>-0.048 (0.005)</td>
<td>-0.057 (0.001)</td>
<td>-0.054 (0.002)</td>
<td>-0.062 (0.002)</td>
</tr>
<tr>
<td>Ieduc_2</td>
<td>0.520 (0.053)</td>
<td>0.675 (0.086)</td>
<td>0.356 (0.065)</td>
<td>0.147 (0.014)</td>
<td>0.154 (0.022)</td>
<td>0.119 (0.018)</td>
</tr>
<tr>
<td>Ieduc_3</td>
<td>0.846 (0.054)</td>
<td>1.017 (0.088)</td>
<td>0.680 (0.067)</td>
<td>0.398 (0.015)</td>
<td>0.391 (0.023)</td>
<td>0.385 (0.020)</td>
</tr>
<tr>
<td>Ieduc_4</td>
<td>1.236 (0.054)</td>
<td>1.400 (0.087)</td>
<td>1.047 (0.067)</td>
<td>0.688 (0.016)</td>
<td>0.673 (0.024)</td>
<td>0.677 (0.021)</td>
</tr>
<tr>
<td>Ieduc_5</td>
<td>1.502 (0.056)</td>
<td>1.653 (0.090)</td>
<td>1.347 (0.071)</td>
<td>0.898 (0.016)</td>
<td>0.866 (0.024)</td>
<td>0.874 (0.022)</td>
</tr>
<tr>
<td>Ieduc_6</td>
<td>1.807 (0.056)</td>
<td>1.947 (0.090)</td>
<td>1.627 (0.071)</td>
<td>1.155 (0.018)</td>
<td>1.158 (0.027)</td>
<td>1.118 (0.025)</td>
</tr>
<tr>
<td>Ieduc_7</td>
<td>2.102 (0.056)</td>
<td>2.270 (0.089)</td>
<td>1.895 (0.070)</td>
<td>1.473 (0.018)</td>
<td>1.488 (0.027)</td>
<td>1.418 (0.024)</td>
</tr>
<tr>
<td>married</td>
<td>0.071 (0.012)</td>
<td>0.086 (0.019)</td>
<td>0.062 (0.015)</td>
<td>0.054 (0.006)</td>
<td>0.071 (0.009)</td>
<td>0.032 (0.008)</td>
</tr>
<tr>
<td>Ireg_1</td>
<td>0.078 (0.016)</td>
<td>0.083 (0.022)</td>
<td>0.076 (0.021)</td>
<td>0.384 (0.010)</td>
<td>0.375 (0.014)</td>
<td>0.395 (0.013)</td>
</tr>
<tr>
<td>Ireg_2</td>
<td>0.031 (0.015)</td>
<td>0.060 (0.021)</td>
<td>0.067 (0.021)</td>
<td>0.219 (0.007)</td>
<td>0.217 (0.009)</td>
<td>0.225 (0.009)</td>
</tr>
<tr>
<td>Ireg_4</td>
<td>0.072 (0.015)</td>
<td>0.070 (0.021)</td>
<td>0.086 (0.020)</td>
<td>0.021 (0.008)</td>
<td>0.010 (0.011)</td>
<td>0.032 (0.011)</td>
</tr>
<tr>
<td>Ireg_5</td>
<td>0.469 (0.019)</td>
<td>0.447 (0.028)</td>
<td>0.506 (0.026)</td>
<td>0.240 (0.008)</td>
<td>0.244 (0.011)</td>
<td>0.232 (0.011)</td>
</tr>
<tr>
<td>urban</td>
<td>0.050 (0.010)</td>
<td>0.083 (0.015)</td>
<td>0.000 (0.015)</td>
<td>0.071 (0.005)</td>
<td>0.082 (0.007)</td>
<td>0.057 (0.007)</td>
</tr>
<tr>
<td>Iloc_1</td>
<td>0.518 (0.029)</td>
<td>0.287 (0.040)</td>
<td>0.800 (0.045)</td>
<td>0.750 (0.016)</td>
<td>0.724 (0.019)</td>
<td>0.826 (0.028)</td>
</tr>
<tr>
<td>Iloc_2</td>
<td>0.579 (0.023)</td>
<td>0.355 (0.036)</td>
<td>0.736 (0.032)</td>
<td>0.781 (0.013)</td>
<td>0.693 (0.018)</td>
<td>0.837 (0.018)</td>
</tr>
<tr>
<td>Iloc_3</td>
<td>0.298 (0.024)</td>
<td>0.127 (0.033)</td>
<td>0.611 (0.054)</td>
<td>0.592 (0.011)</td>
<td>0.646 (0.015)</td>
<td>0.525 (0.016)</td>
</tr>
<tr>
<td>Iloc_4</td>
<td>0.331 (0.021)</td>
<td>0.187 (0.034)</td>
<td>0.414 (0.029)</td>
<td>0.456 (0.011)</td>
<td>0.493 (0.017)</td>
<td>0.425 (0.015)</td>
</tr>
<tr>
<td>Iloc_5</td>
<td>0.088 (0.020)</td>
<td>-0.016 (0.030)</td>
<td>0.090 (0.029)</td>
<td>0.201 (0.009)</td>
<td>0.289 (0.014)</td>
<td>0.128 (0.012)</td>
</tr>
<tr>
<td>Iloc_6</td>
<td>0.005 (0.046)</td>
<td>-0.177 (0.057)</td>
<td>0.248 (0.091)</td>
<td>0.073 (0.010)</td>
<td>0.047 (0.014)</td>
<td>0.117 (0.015)</td>
</tr>
<tr>
<td>Iloc_7</td>
<td>-0.011 (0.021)</td>
<td>-0.065 (0.031)</td>
<td>-0.144 (0.031)</td>
<td>0.121 (0.008)</td>
<td>0.175 (0.010)</td>
<td>-0.003 (0.013)</td>
</tr>
<tr>
<td>Iloc_8</td>
<td>0.040 (0.039)</td>
<td>-0.043 (0.051)</td>
<td>-0.050 (0.068)</td>
<td>0.268 (0.009)</td>
<td>0.291 (0.012)</td>
<td>0.216 (0.014)</td>
</tr>
<tr>
<td>Iestsize_2</td>
<td>-0.146 (0.014)</td>
<td>-0.194 (0.019)</td>
<td>-0.098 (0.019)</td>
<td>-0.095 (0.006)</td>
<td>-0.103 (0.008)</td>
<td>-0.080 (0.009)</td>
</tr>
<tr>
<td>Iestsize_3</td>
<td>-0.038 (0.015)</td>
<td>-0.034 (0.021)</td>
<td>0.022 (0.021)</td>
<td>0.041 (0.007)</td>
<td>0.000 (0.010)</td>
<td>0.092 (0.010)</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.019 (0.056)</td>
<td>1.128 (0.092)</td>
<td>1.174 (0.070)</td>
<td>1.348 (0.018)</td>
<td>1.511 (0.026)</td>
<td>1.426 (0.024)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1997</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>1999</td>
<td>2006</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.600</td>
<td>0.609</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1997 (M,F)</th>
<th>2006 (M,F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>chi2(16)</td>
<td>183.13</td>
<td>122.66</td>
</tr>
<tr>
<td>Prob &gt; chi2</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

### Table 14. Oaxaca decomposition 1997 and 2006

<table>
<thead>
<tr>
<th></th>
<th>1997 Male-Female Differential</th>
<th>2006 Male-Female Differential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Monthly Wage gap:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male mean prediction</td>
<td>3.875</td>
<td>3.444</td>
</tr>
<tr>
<td>Female mean prediction</td>
<td>3.752</td>
<td>3.372</td>
</tr>
<tr>
<td>Unadjusted gap</td>
<td>0.123</td>
<td>0.072</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Male coefficient</th>
<th>Explained</th>
<th>Unexplained</th>
<th>Explained</th>
<th>Unexplained</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Log Monthly Wage gap</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>exp</td>
<td>0.262</td>
<td>0.145</td>
<td>0.083</td>
<td>-0.075</td>
</tr>
<tr>
<td>exp2</td>
<td>-0.108</td>
<td>-0.062</td>
<td>-0.041</td>
<td>0.088</td>
</tr>
<tr>
<td>educ</td>
<td>-0.205</td>
<td>0.142</td>
<td>-0.146</td>
<td>0.014</td>
</tr>
<tr>
<td>married</td>
<td>0.008</td>
<td>0.019</td>
<td>0.001</td>
<td>0.042</td>
</tr>
<tr>
<td>reg</td>
<td>-0.005</td>
<td>-0.013</td>
<td>-0.005</td>
<td>-0.014</td>
</tr>
<tr>
<td>urban</td>
<td>-0.002</td>
<td>0.052</td>
<td>-0.002</td>
<td>0.028</td>
</tr>
<tr>
<td>All variables</td>
<td>-0.049</td>
<td>0.172</td>
<td>-0.112</td>
<td>0.184</td>
</tr>
</tbody>
</table>

| Percentage by exp      | 212.83%   | 117.74%     | 114.80%   | -104.24%    |
| Percentage by exp2     | -87.82%   | -50.03%     | -57.24%   | 122.31%     |
| Percentage by educ     | -166.48%  | 115.83%     | -203.48%  | 18.85%      |
| Percentage by married  | 6.59%     | 15.14%      | 0.86%     | 57.67%      |
| Percentage by reg      | -3.69%    | -10.80%     | -7.10%    | -18.98%     |
| Percentage by urban    | -1.46%    | 42.43%      | -3.02%    | 38.72%      |
| Percentage by all variables | -40.02%  | 140.02%     | -155.18%  | 255.18%     |

### Table 15. Oaxaca decomposition 1997 and 2006 (Occupation)

<table>
<thead>
<tr>
<th></th>
<th>1997</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male-Female Differential</td>
<td>Male-Female Differential</td>
</tr>
<tr>
<td>Log Hourly Wage gap:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male mean prediction</td>
<td>3.875</td>
<td>3.444</td>
</tr>
<tr>
<td>Female mean prediction</td>
<td>3.752</td>
<td>3.372</td>
</tr>
<tr>
<td>Unadjusted gap</td>
<td>0.123</td>
<td>0.072</td>
</tr>
<tr>
<td>Male coefficient</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explained</td>
<td>0.228</td>
<td>0.211</td>
</tr>
<tr>
<td>Unexplained</td>
<td>-0.095</td>
<td>-0.082</td>
</tr>
<tr>
<td>exp</td>
<td>-0.152</td>
<td>0.365</td>
</tr>
<tr>
<td>educ</td>
<td>0.009</td>
<td>0.017</td>
</tr>
<tr>
<td>married</td>
<td>-0.005</td>
<td>-0.036</td>
</tr>
<tr>
<td>reg</td>
<td>-0.002</td>
<td>0.058</td>
</tr>
<tr>
<td>urban</td>
<td>-0.076</td>
<td>0.277</td>
</tr>
<tr>
<td>occ</td>
<td>-0.093</td>
<td>0.216</td>
</tr>
<tr>
<td>All variables</td>
<td>-0.223</td>
<td>-0.016</td>
</tr>
</tbody>
</table>

| Percentage by exp    | 185.65%       | 171.62%       |
| Percentage by exp2   | -77.40%       | -66.45%       |
| Percentage by educ   | -123.56%      | 296.99%       |
| Percentage by married| 7.22%         | 13.93%        |
| Percentage by reg    | -4.41%        | -29.48%       |
| Percentage by urban  | -1.62%        | 47.20%        |
| Percentage by occ    | -61.71%       | 225.34%       |
| Percentage by all variables | -75.84% | 175.84% |

Figure 1. 1997 and 2006 Actual Gender Log Wage Gap by Percentile

Thai log hourly wage gap by percentile.

Sources: LFS 1997 and 2006.
Figure 2. 2006 Counterfactual Wage Gap DiNardo, Fortin, Lemieux Decomposition

Results:

(A) Occupation and Actual

(B) Education and (A) Occupation

(C) Experience and (B) Education

(D) Urban and (C) Experience

(E) Region and (D) Urban

Sources: LFS 2006

Notes: The log hourly wage gap (y-axis) at each wage percentile (x-axis) is defined as the counterfactual female log hourly wage at that percentile (i.e., the conditional distribution of indicated characteristics of females are changed so they are equivalent to their male counterparts) minus the actual male log hourly wage at the same percentile.
Figure 3. 1997 Counterfactual Wage Gap DiNardo, Fortin, Lemieux Decomposition Results:

Sources: LFS 1997
Notes: The log hourly wage gap (y-axis) at each wage percentile (x-axis) is defined as the counterfactual female log hourly wage at that percentile (i.e., the conditional distribution of indicated characteristics of females are changed so they are equivalent to their male counterparts) minus the actual male log hourly wage at the same percentile.
Figure 4. Probability Distribution of Log Hourly Earnings.
Panel (A) 1997 Actual Wage Density for Thai Female and Male Workers

Panel (B) 2006 Actual Wage Density for Thai Female and Male Workers

Sources: LFS 1997 and 2006.
Notes: For top figure, the blue line represents the female wage density, while the red dash line represents the male wage density.
Figure 5. Panel (A-G). Counterfactual Probability Distribution of Log Hourly Earnings. 1997 Female Counterfactual Wage Density
Sources: LFS 1997.
Notes: The counterfactual density is density of female workers that would have prevailed suppose women are paid with their wage function (face the same wage structure) and women have the observable characteristic of men.
Sources: LFS 2006.
Notes: The counterfactual density is density of female workers that would have prevailed suppose women are paid with their wage function (face the same wage structure) and women have the observable characteristic of men.
STATA code for Figure 1. 1997 and 2006 Actual Gender Log Wage Gap by Percentile

******************************************************************************
/* I. Determine sample ***/
******************************************************************************
tab female
gen hweight=total_hr

******************************************************************************
/* (0) determine the weight associated with covariates of men */
******************************************************************************
logit female
predict pr_f if female==1, p
gen psi_fem = ((1-pr_f)/pr_f) if female==1

******************************************************************************
/* (1) determine the weight associated with covariates of men given all */
******************************************************************************
xi: logit female i.occ i.educ exp urban i.reg
predict pr_f_all if female==1, p
gen psi_all = ((1-pr_f_all)/pr_f_all) if female==1

******************************************************************************
/* (2) occ given i.educ exp urban i.reg */
******************************************************************************
set more off
xi: mlogit occ i.educ i.expc urban i.reg
predict pr_f_occ1 pr_f_occ2 pr_f_occ3 pr_f_occ4 pr_f_occ5 ///
pr_f_occ6 pr_f_occ7 pr_f_occ8 pr_f_occ9 if female==1,p

gen pr_f_occ = pr_f_occ1 if female==1 & occ==1
replace pr_f_occ = pr_f_occ2 if female==1 & occ==2
replace pr_f_occ = pr_f_occ3 if female==1 & occ==3
replace pr_f_occ = pr_f_occ4 if female==1 & occ==4
replace pr_f_occ = pr_f_occ5 if female==1 & occ==5
replace pr_f_occ = pr_f_occ6 if female==1 & occ==6
replace pr_f_occ = pr_f_occ7 if female==1 & occ==7
replace pr_f_occ = pr_f_occ8 if female==1 & occ==8
replace pr_f_occ = pr_f_occ9 if female==1 & occ==9

xi: mlogit occ i.educ i.expc urban i.reg
predict pr_f_occ1 pr_f_occ2 pr_f_occ3 pr_f_occ4 pr_f_occ5 ///
pr_f_occ6 pr_f_occ7 pr_f_occ8 pr_f_occ9 if female==1,p

gen pr_f_occ = pr_f_occ1 if female==1 & occ==1
replace pr_f_occ = pr_f_occ2 if female==1 & occ==2
replace pr_f_occ = pr_f_occ3 if female==1 & occ==3
replace pr_f_occ = pr_f_occ4 if female==1 & occ==4
replace pr_f_occ = pr_f_occ5 if female==1 & occ==5
replace pr_f_occ = pr_f_occ6 if female==1 & occ==6
replace pr_f_occ = pr_f_occ7 if female==1 & occ==7
replace pr_f_occ = pr_f_occ8 if female==1 & occ==8
replace pr_f_occ = pr_f_occ9 if female==1 & occ==9

gen psi_occ = pr_m_occ/pr_f_occ if female==1

/*************************************************************/
/* (3) educ (education) given exp urban i.reg ***********/
/*************************************************************/
xi: mlogit educ i.expc urban i.reg if female==0
predict pr_m_edu1 pr_m_edu2 pr_m_edu3 pr_m_edu4 pr_m_edu5 ///
pr_m_edu6 pr_m_edu7 if female==1,p

gen pr_m_edu = pr_m_edu1 if female==1 & educ==1
replace pr_m_edu = pr_m_edu2 if female==1 & educ==2
replace pr_m_edu = pr_m_edu3 if female==1 & educ==3
replace pr_m_edu = pr_m_edu4 if female==1 & educ==4
replace pr_m_edu = pr_m_edu5 if female==1 & educ==5
replace pr_m_edu = pr_m_edu6 if female==1 & educ==6
replace pr_m_edu = pr_m_edu7 if female==1 & educ==7

xi: mlogit educ i.expc urban i.reg if female==1
predict pr_f_edu1 pr_f_edu2 pr_f_edu3 pr_f_edu4 pr_f_edu5 ///
pr_f_edu6 pr_f_edu7 if female==1,p

gen pr_f_edu = pr_f_edu1 if female==1 & educ==1
replace pr_f_edu = pr_f_edu2 if female==1 & educ==2
replace pr_f_edu = pr_f_edu3 if female==1 & educ==3
replace pr_f_edu = pr_f_edu4 if female==1 & educ==4
replace pr_f_edu = pr_f_edu5 if female==1 & educ==5
replace pr_f_edu = pr_f_edu6 if female==1 & educ==6
replace pr_f_edu = pr_f_edu7 if female==1 & educ==7

gen psi_edu=pr_m_edu/pr_f_edu if female==1

/*************************************************************/
/* (4) exp (experience) given urban i.reg ***********/
/*************************************************************/
xi: mlogit expc urban i.reg if female==0
predict pr_m_exp1 pr_m_exp2 pr_m_exp3 pr_m_exp4 pr_m_exp5 ///
pr_m_exp6 pr_m_exp7 if female==1,p

gen pr_m_exp = pr_m_exp1 if female==1 & expc==1
replace pr_m_exp = pr_m_exp2 if female==1 & expc==2
replace pr_m_exp = pr_m_exp3 if female==1 & expc==3
replace pr_m_exp = pr_m_exp4 if female==1 & expc==4
replace pr_m_exp = pr_m_exp5 if female==1 & expc==5
replace pr_m_exp = pr_m_exp6 if female==1 & expc==6
replace pr_m_exp = pr_m_exp7 if female==1 & expc==7

57
xi: mlogit expc urban i.reg if female==1
predict pr_f_exp1 pr_f_exp2 pr_f_exp3 pr_f_exp4 pr_f_exp5 pr_f_exp6 pr_f_exp7 if female==1,p

gen    pr_f_exp = pr_f_exp1 if female==1 & expc==1
replace pr_f_exp = pr_f_exp2 if female==1 & expc==2
replace pr_f_exp = pr_f_exp3 if female==1 & expc==3
replace pr_f_exp = pr_f_exp4 if female==1 & expc==4
replace pr_f_exp = pr_f_exp5 if female==1 & expc==5
replace pr_f_exp = pr_f_exp6 if female==1 & expc==6
replace pr_f_exp = pr_f_exp7 if female==1 & expc==7

gen psi_exp=pr_m_exp/pr_f_exp if female==1

/***************************************************************/
/* (5) urb (urban) given reg                                  */
/***************************************************************/
xi: mlogit urban i.reg if female==0
predict pr_m_urb0 pr_m_urb1 if female==1,p

gen    pr_m_urb = pr_m_urb0 if female==1 & urban==0
replace pr_m_urb = pr_m_urb1 if female==1 & urban==1

xi: mlogit urban i.reg if female==1
predict pr_f_urb0 pr_f_urb1 if female==1,p

gen    pr_f_urb = pr_f_urb0 if female==1 & urban==0
replace pr_f_urb = pr_f_urb1 if female==1 & urban==1

gen psi_urb=pr_m_urb/pr_f_urb if female==1

/***************************************************************/
/* (6) reg (region)                                           */
/***************************************************************/
xi: mlogit reg if female==0
predict pr_m_reg1 pr_m_reg2 pr_m_reg3 pr_m_reg4 pr_m_reg5 if female==1,p

gen    pr_m_reg = pr_m_reg1 if female==1 & reg==1
replace pr_m_reg = pr_m_reg2 if female==1 & reg==2
replace pr_m_reg = pr_m_reg3 if female==1 & reg==3
replace pr_m_reg = pr_m_reg4 if female==1 & reg==4
replace pr_m_reg = pr_m_reg5 if female==1 & reg==5

xi: mlogit reg if female==1
predict pr_f_reg1 pr_f_reg2 pr_f_reg3 pr_f_reg4 pr_f_reg5 if female==1,p
gen pr_f_reg = pr_f_reg1 if female==1 & reg==1
replace pr_f_reg = pr_f_reg2 if female==1 & reg==2
replace pr_f_reg = pr_f_reg3 if female==1 & reg==3
replace pr_f_reg = pr_f_reg4 if female==1 & reg==4
replace pr_f_reg = pr_f_reg5 if female==1 & reg==5

gen psi_reg=pr_m_reg/pr_f_reg if female==1

/*****************************/
/* III. Generate psi */
/*****************************/

/*Actual Weight*/
gen psia00=hweight*pr_f /*same as aweight=hweight*/
gen psia0= hweight*pr_f_all
gen psial= hweight*pr_f_occ
gen psia2= hweight*pr_f_occ*pr_f_edu
gen psia3= hweight*pr_f_occ*pr_f_edu*pr_f_exp
gen psia4= hweight*pr_f_occ*pr_f_edu*pr_f_exp*pr_f_urb
gen psia5= hweight*pr_f_occ*pr_f_edu*pr_f_exp*pr_f_urb*pr_f_reg

/*Counter Weight*/
gen psi00=hweight*psi_fem
gen psi0 =hweight*psi_all
gen psi1 =hweight*psi_occ
gen psi2 =hweight*psi_occ*psi_edu
gen psi3 =hweight*psi_occ*psi_edu*psi_exp
gen psi4 =hweight*psi_occ*psi_edu*psi_exp*psi_urb
gen psi5 =hweight*psi_occ*psi_edu*psi_exp*psi_urb*psi_reg

/*****************************/
/* IV. Determin Kernal Density***/
/****Actual Densities************/  
/*****************************/
pctile newx = lwage, nq(300)
kdensity lwage [aweight=hweight] if female==0, ///
at(newx) w(0.065) gen(m00b d_xm00b) nograph /**male**/

ekdensity lwage [aweight=psia00] if female==1, ///
at(newx) w(0.065) gen(f00b d_xf00b) nograph /**female**/
kdensity lwage [aweight=psia0] if female==1, ///
at(newx) w(0.065) gen(f0b d_xf0b) nograph /**all**/
kdensity lwage [aweight=psial] if female==1 , ///
at(newx) w(0.065) gen(f1b d_xf1b) nograph /**Occupation**/
kdensity lwage [aweight=psia2] if female==1 , ///
at(newx) w(0.065) gen(f2b d_xf2b) nograph /**education**/
kdensity lwage [aweight=psia3] if female==1 , ///
at(newx) w(0.065) gen(f3b d_xf3b) nograph /**experience**/
kdensity lwage [aweight=psia4] if female==1 , ///
at(newx) w(0.065) gen(f4b d_xf4b) nograph /**urban**/
kdensity lwage [aweight=psia5] if female==1 , ///
at(newx) w(0.065) gen(f5b d_xf5b) nograph /**region**/
/*Counterfactual Densities*/
/***********************************************************/
/*female distribution of wages with distribution of covariates of men*/
kdensity lwage [aweight=psi00] if female==1 , ///
at(newx) w(0.065) gen(f00a d_xf00a) nograph /**female--male same as female**/
kdensity lwage [aweight=psi0] if female==1 , ///
at(newx) w(0.065) gen(f0a d_xf0a) nograph /**all**/
kdensity lwage [aweight=psi1] if female==1 , ///
at(newx) w(0.065) gen(f1a d_xf1a) nograph /**occupation**/
kdensity lwage [aweight=psi2] if female==1 , ///
at(newx) w(0.065) gen(f2a d_xf2a) nograph /**education**/
kdensity lwage [aweight=psi3] if female==1 , ///
at(newx) w(0.065) gen(f3a d_xf3a) nograph /**experience**/
kdensity lwage [aweight=psi4] if female==1 , ///
at(newx) w(0.065) gen(f4a d_xf4a) nograph /**urban**/
kdensity lwage [aweight=psi5] if female==1 , ///
at(newx) w(0.065) gen(f5a d_xf5a) nograph /**region**/
/***********************************************************/
/* VII. GRAPH wage gap**/
/***********************************************************/
/* 1. Generate wage gap***** /
/* 1a. Male actual vs. Female actual***** /
gen x=_n
pctile m_a = lwage if female==0, nquantiles(100)
pctile f_a = lwage if female==1, nquantiles(100)
gen w_gap_2006=f_a-m_a
/** 1b. ALL**/
pctile all_fa = lwage[w=psi0] if female==1, nquantiles(100)
gen w_all_a=all_fa-m_a
/** 1c. Occupation**/
pctile occ_fa = lwage[w=psi1] if female==1, nquantiles(100)
gen w_occ_a=occ_fa-m_a
/** 1d. Education**/
pctile edu_fa = lwage[w=psi2] if female==1, nquantiles(100)
gen w_edu_a=edu_fa-m_a
/** 1e. Experience**/
pctile exp_fa = lwage[w=psi3] if female==1, nquantiles(100)
gen w_exp_a=exp_fa-m_a
/** 1f. Urban**/
pctile urb_fa = lwage[w=psi4] if female==1, nquantiles(100)
gen w_urb_a=urb_fa-m_a
/** 1g. Region**/
pctile reg_fa = lwage[w=psi5] if female==1, nquantiles(100)
gen w_reg_a = reg_fa - m_a

*list x w_gap w_all_a w_occ_a w_est_a w_reg_a if x<20, separator(0)

/*/ 2. Graph wage gap**/
/**2a. Male actual vs. Female actual**/

/* Graph wage gap**/
graph twoway ///
   (line  w_gap$qrt  x,legend(off) xtitle("Percentile") ///
     ytitle("Log hourly Wage Gap") ///
     subtitle("Thai Gender Log hourly Wage Gap by Percentile") ///
     xmtick(##10) ylabel(-.45(.05)0) xlabel(0(10)100)clcolor(blue) ///
     clwidth(medium) note("Source: LFS, ${year}.") ///
     "Thai log hourly wage gap by percentile.") ///
   msymbol(i) clwidth(medium)) if x<100, ///
yline(0) scheme(s1color) name(gw0a,replace)

/**2i. Adjusted Wage Gap-- 1997 vs. 2006**/

graph twoway ///
   (line w_gap$1997 x,legend(label(1 "1997") ring(0) pos(8) ///
     rows(2) size(vsmall)symxsize(3)) xtitle("") ytitle("") ///
     yline(0, lcolor(black)) subtitle("") clcolor(blue) ///
     clwidth(medium) xmtick(##10)) ///
   (line w_gap$2006 x,legend(label(2 "2006") ring(0) pos(8) ///
     rows(2) size(vsmall)symxsize(3)) xtitle("Percentile") ytitle("Log 
hourly Wage Gap") ///
     yline(0, lcolor(black)) subtitle("Log Wage Gap 1997 and 2006")
     clcolor(red) clwidth(medium) clpattern(dash) ///
     note("Source: LFS, 1997 and 2006." "Thai log hourly wage gap by 
percentile.") ///
     xmtick(##10) ylabel(-.45(.05)0) xlabel(0(10)100)clcolor(red) ///
     clwidth(medium) xmtick(##10)) ///
   if x<100, scheme(s1color) name(gw8a,replace)

STATA code for Figure 2 and 3. 2006 Counterfactual Wage Gap DiNardo, Fortin, Lemieux Decomposition Results

/**2b. Adjusted Wage Gap-- Act vs. All**/

graph twoway ///
   (line w_gap$qrt x,legend(label(1 "Before Adjustment") ring(0) ///
     pos(8) ///
     rows(2) size(vsmall)symxsize(3)) xtitle("") ytitle("") ///
     yline(0, lcolor(black)) subtitle("") clcolor(blue) ///
     clwidth(medium) xmtick(##10)) ///
   (line w_all_a x,legend(label(2 "After Adjustment") ring(0) ///
     pos(8) ///
     rows(2) size(vsmall)symxsize(3)) xtitle("Percentile") ytitle("Log 
hourly Wage Gap") ///
     yline(0, lcolor(black)) subtitle("(A) all") clcolor(red)
     clwidth(medium) clpattern(dash) ///
STATA code for Figure 4. Probability Distribution of Log Hourly Earnings. Panel (A) 2006 Actual Wage Density for Thai Female and Male Workers

/*1. Male actual vs. Female actual*/
/* d_xf0b=d_xf00b=d_xf00a are the same */
graph twoway
   (line d_xf00b newx, legend(label(1 "Female Actual") ring(0))
   pos(1) ///
   rows(2) size(vsmall) symsize(3) clcolor(blue) clwidth(thick))
   ///
   (line d_xm00b newx, legend(label(2 "Male Actual") ring(0))
   pos(1) ///
   rows(2) size(vsmall) symsize(3) clcolor(red) clwidth(medium) clpattern(dash)), ///
   xtitle("Log Hourly Wage (${year})") ytitle("Density") subtitle("") ///
scheme(s1color) saving(g00,replace)

STATA code for Figure 5 and 6. Panel (A-G). Counterfactual Probability Distribution of Log Hourly Earnings. 2006 Female Counterfactual Wage Density

///**Before and After adjustment of Densities**///
/*1. Male actual vs. Female actual*/
graph twoway
   (line d_xf00b newx, legend(label(1 "Female Actual") ring(0))
   pos(1) ///
   rows(2) size(vsmall) symsize(3) clcolor(blue) clwidth(thick))
   ///
   (line d_xm00b newx, legend(label(2 "Male Actual") ring(0))
   pos(1) ///
   rows(2) size(vsmall) symsize(3) clcolor(red) clwidth(medium) clpattern(dash)), ///
   xtitle("Log Hourly Wage (${year})") ytitle("Density") subtitle("") ///
scheme(s1color) saving(g00,replace)
/*All: Female actual vs. Female char male all*/
graph twoway ///
(line d_xf00b newx, legend(label(1 "Female Actual") ring(0)
pos(1) ///
rows(2) size(vsmall) symxsize(3) clcolor(blue) clwidth(thick)) ///
(line d_xf0a newx, legend(label(2 "Female WT All") ring(0)
pos(1) ///
rows(2) size(vsmall) symxsize(3) clcolor(red) clwidth(medium)
clpattern(dash)), ///
xtitle("Log Hourly Wage (${year})") ytitle("Density")
subtitle("") ///
scheme(s1color) saving(g0,replace)

/*Occ: Female actual vs. Female char male occ*/
graph twoway ///
(line d_xf00b newx, legend(label(1 "Female Actual") ring(0)
pos(1) ///
rows(2) size(vsmall) symxsize(3) clcolor(blue) clwidth(thick)) ///
(line d_xf1a newx, legend(label(2 "Female WT +Occ") ring(0)
pos(1) ///
rows(2) size(vsmall) symxsize(3) clcolor(red) clwidth(medium)
clpattern(dash)), ///
xtitle("Log Hourly Wage (${year})") ytitle("Density")
subtitle("") ///
scheme(s1color) saving(g1,replace)

/*Est: Female WT Occ vs. Female char male Edu*/
graph twoway ///
(line d_xf1a newx, legend(label(1 "Female WT +Occ") ring(0)
pos(1) ///
rows(2) size(vsmall) symxsize(3) clcolor(blue) clwidth(thick)) ///
(line d_xf2a newx, legend(label(2 "Female WT +Edu") ring(0)
pos(1) ///
rows(2) size(vsmall) symxsize(3) clcolor(red) clwidth(medium)
clpattern(dash)), ///
xtitle("Log Hourly Wage (${year})") ytitle("Density")
subtitle("") ///
scheme(s1color) saving(g2,replace)

/*Edu: Female WT Edu vs. Female char male Exp*/
graph twoway ///
(line d_xf2a newx, legend(label(1 "Female WT +Edu") ring(0)
pos(1) ///
rows(2) size(vsmall) symxsize(3) clcolor(blue) clwidth(thick)) ///
(line d_xf3a newx, legend(label(2 "Female WT +Exp") ring(0)
pos(1) ///
rows(2) size(vsmall) symxsize(3) clcolor(red) clwidth(medium)
clpattern(dash)), ///
xtile("Log Hourly Wage ($\{year\})") ytitle("Density") subtitle("") ///
scheme(s1color) saving(g3,replace)

/*Age: Female WT Exp vs. Female char male Urb*/
graph twoway ///
(line d_xf3a newx ,legend(label(1 "Female WT +Exp") ring(0)
pos(1) ///
    rows(2) size(vsmall) symxsize(3)) clcolor(blue) clwidth(thick))
///
(line d_xf4a newx ,legend(label(2 "Female WT +Urb") ring(0)
pos(1) ///
    rows(2) size(vsmall) symxsize(3)) clcolor(red) clwidth(medium)
clpattern(dash)), ///
xtile("Log Hourly Wage ($\{year\})") ytitle("Density")
scheme(s1color) saving(g4,replace)

/*Reg: Female WT Urb vs. Female char male Reg*/
graph twoway ///
(line d_xf4a newx ,legend(label(1 "Female WT +Urb") ring(0)
pos(1) ///
    rows(2) size(vsmall) symxsize(3)) clcolor(blue) clwidth(thick))
///
(line d_xf5a newx ,legend(label(2 "Female WT +Reg") ring(0)
pos(1) ///
    rows(2) size(vsmall) symxsize(3)) clcolor(red) clwidth(medium)
clpattern(dash)), ///
xtile("Log Hourly Wage ($\{year\})") ytitle("Density")
scheme(s1color) saving(g5,replace)

/*Act vs. All*/
graph twoway ///
(line d_xm00b newx ,legend(label(1 "A_male") ring(0) pos(1) ///
    rows(2) size(vsmall) symxsize(3)) clcolor(blue) clwidth(thick))
///
(line d_xf00b newx ,legend(label(2 "A_female") ring(0) pos(1) ///
    rows(2) size(vsmall) symxsize(3)) clcolor(pink) clwidth(thick))
///
(line d_xf0a newx ,legend(label(3 "WT all0") ring(0) pos(1) ///
    rows(2) size(vsmall) symxsize(3)) clcolor(green) clwidth(medium)
clpattern(dash)) ///
(line d_xf5a newx ,legend(label(4 "WT all5") ring(0) pos(1) ///
    rows(2) size(vsmall) symxsize(3)) clcolor(red) clwidth(medium)
clpattern(dash)), ///
xtitle("Log Hourly Wage ()") ytitle("Density") subtitle("") ///
scheme(s1color) saving(temp,replace)
Chapter 3

The Determinants and Effects of Job Training on U.S. Immigrant Workers

with Hung-lin Chen

3.1 Introduction

In recent years, there have been a growing number of studies on job training and program evaluation in the United States. Despite the large and steadily increasing population of immigrants working in the United States, very few training studies focus on immigrants. ‘Immigrants make up one in nine U.S. residents, one in seven U.S. workers, and one in five low-wage workers’ (Capps et al., 2003). Since the year 2000, legal immigrants that came to the U.S. have numbered approximately one million per year (Borjas, 2008). Surprisingly, only modest knowledge of the effect of training programs on minorities in the labor market is known, and the evidence of these programs on immigrants as a whole is even more limited (Ashenfelter 1983, Capps et al., 2003, Flores-Lagunes et al., 2007). This paper aims to improve our understanding of the effect of job training on immigrant workers.

The main incentives of job training provided by employers is not only to increase worker’s productivity, but also to minimize the adoption time when new technology arrives. These improvements are in line with business goals, including implementation in order to reduce
costs of job turnover. It is widely known in the labor literature that firms which provide job
training have more productive workers and less expenses associated with turnover (Frazis et
al., 1998). These firms can earn more profit from the increase of workers’ productivity. Thus,
it is likely that companies would want to invest in their employees, particularly immigrants
who tend to possess fewer skills, limited English proficiency and less formal education;
shortage, they are opportunities for productivity gains. Immigrants also need training to help
them adapt to changing work environments. However, our results suggest that immigrant
workers receive considerably less training than native workers—21.6 percent and 39.1 percent
respectively.

Job training is an important instrument that can be used to assist both native and
immigrant workers improve their welfare. Most immigrants, like other low-skill workers, lack
opportunities to learn new skills and to benefit from employer-provided training programs
(Ahlstrand et al., 2001). In the private sector, it is often the case that not all employees
receive equal training opportunities. More specifically, employers prefer to provide job
training to workers who are more likely to stay with the firms for a relatively long period of
time. With evidences from the US and Canada showing that workers are not likely to pay
for their own training (Parent, 1999), immigrants with greatest need of general training are
often unable or unwilling to invest in training themselves. Hence, these immigrants, who
are low-skilled and are perceived to have short-term tenure, are in need of general training,
and yet they end up losing out on opportunities to be trained.

In the case of government sponsored training, these programs often have stringent re-
requirements that reduce immigrants’ chances of attaining job training. Some studies have
found that low-wage immigrants are often under-served due to the lack of proficiency in En-
glish (Tumin and Zimmermann, 2003). The main purpose of government training programs
is to prepare economically disadvantaged individuals, such as welfare program recipients for
the labor market, yet it is common to find certain minimum requirements that prevent
low-skilled workers from obtaining these training opportunities (Capps et al., 2003). For
instance, it is frequently observed that programs require participants to have at least ninth-
grade level literacy, and they should have numeracy ability and basic English skills.

According to the March 2002 Current Population Survey (CPS), 18 percent of all immigrant workers and 28 percent of all low-wage immigrant workers have educational attainment less than ninth grade, while 1 percent of all native workers and 2 percent of all low-wage native workers have educational attainment less than ninth grade (Capps et al., 2003). Thus, by not meeting the minimum requirements, many immigrants lose out on the opportunity to be trained by the programs that are sponsored by government.

Immigrant workers are important assets in the U.S. labor market, representing 14 percent of the total U.S. labor force and 20 percent of low-wage earning workers (Capps et al., 2003). Immigrant workers represent a large share of low-skilled workers, who engage in low paying jobs that are necessary for our economy. Another important fact is that nearly half of immigrants earn less than two times the minimum wage, compared to less than one-third for native workers (Capps et al. 2003).

Policy makers also might be interested in the labor market outcomes of immigrants besides earnings. Immigrants face not only low wage, but also problems of low education, limited English proficiency and lack of formal training. At the same time, the immigrant workforce is also confronted by basic problems such as inadequate healthcare, transportation and childcare (Edid, 2007). A better earning capability of immigrants would likely result in a reduction of social problems such as unemployment, gangs and crimes activities. Non-experimental data have shown that increased earnings reduce criminal activity (Lalonde, 1986) and lower the number of crimes such as murders (Donohue and Siegelman, 1998).

To summarize, we know that training is important to immigrants. Yet, immigrants, as a minority group who is in need of training, are receiving much lesser training than natives. Hence, to help policy makers improve their decisions regarding training provided for immigrant workers, we need to gain better understanding of the effect of training on immigrants.

The contribution of this paper is that it is one of the few papers that inquires into the effect of training on immigrants, and it is the first paper that proposes an economic
model that analyzes the effect of job training on immigrant workers in the United States. This is the first paper that looks at both the mean and distributional effect of job training on immigrant workers in the United States. It is the first paper that studies the effect of training on immigrant workers using the 2004, 2001 and 1996 Survey of Income and Program Participation (SIPP) data, and we do this using a Quantile Regression (QREQ) model, a semi-parametric reweighting DiNardo, Fortin and Lemieux (DFL) method, and propensity score matching method. Another contribution is that this is one of the few studies to explore the effect of training on native born workers excluding foreign born workers.

This paper: (1) compares the average impact of job training on earnings of native and immigrant workers; (2) explores the distributional impacts of job training on native and immigrant workers’ earnings and (3) examines the counterfactual distributional impact of job training if trained and untrained workers have similar observable characteristics.

3.1.1 Issues on Training

In general, there are two main problems when measuring the impact of training: unobserved heterogeneity and selection. First, an unobservable heterogeneity problem occurs when the model suffers from omitted variables bias. In our study, the omitted variables come from the difficulty of measuring characteristics such as intelligence, motivation and obtaining the opportunity cost of participating in training.

Since the focus of our paper is to study the comparative effect of training on natives and immigrants, obtaining the exact estimate is not our biggest problem. There might exist upward bias in the estimated training coefficients, yet it will not change our result. Since, we are looking at the result of the comparison which is the differences in return to training between natives and immigrants.

Since this paper is the first paper on the topic, we believe it is more important to layout the fundamental results rather than getting into a sophisticated econometric model.

\footnote{Like other studies on training, we acknowledge that we will not be able to resolve all unobserved heterogeneity problems.}
Therefore, the other issue of training evaluation such as selection problem is beyond the scope of this paper. Yet, we will discuss how to overcome some of the selection problems. Also, we plan to tackle the selection issue in our future work.

Lastly, there is an issue of the importance of the distributional effects of training. It is known that the effects of training often vary across the wage distribution that are not captured by the mean; examples of these studies include Lalonde (1986) or Abadie et al. (2002). The distributional outcomes beyond simple averages are of fundamental interest in the policy analysis of welfare implications such as transfer, education and training programs (Abadie et al., 2002). The results from the distributional analysis highlight the important differences for low-wage and low-skill workers.

Given the distinct characteristics of natives and immigrants, we might expect differences in the role and effect of job training on the labor market. First, we know that immigrants have less working experience in the US than natives since they have shorter residency. Also, immigrants are likely to have lower education and English proficiency. Since immigrants possess lower human capital accumulation that is difficult to quantify in the wage equation, we expect higher return to job training for native workers.

Next, we know that certain groups of immigrants may not plan to be in the US for a long period of time. Cortes (2004) examined the differences in earning between the economic immigrants and refugees. Looking at the investment in human capital accumulation, she found that the economic immigrants may opt to return to their birth country after accumulated wealth. In general, due to instability of these immigrants, firms are less likely to invest in specific job training resulting in less training for immigrant workers. These immigrants may also be less willing to invest in job training because they will not remain with a firm for a long period of time. Immigrants that opt for training may also put in less effort. Hence, even for immigrants that are willing to obtain training, the amount of training and the rate of return to training will be different.

Assuming the output of job training is a function of human capital accumulation and inputs (the effort and the hours of job training from the firm) we expect a lower return
to job training for immigrant workers. Furthermore, it is likely that the rate of return to job training does not increase linearly with the amount of accumulated job training. Since immigrants are endowed with lower human capital accumulation and may put in less effort, we expect that the rate of return for immigrants will be relatively lower at lower level and much lower at the high level than that of natives.

The paper is structured as follows. Section 3.2 reviews the literature on job training. Section 3.4 describes the data used in our analysis. Section 3.3 explores the methods that will be applied. Section 3.5 discusses our analysis and findings. Section 3.6 summarizes our findings and comments on policy implications.

3.2 Literature Review

Despite the fact that there are many studies exploring the different aspects of immigrant workers in the United States and numerous studies measuring the effect of training individually, we know of only a few studies that inquire into the impact of training on immigrants. We suspect that the reasons for the small number of studies could be twofold: because of poor data sources and because there are very few training programs that target immigrants specifically. Few existing studies that inquire into the impact of training on immigrants present only basic results such as composition tables and graphs. For instance, Capps et al. (2003) compare the profile of low-wage immigrant workers and native workers using composition tables and graphs on CPS data. They argue in favor of revamping training requirements to increase access to training for immigrant workers. Another source such as Edid (2007) evaluates the training needs of immigrant workers in Syracuse by using interviews and composition graphs on Census data. She supports the improvement of English proficiency among among other skills needed for immigrant workers.

Fortinand and Parent (2005) explore the incidence of training in US and Canada. Although the paper does not focus on immigrants, the authors note that immigrant status is significant in the determinants of training. They also found that immigrants in the U.S.
are more likely to have shorter training than those in Canada. Though these papers did not evaluate the effect of training on immigrants, they provide the cornerstone question for us to explore further. We will improve upon their studies by setting up our training evaluation model, comparing the differences in the returns to training for immigrant and native workers, doing so using econometric tools. In the rest of our literature review, we will discuss the general training aspects and the relevant econometric evaluation methods.

Most studies on the effect of job training are from the United States and Europe. The majority of the studies are empirical works. They focus on questions such as (1) do participants benefit from job training, (2) is there social merit in job training or (3) what are the determinants of job training? The general consensus is that government sponsored job training programs are ineffective, resulting in a small positive or even negative net benefits, yet with great heterogeneity (Heckman et al., 1999). Furthermore, it appears that there is no consensus on the model to measure the effect of job training program. Due to the inconsistencies of measurement findings, there is an increase in econometric methodological studies in recent years. Generally, the results from experimental studies yield impact of job training programs on earnings that range from minus 15 to plus 70 percent (Heckman et al., 1999).

The focus of our paper is on non-experimental data, focusing on the impact of training of disadvantage workers in the United States. Notable studies include economically disadvantaged adult participants (Ashenfelter and Card, 1985; Bassi, 1983; Dickinson et al., 1986), displaced workers (Bloom, 1990; Decker and Corson, 1995) and economically disadvantaged youth (Dickinson et al., 1986; Bryant and Rupp, 1987; Bassi et al., 1984).

Besides the studies that focus on displaced workers and youths, there are many papers that explore the effect of training on specific minority groups. For example, there are papers that focus on the effect of training on blacks (Butler and Heckman, 1977; Smith and Welch).
and the effect of training on Hispanic workers (Schochet et al., 2001; Flores-Lagunes et al., 2006, 2007). These literatures lead us to believe that there are heterogeneous treatment effects in these special groups, since women, disabled, youth and other minority groups have different treatment effects. Although these papers above discuss the effect of training on different races, the studies on immigrants, which is a diverse minority group that has different characteristics than natives, have not yet been fully investigated as we mentioned before. Below, we discuss the literature of the most relevant approaches to this paper, studying the average effect and the distributional effect.

### 3.2.1 Study of Average Treatment Effect on Treated

For the average effect of training on workers in the US, the conventional methods use RE or FE models on panel data. Panel data can be used to control for unobserved omitted variables. Furthermore, recent studies estimate the effect of job training using a propensity score matching (PSM) method, first proposed by Rosenbaum and Rubin (1983). It assumes that the distributional outcome of the treated is not statistically different from the distributional outcome of the untreated. PSM resolves selection on combinations and interactions of observable characteristics, but does little to address bias due to unobserved heterogeneity.

There have been many studies using PSM, including the main contribution to the training literature by Dehejia and Wahba (1999, 2002), Heckman et al. (1998b), and Heckman et al. (1997, 1998a). Dehejia and Wahba (1999), using Lalonde (1986) study of NationalSupported Work (NSW), CPS and PSID data, found a reduction in the treated coefficient estimates of experimental data when using PSM. In addition, Dehejia and Wahba (2002) further contribute to the literature by using several matching methods such as nearest neighbor and radius (caliper). Other main extensions of original PSM include the kernel and stratification matching methods. When choosing among the existing matching methods,
the general consensus is that there is no preferred matching method (Becker and Ichino, 2002). Dehejia and Wahba (1999, 2002) concluded that "propensity score matching methods provide a natural weighting scheme that yields unbiased estimates of the treatment impact for nonexperimental approaches."

### 3.2.2 Study of the Distribution Effect

The effect of job training on the income distribution has become of great interest in recent years. This is due to the importance of the distributional effect of training, which is not captured by the mean impact. The most predominant method is applying the basic framework of quantile regression that was developed by Koenker and Gilbert Bassett (1978). QREQ assumes conditional treatment and no selection problem. The latest distribution studies focus on resolving the treatment (selection problems) on both conditional and unconditional effects, where exogenous treatment choices assume selection on observable and endogenous treatment choices assume selection on unobservable (Frolich and Melly, 2008).

Some recent distributional studies focus on conditional treatment with selection on observables only. Quantile treatment effect (QTE), proposed by Abadie et al. (2002), who study the effect of the JTPA training program using IV estimator on conditional quantiles in order to deal with bias due to unobservable. Using "indicators of the randomized offered training as binary instrument variable" in QTE, they found the largest impact of job training at low quantiles for women and the only positive impact in the upper half distribution of men with JTPA data. As a benchmark, treatment 2SLS estimates a 15 percent increase in earning for women and a 9 percent increase in earning for men (Orr et al., 1996). Unfortunately, our non-experimental data does not possess the IV that was used in Abadie et al. (2002). Our paper extends the current literature by analyzing the effect of training on immigrants in the United States using quantile regression, reweighting methodology, and propensity score matching method.

---

5 Similar to our average effect study, we acknowledge that we will not be able to resolve all unobserved heterogeneity problems in our distributional effect study.
3.3 Empirical Model

3.3.1 A Model of Quantile Regression

In this section, we examine the estimator of the Koenker and Gilbert Bassett (1978) quantile regression model. The quantile regression model has outcome variable, $Y$, binary treatment indicator, $D$, and a vector of covariates, $X$. In our empirical study of job training, $Y$ is workers’ earnings and $D$ is an indicator of exposure to job training. $X$ indicates the observable characteristics of the workers (occupation sorting, demographic differences and human capital differences). For $n$ observations, individual workers’ outcomes can be expressed as follows:

\[
Y_i = Y_i^1 D_i + Y_i^0 (1 - D_i)
\]  

(3.1)

where $Y_i^1$ is the indicator of potential outcome if workers received treatment (potential earning if workers received training) and $Y_i^0$ is indicator of potential outcome if workers did not receive treatment (potential earning if workers did not receive training) for the entire wage distribution function. Quantile regression model has the following basic assumptions Frolich and Melly (2008):

Assumption (1): Suppose the outcome is a linear function of $X$ and $D$, the outcome can be expressed as follows:

\[
Y_i^d = X_i \beta^d + D \delta^t + \varepsilon_i
\]  

(3.2)

\[
Q_{t \varepsilon} = 0
\]  

(3.3)

where $Q_{t \varepsilon}$ is the $t^{th}$ quantile. We assume $D$ is uncorrelated with error term $\varepsilon$

Assumption (2): Independence:
\((Y^0, Y^1)\) is jointly independent of \(D|X\)

Independence assumption indicates that the potential outcomes are not affected by treatment on unobservable.

In QREQ model, we assume selection is exogenous on observable characteristics, \((Y^0, Y^1)|X\). Hence, we assume that workers’ earnings for both the treated and the non-treated group are not affected by exposure or self selection to job training conditional on observable characteristics. The classical quantile regression can then be computed with the following formula:

\[
(\beta^t, \delta^t) = \arg \min_{\beta, \delta} \sum \rho_t(Y_i - X_i \beta - D_i \delta)
\]

(3.4)

### 3.3.2 Counterfactual Study

In this section, we review the reweighting technique of the DFL model. Using the DFL reweighting technique, we simulate the counterfactual earnings of native (immigrant) workers along the entire distribution, if these workers, who did not receive job training \((D=0)\), have similar observable characteristics as workers that received training \((D=1)\). Hence, we are comparing the earnings of two groups of workers that possess similar observable characteristics, except that one group has training.

Suppose \((W, Z, D)\) is a vector representing each worker, where \(W\) indicates earnings of the workers, \(Z\) is a vector of worker observable attributes (e.g. occupation, firm estimated size, education, age and metropolitan area) and \(D\) is the training indicator \((D=1\) for received training or \(D=0\) for did not receive training). The probability of workers not received training conditioned on workers’ observable characteristics can be estimated using logit or probit model as \(f(D = 0|Z)\). The "reweighting function," \(\Psi(Z)\), is the counterfactual weight of untrained workers that would have prevailed if untrained workers possessed observable characteristics of trained workers:
The reweighting function can simply be calculated by the product of the sample weight and \[p/(1-p)\], where \(p\) is the predicted probability of being untrained workers conditioning on their observable attributes. The intuition here is that we are making a better comparison group of untrained workers that look more similar to trained workers by using the reweighting function that allocates additional weight to the observations that belong to the minority categories. For example, since immigrants with no training have much lower education than immigrants with training, more weight is allocated to the higher education untrained immigrant workers. Finally, the hypothetical quantile training premium is simply the difference between actual earnings of workers with training and counterfactual earnings of workers without training.

3.3.3 Estimating with Propensity Score Matching

Propensity Score

Propensity score matching (PSM) is a method that helps reduce sample selection bias of treatment effect due to significant differences between characteristics of a treatment and a no treatment group (Dehejei and Wahba, 2002). The results from the PSM treatment effect approximate the results from a randomized trial or random experiment, where matching method yield an unbiased estimate of the treatment impact. The PSM method helps reduce the bias when there is not a sufficient overlap in a treatment and the comparison group and when there are differences in the distributions of the observable characteristics for the two groups (Heckman et al., 1998a).

The intuition here is that we are making a better comparison of the two groups that are more homogeneous by comparing a no treatment group that have propensity score similar

\[
\Psi(Z) = \frac{dF(Z|D = 1)}{dF(Z|D = 0)} = \frac{f(Z|D = 1)}{f(Z|D = 0)} = \frac{f(D = 1|Z)/f(D = 1)}{f(D = 0|Z)/f(D = 0)}
\] (3.5)
to treatment group. PSM matches an observation from the treatment group to observations from the comparison group that have similar observable characteristics. The unmatched observations in comparison group are removed so that they are not used in estimating the treatment impact.

There are several limitations of the matching method. First, though PSM helps reduce selection problem, it does not eliminate bias due to unobserved omitted variables. Next, PSM introduces error to the estimation of treatment impact, if the treatment and the comparison group do not have large number of observations overlap. For example, if the worst observations from the treated group are compared to the best observations of the comparison group, we would observe substantial bias. The bias can only be fully eliminated when the treatment group is truly random.

Rosenbaum and Rubin (1983) proposed the propensity score matching methodology in 1983 and defined the propensity score, \( p(X) \), as the conditional probability of receiving a treatment given a set of observed covariates.

\[
p(X) \equiv Pr\{D = 1|X\} = E\{D|X\} \tag{3.6}
\]

where \( D = 1, 0 \) is an indicator for receiving the treatment (job training) or not receiving the treatment (no job training). \( X \) is a set of observed covariates.

For a given propensity score, we can estimate the average treatment effect on the treated (ATT). ATT is the mean effect of treatment on those who receive treatment compared to those who do not receive treatment given the propensity score,

\[
ATT \quad \equiv \quad E\{Y_1 - Y_0|D = 1, X\} \\
= \quad E\{E\{Y_1 - Y_0|D = 1, p(X)\}\} \tag{3.7} \\
= \quad E\{E\{Y_1|D = 1, p(X)\}\} - E\{Y_0|D = 0, p(X)|D = 1\}
\]

where \( Y_1 \) and \( Y_0 \) are log hourly wages (potential outcomes) in the treatment group and
control group, respectively.

For the propensity score matching method, there are two fundamental assumptions:

Assumption 1: For a given propensity score \( p(X) \), the set of observed covariates is balanced. In other words, a set of observed covariates is independent of a training variable with the same propensity score.

\[
D \perp X \mid p(X)
\]  

(3.8)

Assumption 2: Unconfoundedness is given the propensity score:

\[
Y_1, Y_0 \perp D \mid X
\]  

(3.9)

\[
Y_1, Y_0 \perp D \mid p(X)
\]  

(3.10)

Rosenbaum and Rubin (1983) pointed out that "if receiving the treatment is random within cells defined by \( X \), it is also random within cells defined by the values of the monodimensional variable \( p(X) \)". Therefore, the potential outcomes are also independent of training variables conditional upon the same propensity score \( p(X) \).

In sum, if receiving the training is random, treatment and control groups should be identically averaged after giving the propensity score (Chen and Zeiser, 2008). Eren (2007) mentioned that matching is a powerful methodology because it can solve the first two bias problems which are the bias due to a lack of sufficient overlap in the two groups and the bias due to differences in the distributions of the \( X \)s under the common region (Heckman et al., 1998a). Both problems are sometimes found to occur in the OLS models.

Matching with Propensity Score

The two most common matching methods used to estimate ATT, given the propensity scores, are Nearest Neighbor Matching and Kernel Matching.
In Nearest Neighbor Matching, a treatment unit is matched to a control unit with the nearest propensity score. $T$ and $C$ denote the treatment and control sets. $Y^T_i$ and $Y^C_j$ refer to log hourly wages of the treatment and control units. $C(i)$ denotes the set of control units that are matched to the treatment units given the propensity score ($p(X_i)$),

$$C(i) = \min_j \|p(X_i) - p(X_j)\|$$  \hspace{1cm} (3.11)

The average treatment effect on the treated (ATT) is

$$ATT^N = \frac{1}{N_T} \sum_{i \in T} \{Y^T_i - Y^C_j\}$$  \hspace{1cm} (3.12)

where $N_T$ is the number of treated units and $T$ denotes all treated observations.

In Kernel Matching, the outcome of a treated unit is matched to a weighted average of the outcomes of all control units.

$$ATT^K = \frac{1}{N_T} \sum_{i \in T} \left[ Y^T_i - \sum_{j \in C} g_{ij} Y^C_j \right]$$  \hspace{1cm} (3.13)

where $g_{ij}$ is the weight.

According to Becker and Ichino (2002)’s paper, propensity score matching methods only reduce, but do not eliminate, the bias from omitted variables. The bias can only be fully eliminated if receiving the job training is truly random among workers who have the same propensity score. They also point out that there is no best propensity score matching method and they also describe some pitfalls for each matching method. For instance, the nearest neighbor matching method tries to match all treated units to control units with the nearest propensity score. Some of these matches might be poor because the nearest control units might have matches of low quality.
3.3.4 Sensitivity Analysis for Average treatment Effects on the Treated

Since propensity score matching has become increasingly popular to evaluate treatment effects, checking the sensitivity of estimated treatment effects on the treated has become an important topic lately. Researchers are interested in what happens to the estimated results when there are deviations from the underlying identifying conditional independence assumption.

Model

According to Becker and Caliendo (2007), they assume that the participation probability is given by \[ P_i = P(x_i, u_i) = P(D_i = 1, x_i, u_i) = F(\beta x_i + \gamma u_i), \] where \( x_i \) are the observed variables for individual \( i \), \( u_i \) is the unobserved variable, and \( \gamma \) is the effect on the participation decision. If there is no unobserved bias, \( \gamma \) will be zero. The probability of receiving treatment will only be determined by \( x_i \). However, if there is unobserved bias, two individuals with the same observed variable \( x \) have different probability of receiving treatment. They assume that a matched pair of individuals \( i \) and \( j \) and \( F \) is the logistic distribution. The odds that individuals receive treatment are then given by \( P_i/1 - P_i \) and \( P_j/1 - P_j \), and the odds ratio is given by

\[
\frac{P_i}{1-P_i} = \frac{P_j(1-P_j)}{P_j(1-P_i)} = \frac{\exp(\beta x_i + \gamma u_i)}{\exp(\beta x_j + \gamma u_j)}
\] (3.14)

If both individuals have identical observed variables \( (x_i) \), the \( x \) vector cancels out, then the odds ratio becomes

\[
\frac{\exp(\beta x_i + \gamma u_i)}{\exp(\beta x_j + \gamma u_j)} = \exp \{ \gamma (u_i - u_j) \}
\] (3.15)

If there are no differences in unobserved variables \( (u_i = u_j) \), the odds ratio is one which means there is no unobserved selection bias. Likewise, if unobserved variables have no
influence on the probability of receiving treatment ($\gamma = 0$), the odds ratio is also equal to one. Sensitivity analysis now evaluates different $\gamma$ and $u_i - u_j$ to find out how they alter the estimated treatment effects. Becker and Caliendo (2007) follow Aakvik (2001)'s paper and assume that the unobserved covariate is a dummy variable with $u \in \{0, 1\}$. Rosenbaum (2002) shows that (3.14) implies the following bounds on the odds ratio that either of the two matched individuals will receive treatment:

$$\frac{1}{e^\gamma} \leq \frac{P_i(1 - P_j)}{P_j(1 - P_i)} \leq e^\gamma$$

(3.16)

When $e^\gamma = 1$, both matched individuals have the same probability of receiving treatment. Otherwise, if for example $e^\gamma = 2$, individuals who appear to be similar (in terms of $x$) could differ in their odds of receiving the treatment by as much as a factor of 2. Thus, Rosenbaum (2002) determined that $e^\gamma$ is a measure of the degree of departure from a study that is without unobservable bias.

**MH test statistic**

For binary outcomes, Aakvik (2001) suggests using the Mantel and Haenszel (1959) test statistic. The MH nonparametric test compares the matched individuals in the treatment group and control group with the same expected number. According to Becker and Caliendo (2007)'s paper, researchers must make the individuals in the treatment and control groups as similar as possible because this test is based on random sampling. Rosenbaum (2002) shows that the test statistic $Q_{MH}$ can be bounded by two known distributions. If $e^\gamma = 1$ the bounds are equal to the base scenario of no hidden bias. With increasing $e^\gamma$, the bounds move apart, reflecting uncertainty about the test statistics in the presence of unobserved selection bias. Let $Q_{MH}^+$ be the test statistic, given that we have overestimated the treatment effect, and $Q_{MH}^-$, the case where we have underestimated the treatment effect. The two bounds are then given by
\[ Q_{MH}^+ = \frac{Y_1 - \sum_{s=1}^{S} \tilde{E}_s^+ - 0.5}{\sqrt{\sum_{s=1}^{S} \text{Var}(\tilde{E}_s^+)}} \] (3.17)

\[ Q_{MH}^- = \frac{Y_1 - \sum_{s=1}^{S} \tilde{E}_s^- - 0.5}{\sqrt{\sum_{s=1}^{S} \text{Var}(\tilde{E}_s^-)}} \] (3.18)

where \( \tilde{E}_s \) and \( \text{Var}(\tilde{E}_s) \) are the large-sample approximations to the expectation and variance of the number of successful participants when \( u \) is binary and for given \( \gamma \). \( y \) is the outcome for both treated and control groups and \( s \) is stratum.

### 3.4 Data Source

In our empirical study, we utilize data from the Survey of Income and Program Participation (SIPP) in the years 1996, 2001 and 2004. SIPP, funded by US Census Bureau, collects a variety of information, such as income, labor force participation, types of jobs, program participation and demographic data. A main objective of SIPP is to forecast the cost and evaluate the impact of government and other social programs in the United States. Yet, numerous studies on private sector also use SIPP data. The dataset contains unique information on job training and immigrants, as well as income, other human capital and occupational information. SIPP has abundant observations, unlike many other datasets that have small sample size issues.

The SIPP surveys, which are conducted by personal visits and by telephone interviews, were first administered in 1984. The surveys interview approximately 14,000 to 36,700 households of individuals 15 years of age and older, civilian non-institutionalized, conducting monthly questionnaires. The SIPP, currently containing 12 waves for each year surveyed, is a panel dataset, collecting data once every few years. Each wave includes the "Core" (mainly containing wave standard variables that evaluate economic situation in the US) and the 'Topical Modules' (containing different variables depending on the wave). With 12
individual waves for each year that it conducts the survey, SIPP data has more information than other resources, since the survey can ask different sets of questions to interviewees of different waves. Yet, a drawback is that we do not always observe the same person for each set of questions.

The diverse variable availability in SIPP data and its being "rich enough to determine program eligibility" \cite{Heckman2009} benefit our analysis because only a few databases contain enough information on both job training and immigrant status. Yet, another problem of using SIPP is that variables are sometime removed or transferred to different waves of questionnaires, changing from one year to another. Particularly, only the 2004 data have the English ability variable and job training variable in the same wave.

From SIPP data, we use the combination of first (Core) wave and second (Module Two) wave. Since it is well documented that there are sizeable differences between the impact of job training on men versus women and adults versus youths \cite{Heckman2009}, we use a sample that include all the adult males between 22 to 65 years of age\footnote{Even though we realize that lower incidences of training may occur at later age, older individuals are included in the sample to capture possible retraining of recent technical advancement, such as computer skills.}. In addition, we eliminate observations that have hourly earnings of zero dollars or less. The variables that we are most concerned with are wage and job training. The dependent variable wage is the log of workers’ earnings per hour, while the job training variable indicates whether the workers had any job training in the last 10 years\footnote{Unfortunately, we are not able to differentiate between the effect of job training accumulated much earlier or recently and location of training.}.

Even though SIPP is a panel dataset, it is an unbalanced panel. One drawback of this unbalanced panel is that not all individuals are observed every year. Particularly Module 2 of the SIPP which does not contain sufficient number of repeated observations to use RE and FE model.

The observable covariates used in our wage equation includes year dummies, race (White, Black and Hispanic) dummies, seniority dummies, seniority squared divide by 100, years of education categorical variables, marital status dummies, dummies variable indicating...
whether the individual lives in the female headed household, having children younger than 18 living in the family dummies, metropolitan, private firm, firm size dummies, dummy variable denoting possession of health insurance, union dummies, state dummies, industry dummies and occupation dummy. To resolve the concerns relating to changes in some of the variables’ meanings and their categories, we re-categorize 1996 and 2001 data to match the 2004 definition.

3.4.1 Statistics Summary

Table 1 provide the summary statistics of the variables used in our analysis. Specifically, Table 3.1 displays the mean values of total native sample (column 3), native sample which have training (column 6) and native sample which did not have training (column 9). Table 3.2 displays similar statistical summaries for immigrants. The summary tables show that natives and immigrants possess different characteristics. Also, the tables show that workers who have training possess different characteristics than workers who do not have training.

One notable difference is that native workers (at 39.1 percent) received more training in the past ten years than the immigrant workers (at 21.6 percent) on average. Overall, natives earn higher wages than immigrants, while on average, trained immigrants earn marginally higher wages than untrained natives. Thirdly, the average age of natives is almost two years older than immigrants. Since seniority increases the chance of obtaining training, age differences may lead to a small upward bias in the training estimate for natives. Hence, we control for seniority with age variable.

Next, we observe that immigrant workers have accumulated less education than native born workers (26 percent of immigrants have nine years of education or less compared to only 2 percent of natives). Yet, at the higher education level, immigrants and natives have similar education attainment. The education differences in the immigrants may contribute to the heterogeneous effects in the earnings that we will explore in our distributional study.

\footnote{We use age as proxy for seniority.}
In addition, we observe that trained workers have accumulated more education than not trained workers for both natives and immigrants.

We observe that more natives have female heads of households and more natives have health insurance than their immigrant counterparts. Yet, more immigrants are likely to be married, are likely to have more children and are likely to live in metropolitan cities. Generally, being married, having fewer children and living in metropolitan cities increase the probability of obtaining training. Last, more immigrants work in private sector and smaller size firms than natives. Since private sector and small size firms tend to provide less job training, it is possible that job selection of immigrants contribute to immigrants receiving less job training. Consequently, we note that the above covariates are important and need to be controlled for in our wage equation.

3.5 Analysis

This section explores the effect of job training, comparing native and immigrant workers’ wage premium on training. We divide this section into four main subsections: unconditional effect (mean and quantiles), conditional effect (quantile regression), the DFL reweighting counterfactual method and propensity score matching method.

3.5.1 Unconditional Effect

Table 3.3 exhibits results of the unconditional effect of job training on earnings. It shows that though, on average, immigrants earn less than natives, the unconditional training wage premium for immigrants is relatively larger than natives. In terms of monthly earnings, on average, immigrant workers who received training earned 1480 dollars more than immigrant workers who did not receive training, while trained native workers earned 848 dollars more than not trained native workers. The unconditional training wage premium for natives is
17.7 percent, while the unconditional training wage premium for immigrants is 19.7 percent (Table 3.4). Hence, on average, the unconditional wage premium on training for immigrants is 2 percent higher than that for natives.

Figure 3.1 presents the distributional effect. The patterns displayed from the effect of training on the income distribution for immigrant and native workers are quite different from the effect of training at the mean and each other. For the unconditional training premium, the results show that training has largest proportional impact at the upper quantiles for immigrants and largest proportional impact at the middle quantiles for natives. For natives, it shows the differences in results across quantiles where the unconditional wage premium on training increase from lower to median quantiles and slowly decrease from median to upper quantiles. We observe the highest wage gain from training for middle-income workers and the lowest wage gain from training for low-income and upper-income workers. The .50 quantile natives experience earnings increase of around 21.5 percent, while the .05 and .95 quantile natives experience earnings increase of only 12.5 percent.

Unlike natives, the unconditional wage premium on training for immigrants monotonically increases from lower to upper quantiles. Immigrants enjoy highest unconditional wage premium on training at the upper part of the income distribution and receive lowest wage premium at the lower part of the income distribution. The .10 quantile immigrants experience earnings increase of around 8 percent, while the .95 quantile immigrants experience a substantial increase in earnings of 28 percent.

When comparing the training premiums of natives and immigrants across quantiles, it is notable that the unconditional training premium for immigrants is relatively larger than that of natives at the upper part of the income distribution and similar to natives at the lower and middle quantiles (Figure 3.2). While the unconditional training wage premium for natives is more or less constant across the quantiles, the effect of training on immigrants is more remarkable. A possible reason that there are differences in training premium between natives and immigrants at the higher quantiles is that high-skilled immigrants benefit more from training since they started with lower human capital, in terms of language literacy.
For instance, an immigrant doctor will obtain a sizeable income raise once receiving training and English ability is good enough to deal with patients, while an immigrant farmer will earn similar wage increase as a native farmer even after receiving training and English skill is improved.

3.5.2 Conditional Effect

OLS Model

Table 3.5 show the results from pooled OLS controlling for race, age, and education. The estimated training coefficients are positive for the pooled OLS model. The conditional training wage premium for natives is 13.9 percent, while the conditional training wage premium for immigrants is 11 percent. Hence, on average, the conditional wage premium on training for immigrants is 3 percent less than that for natives.

Table 3.6 displays the estimated training coefficients using the pooled OLS estimates of the conditional effect of job training on wages when adding more variables on the right hand side. The results from the OLS model suggests that when we include more observable characteristics to wage equation, the effect of training on earnings reduces considerably. Yet, the estimated training coefficients remain positive and significant. Rows 2 to 10 show the training premium conditioning on different covariates (human capital, demographic and occupation sorting) that influence earnings. Row 2 adds the yearly dummies. Row 3 includes racial dummies. It reports that the effect of training drops to 16.7 percent for natives and to 15.4 percent for immigrants when racial dummies are included. The racial dummies have larger influence on estimated training coefficients for immigrants than natives; four percent and one percent reduction respectively.

Row 4 controls for age and education. The magnitudes of estimated coefficients reduce to 11 percent for immigrants and to 13.9 percent for natives. It is noteworthy that almost half of the training effect for immigrants and one-eighth of the training effect for natives are
due to the effects of schooling and work experience. Row 5 shows that when controlling for marital status, female head of household and having children under age of 18, we only find marginal changes in the estimated training coefficients. As expected, while being married and having children have positive impact on earnings, living in the female head of household house has negative impact on earnings. Row 6 adds private firm, metropolitan and firm size dummy variables to the wage equation. Surprisingly, we also found only minor impacts from metropolitan and firm size despite both variables being significant and positively correlated with earnings.

Row 7 includes a health insurance variable. Controlling for health insurance, we observe 3 percent reduction in estimated training coefficient for immigrants. Row 8 includes the union dummy variable. It shows that while being in the union increases wage earning around 14 percent, controlling for union status reduces the effects of training by 0.1 percent for natives and 0.3 percent for immigrants. Rows 9 and 10 show that estimated training coefficients stay relatively the same when state-level and industrial dummy variables are included. Row 11 indicates that adding an occupation covariate reduces the effect of training by 1.3 percent for natives and 1.9 percent for immigrants.

Next, we include interaction terms to the wage equation. When we include industry interaction dummy variables to the wage equation, we found a small reduction in the effect of training for natives. For immigrants, we also found a reduction in the effect of training, yet the training variable become statistically insignificant\textsuperscript{10} When we include occupation interaction dummy variables to the wage equation, we found a small increase in the effect of training for natives and again statistically insignificant coefficient for immigrants. In general, adding interaction covariate terms to the wage equation adds very little change to the training coefficients.

Consequently, we conclude that the differences in racial, schooling, experiences and occupation are the most relevant observable covariates that account for the majority of effect of training. After conditioning on all observable covariates, the OLS model indicates

\textsuperscript{10}Interaction result table is available upon request.
that the training premium for immigrants is lesser than natives.

**Sensitivity analysis**

**Education**

This section discusses the result from the sensitivity analysis. It is possible that the training variable is correlated with the covariates, causing spurious training coefficients. Due to the small sample size of immigrants, we perform the sensitivity analysis using the OLS model. In general, we perform the robustness check by modifying our original model, testing our model for high school graduate, college graduate, married and non-married groups. We know from the summary table that native college graduates received more training than high school graduates (nearly 10 percent more), and native high school graduates account for the most untrained native workers (36.9 percent) (Table 3.1). Similarly, we find that immigrant college graduates received more training than high school graduates (nearly 6 percent more), and immigrant high school graduates also account for the most untrained immigrant workers (nearly 25.4 percent) (Table 3.2).

Table 3.7 displays the estimated training coefficients for workers who graduated from high school versus workers who graduated from college, and married versus non-married workers. Row 1 and 2 show that training premium is relatively higher for natives who have bachelor degrees compared to those native who have high school diplomas. Surprisingly, we find the opposite effect for immigrants, particularly negative effect for immigrants with bachelor degrees. However, we note that the estimated training coefficient for immigrants with bachelor degrees is not statistically significant, and the sample size for immigrants in this group is very small.

Comparing natives and immigrants, we still find that natives enjoy higher training premium than immigrants. Training increases natives’ earnings around 5.6 percent versus immigrants’ earnings around 5.3 percent for high school graduates. For college graduates, training increases natives’ earnings around 6.4 percent, while training reduces earnings for
immigrants with college degree around 13.4 percent. The negative return on training for these immigrants is likely due to the small sample size of immigrants with college degree. We summarize that for this sensitive analysis, the effect of training for natives is strong, while the effect of training for immigrants is somewhat weak and ambiguous. Furthermore, it is notable that in general, college graduates have more training than high school graduates for both natives and immigrants, and the effect of training for both secondary school and college graduates is robust with natives enjoying higher training premium.

**Marital Status**

From the summary Table 3.1, we observe that the native married workers received more training than native non-married workers (nearly 9 percent more). For immigrants, non-married workers received slightly more training than married workers (Table 3.2). Similar to education groupings, we study the effect of training for married and non-married using separate wage equations. Comparing married and non-married, row 3 and 4 show that the effect of training on wage is relatively larger for non-married native workers, while the effect of training on wage is much smaller for non-married immigrants (Table 3.7). Comparing natives and immigrants, the training premium enjoyed by married workers is relatively the same for both natives and immigrants, while the training premium enjoyed by non-married workers is relatively the larger for the natives. However, the estimated training coefficient for immigrants is negative and not statistically significant.

For row 5 and 6, we remove health insurance, union, state, industry and occupation dummies from the wage equation for married and non-married workers (Table 3.7). We obtain statistically significant estimated training coefficient for immigrants, yet the main results do not change. Similar to the result from education, the effect of training for natives is strong, while the effect of training for immigrants is somewhat weak and ambiguous. From the result of the sensitivity analysis, we conclude that our original model is relatively robust where natives enjoy higher training premium than immigrants for the majority of
Looking at the sensitivity test, there is reason to believe that we will have a heterogeneous outcome. For example, within the high school graduate, both married and older cohort, the differences in return to training are negligible, while for the differences within college graduate, which are not married and younger, they are relatively large. In addition, when we study the average effect using econometric tools such as OLS, it is required that the wage density conform to the normal distribution condition. Yet, this may not always hold. Hence, it is important to explore the distributional effects.

**Quantile Regression (QREQ)**

Since one of our main concerns is the welfare of different income groups, especially, low-wage workers, the study of the distributional effect of training is particularly important. First, we explore the distributional impact of training on native workers using quantile regression. Figure 3.3 presents (solid line) conditional training wage premium of native workers across the quantiles distribution (the top figure), conditioning on race, age and education covariates, (dashed line) unconditional training wage premium of native workers across quantiles (trained workers' mean wage minus untrained workers' mean wage) and (light straight line) estimated OLS coefficient. It shows that the magnitude of estimated training coefficients are considerably lowered across the distribution after we include yearly, race, age and education dummies to the wage equation. The reduction of training premium amplifies at the upper half of the wage distribution, becoming more uniformly distributed with a small dip at the highest quantiles.

Figure 3.4 displays conditional training premium of natives when we include all observable characteristics to the wage equation (the top figure). The training effect on natives reduces more drastically when we include all observable covariates to the wage equation, becoming even more uniformly distributed, almost identical to the OLS line. Nevertheless, it is noticeable that despite the reduction in wage premium, we still observe that train-
ing raises wages throughout the quantiles distribution, at an increase of nearly 8 percent. Hence, according to Quantile regression, we find that training raises wage premium for all native workers in the distribution in a similar fashion.

Next, we study the impact of training on immigrant workers across quantiles. Figure 3.3 displays (solid line) training wage premium of immigrant workers across the quantiles distribution conditioning on race, age and education covariates, (dashed line) unconditional training wage premium of immigrants and (light straight line) estimated OLS coefficient (the bottom figure). Similar to native workers, we observe that the conditioned distributional training premium of immigrant workers is reduced considerably. Yet, the reduction is much more apparent than that of natives, especially at the highest quantiles.

When we include all observable characteristics to the wage equation, we observe that the quantiles wage premium for immigrants drop further (the bottom figure). It changes from rapid monotonically increase of unconditional training premium across the quantiles distribution to slow monotonically decrease of conditional training premium (Figure 3.4). Despite the reduction in training premium, we still observe that training increases earnings at the lower and middle quantiles, an increase of nearly 4 percent.

Consequently, according to the QREQ model, low and mid-income immigrants still have lower training premium than natives, while high-income immigrants, who had higher unconditional training premium, now have much lower conditional training premium than natives. It is noteworthy that similar to OLS estimates, the Quantile regression may suffer from upward bias. As mentioned in the methodology section, we concede that we will not be able to resolve unobservable selection problems. Yet, we will explore the counterfactual study alternative. It can be argued that some differences in wage premium between natives and immigrants are due to their observable characteristics. In the next section, we will further explore the training effect on the distribution using the DFL weighting technique.
3.5.3 Counterfactual Study

In the similar spirit of the DFL, this section presents a counterfactual study of job training, simulating the quantile distribution of training premium, supposing both trained and untrained workers have similar observable characteristics. First, using a counterfactual study, we explore the impact of training on native workers. The top of Figure 3.5 and 3.6 present (dashed line) unconditional job training premium of native workers and (solid line) counterfactual training premium of native workers (trained workers’ wage minus untrained workers’ wage supposes these untrained native workers have similar observable characteristics as trained workers) (the top figure). When we corrected for observable characteristics differences between trained and untrained native workers, we found ambiguous results.

When we remove race, age and education differences between trained and untrained native workers, we observe that the counterfactual premium becomes slightly more uniform. We find that training premium drops marginally at the upper half and increases negligibly at the lower half of the income distribution, reducing training premium at upper quantiles around 2 to 3 percent and increasing training premium at lowest quantiles around 2 percent (the top of Figure 3.5). The results show that suppose high income untrained natives have similar education as trained natives, they would receive lower wage.

When we removed all observable characteristic differences between trained and untrained natives, we observed that the counterfactual training premium changed from being relatively uniform to a slight monotonically increase from lower to upper quantiles (the top of Figure 3.6). We find that counterfactual training premium generally remain unchanged at the lower half of the income distribution. Yet, at the upper half of the income distribution, the counterfactual training premium is surprisingly greater than the unconditional wage premium. Suppose high income untrained natives have similar observable characteristics as trained natives, this result indicates that untrained natives will actually receive even lower wage.

Next, continuing to apply the counterfactual framework, we explore the impact of train-
ing on immigrant workers. The bottom of Figure 3.5 displays (dashed line) unconditional training premium of immigrant workers and (solid line) counterfactual training premium of immigrant workers (trained workers’ wage minus untrained workers’ wage supposes these untrained immigrant workers have similar observable characteristics as trained workers). Unlike the income premium distribution of native workers, we observe that training premium of immigrant workers reduces considerably after corrected for observable characteristics differences between trained and untrained workers.

When we corrected for race, age and education differences between trained and untrained immigrant workers, we found that the counterfactual training premium was reduced considerably, particularly, at the upper half of the income distribution. Although the effect is minimal at lower quantiles, the reduction of immigrants’ wage premium is much greater than the natives, dropping the wage premium at .50 Quantile around 8 percent and at .95 Quantile around 13 percent (the bottom of Figure 3.5). Similar to natives, the results show that if high income untrained immigrants have similar education as trained immigrants, they would receive a lower wage, yet in much larger scale.

The bottom of Figure 3.6 displays counterfactual training premium when we removed all observable characteristics differences between trained and untrained immigrants. We find that the counterfactual training premium is still smaller than unconditional training premium, but the effect from correcting for all observable characteristics differences is not as large as for correcting for only race, age and education differences. Yet, we still find a reduction of wage premium, particularly, at the upper half of the income distribution, dropping the wage premium at upper half of the distribution around 5 percent (the bottom of Figure 3.6).

Using the DFL weighting method, we found that after removing all observable characteristics differences between trained and untrained workers, training still increases wage premium for both natives and immigrants throughout the income distribution. Similar to Abadie et al. (2002) that found the impact of training only at the upper half of the income distribution, we observe largest impact of training at upper half of the income distribution.
for both natives and immigrants. Hence, high income workers still benefit most from training. These results suggest that training premium for highly skilled workers is higher than lower skilled workers. Also, training premium is lower for immigrants than natives for the majority of the income distribution, and training has the smallest effect for very low-skilled and low-wage immigrant workers. Nevertheless, we still find that training increases wage premium of low and middle income workers, including immigrants.

3.5.4 Propensity Score Matching

Table 3.8 presents both the OLS and propensity score matching results of training premiums. The OLS results show that the job training premium for foreign-born workers is a positive value of 0.039, whereas for native-born workers it is 0.076. Our results show that there is 4-percent difference in the job training premium between native and immigrant. A common support condition is imposed by propensity score matching to improve the quality of the matches. We present results based on nearest-neighbor matching and kernel matching using the Epanechnikov kernel with a bandwidth of 0.63 which are utilized by Eren (2007). Nearest-neighbor matching indicates a positive value of 0.063 (training premium) for foreign-born workers. Similarly, kernel matching estimate indicates 0.184 (training premium). For native-born workers, the estimate based on nearest neighbor matching is 0.108 and the result of kernel matching is 0.229. All estimates are statistically significant. Those matching results are higher than the OLS results for both native-born and foreign-born workers, especially the results of kernel matching. Our results suggest that OLS estimates underestimate the training premium.

3.6 Conclusion

In this paper, we study the effect of job training on the US immigrant workers, using the 1996, 2001 and 2004, Survey of Income and Program Participation (SIPP) data. Job
training is the essential key for immigrant workers, who often face immense difficulty in the labor market that tends to favor native workers, to improve their standard of living. Training increases life time earning capability of immigrants, which is rewarded in the labor market and helps reduce poverty driven social problems. Since immigrants are important and necessary part of the US labor market and represent a large fraction of the workers, it is important to address and understand the true effect of training on immigrants in the US.

Earlier studies on training rarely look at immigrants, and few studies that look at the effect of training on immigrants utilize economic models, using instead only descriptive and mean table as analytical tools. Hence, this allows us to study different aspects of training and immigrants that have not been explored. As a result, we improve upon prior studies by setting up our training evaluation model, studying the impact of training on both the average and the distributional earning of workers and comparing the differences in the return to training for immigrant and native workers by applying the Quantile regression (QREQ) model, the DiNardo, Fortin and Lemieux (DFL) reweighting methods, and propensity score matching method.

From our mean analysis, we find that training has a positive and significant effect on wages of the average immigrant worker. Looking at the unconditional training premium, our analysis suggests that though natives earn more than immigrants, the training premium for immigrants is relatively larger than natives. In other words, immigrants, who received job training, earn higher wage premium than natives who have received job training.

Our sensitivity analysis results show that our original model is relatively robust with high school graduates, college graduates, married and non-married workers where natives enjoy higher training premium than immigrants for the majority of cases.

From our distribution study, we find that training has a positive effect on wages of immigrant workers for most parts of income distribution. The results suggest that the effect of training across workers income quantiles is relatively different compared to the effect of training at the mean for both immigrants and natives. The differences in the effect of training appear large, interesting and important for welfare consideration when we look
at the effect of training across the different quantiles.

Looking at the unconditional training premium, the results show that training increases earnings throughout the quantiles for both immigrants and natives. We observe that immigrants enjoy largest unconditional training premium at the upper part of workers’ earning quantiles, and they enjoy lowest unconditional training premium at the lower part of workers’ earning quantiles. We find evidence that natives enjoy largest unconditional training premium at the middle of workers’ earning quantiles. Comparing natives and immigrants, it is notable that the unconditional training wage premium for immigrants is considerably larger than natives at upper quantiles and similar to natives at the lower and middle quantiles. As a result, we observe a more remarkable unconditional gain from training for wealthy immigrants and less gain for poorer immigrants.

Examining counterfactual study, the DFL reweighting technique shows that after removing all observable characteristics differences between trained and untrained workers, training still increases wage premium for both natives and immigrants throughout the income distribution. Counterfactual training simulates the quantile distribution of training premium, if untrained workers have similar observable characteristics as trained workers. After controlling for all observable characteristics, we observe a sizeable reduction in training premium for immigrants, yet we note a small increase in training premium for natives.

Similar to Abadie et al. (2002) that found an impact of training only at the upper half of the income distribution, we observe the largest proportional impact of training at upper half of the income distribution for both natives and immigrants. Our analysis provides strong evidence for the hypothesis that after corrected for observable characteristics differences between trained and untrained workers, the effect of training is relatively larger for rich natives, much larger for middle income natives and similar for the poor natives and immigrants. Nevertheless, we still find that training increases wage premium of low and middle income workers, including immigrants.
3.6.1 Policy Implication

There are several proposed initiatives that policy makers can take away from this study. The practical lesson is that job training is beneficial and important to the improvement of immigrants’ well-being, yet many immigrants are still deprived of these much needed training. Although we did not find the largest impact of training for low-skilled and low-wage immigrant workers, we did find a strong and positive training impact for this group. Hence, these low-skilled and low-wage immigrants should be one of the main target groups of training provision, since they need the most assistance in obtaining training and would greatly benefit from the result of training.

Policy makers can restructure the existing programs to allow easier excess for immigrants such as revamping the Workforce Investment Act by changing the English prerequisite. Also, it is important to concentrate on outreach programs that increase awareness to Limited English Proficiency (LEP) workers regarding the availability of job training. Furthermore, realizing that these poor workers earn their living day by day, tangible assistance such as providing of transportation and childcare arrangements during training may be necessary. In addition, policy makers may consider offering English as a Second Language (ESL) classes and training programs simultaneously to immigrants, focusing on providing English literacy for agricultural workers and providing of English and skills training for manufacturing and service workers.

For immigrants that are unable to participate in the training program immediately, policy makers can allocate funds for job fairs that target immigrants, providing help with filling out applications and language assisted interviews. For private sector, government agencies can redirect some resources to give companies incentive to provide training for immigrants. Tax cuts and funding can be used as incentive tools to encourage firms to grant training to immigrants and managers to promote workforce diversity.
3.6.2 Future Work

There remain many facets of the effect of training on immigrants that have not been explored. Our framework can be extended to study other minority groups within immigrants, particularly concentrating on women, youth and other racial ethnic immigrants such as Black, Hispanic and Asian. It is important to pay attention to these subgroups, especially youth, since they are the future workforce and would provide lifetime return on social investment.

From the distribution study, our application of the DFL reweighting technique can further be used to identify the observable characteristics differences between trained and untrained workers that are most influential to training premium at different earning quantiles. Hence, this application is useful in assisting policy makers to pinpoint existing problems. Next, since our DFL reweighting analysis relies on the assumptions that treatment selection is based on observable characteristics, it is possible that selection problem may bias our distribution estimate. An instrument variable to use in quantiles treatment effects (QTE), Abadie et al. (2002) can be another possible research avenue.

In the future, we will try to correct the problem and check whether our treatment effect is significant or not. Furthermore, to further resolve unobserved heterogeneity problem, creating a better panel dataset with the focus on immigrants and training would be very beneficial.

In future work, we plan to explore the effect of training on immigrant work hours and employment. In addition, due to the shortcoming of our SIPP data, we cannot identify time since exposure to training, amount of training received and length of training exposure. With other data set, future research should investigate the effect of length of training exposure similar to Flores-Lagunes et al. (2007).
<table>
<thead>
<tr>
<th>Summary</th>
<th>Obs</th>
<th>Native (All)</th>
<th>Standard Errors</th>
<th>Obs</th>
<th>Training</th>
<th>Standard Errors</th>
<th>Obs</th>
<th>No Training</th>
<th>Standard Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Hourly Wage</td>
<td>24401</td>
<td>2.399</td>
<td>0.456</td>
<td>9519</td>
<td>2.595</td>
<td>0.448</td>
<td>14882</td>
<td>2.437</td>
<td>0.451</td>
</tr>
<tr>
<td>Monthly Income</td>
<td>24401</td>
<td>2482</td>
<td>2134</td>
<td>9519</td>
<td>2824</td>
<td>2156</td>
<td>14882</td>
<td>2262</td>
<td>2091</td>
</tr>
<tr>
<td>Training last 10 year</td>
<td>24401</td>
<td>0.391</td>
<td>0.488</td>
<td>9519</td>
<td>1.000</td>
<td>0.000</td>
<td>14882</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Training last 1 year</td>
<td>24401</td>
<td>0.199</td>
<td>0.399</td>
<td>9519</td>
<td>0.508</td>
<td>0.500</td>
<td>14882</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>1 Day to 1 Week</td>
<td>934</td>
<td>0.361</td>
<td>0.480</td>
<td>934</td>
<td>0.361</td>
<td>0.480</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>More than 1 Week</td>
<td>934</td>
<td>0.333</td>
<td>0.471</td>
<td>934</td>
<td>0.333</td>
<td>0.471</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Currently in Training</td>
<td>934</td>
<td>0.109</td>
<td>0.312</td>
<td>934</td>
<td>0.109</td>
<td>0.312</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>White</td>
<td>23896</td>
<td>0.855</td>
<td>0.352</td>
<td>9300</td>
<td>0.881</td>
<td>0.324</td>
<td>14596</td>
<td>0.839</td>
<td>0.368</td>
</tr>
<tr>
<td>Black</td>
<td>23896</td>
<td>0.134</td>
<td>0.340</td>
<td>9300</td>
<td>0.107</td>
<td>0.309</td>
<td>14596</td>
<td>0.151</td>
<td>0.358</td>
</tr>
<tr>
<td>Hispanic</td>
<td>24401</td>
<td>0.088</td>
<td>0.283</td>
<td>9519</td>
<td>0.067</td>
<td>0.249</td>
<td>14882</td>
<td>0.101</td>
<td>0.302</td>
</tr>
<tr>
<td>Asian</td>
<td>23896</td>
<td>0.011</td>
<td>0.104</td>
<td>9300</td>
<td>0.013</td>
<td>0.111</td>
<td>14596</td>
<td>0.010</td>
<td>0.100</td>
</tr>
<tr>
<td>Age</td>
<td>24401</td>
<td>38.985</td>
<td>11.343</td>
<td>9519</td>
<td>38.957</td>
<td>10.826</td>
<td>14882</td>
<td>39.003</td>
<td>11.663</td>
</tr>
<tr>
<td>Highest grade &lt; 9</td>
<td>23735</td>
<td>0.024</td>
<td>0.152</td>
<td>9141</td>
<td>0.010</td>
<td>0.097</td>
<td>14594</td>
<td>0.032</td>
<td>0.177</td>
</tr>
<tr>
<td>Highest grade &lt; 12</td>
<td>23735</td>
<td>0.093</td>
<td>0.290</td>
<td>9141</td>
<td>0.055</td>
<td>0.228</td>
<td>14594</td>
<td>0.117</td>
<td>0.321</td>
</tr>
<tr>
<td>High school diploma</td>
<td>23735</td>
<td>0.416</td>
<td>0.493</td>
<td>9141</td>
<td>0.347</td>
<td>0.476</td>
<td>14594</td>
<td>0.459</td>
<td>0.498</td>
</tr>
<tr>
<td>Some college</td>
<td>23735</td>
<td>0.555</td>
<td>0.479</td>
<td>9141</td>
<td>0.438</td>
<td>0.496</td>
<td>14594</td>
<td>0.303</td>
<td>0.459</td>
</tr>
<tr>
<td>Bachelor diploma</td>
<td>23735</td>
<td>0.092</td>
<td>0.289</td>
<td>9141</td>
<td>0.121</td>
<td>0.326</td>
<td>14594</td>
<td>0.073</td>
<td>0.261</td>
</tr>
<tr>
<td>Master or higher</td>
<td>23735</td>
<td>0.020</td>
<td>0.139</td>
<td>9141</td>
<td>0.029</td>
<td>0.167</td>
<td>14594</td>
<td>0.014</td>
<td>0.119</td>
</tr>
<tr>
<td>Married</td>
<td>24401</td>
<td>0.557</td>
<td>0.497</td>
<td>9141</td>
<td>0.599</td>
<td>0.490</td>
<td>14594</td>
<td>0.530</td>
<td>0.499</td>
</tr>
<tr>
<td>Female head</td>
<td>24401</td>
<td>0.053</td>
<td>0.223</td>
<td>9141</td>
<td>0.035</td>
<td>0.184</td>
<td>14594</td>
<td>0.064</td>
<td>0.245</td>
</tr>
<tr>
<td>Kids 18 years or less</td>
<td>24401</td>
<td>0.389</td>
<td>0.488</td>
<td>9141</td>
<td>0.406</td>
<td>0.491</td>
<td>14594</td>
<td>0.378</td>
<td>0.485</td>
</tr>
<tr>
<td>Metropolitan area</td>
<td>24075</td>
<td>0.741</td>
<td>0.438</td>
<td>9388</td>
<td>0.740</td>
<td>0.439</td>
<td>14687</td>
<td>0.741</td>
<td>0.438</td>
</tr>
<tr>
<td>25 to 99 employees</td>
<td>24112</td>
<td>0.241</td>
<td>0.428</td>
<td>9409</td>
<td>0.236</td>
<td>0.424</td>
<td>14703</td>
<td>0.244</td>
<td>0.429</td>
</tr>
<tr>
<td>100+ employees</td>
<td>24112</td>
<td>0.423</td>
<td>0.494</td>
<td>9409</td>
<td>0.479</td>
<td>0.500</td>
<td>14703</td>
<td>0.387</td>
<td>0.487</td>
</tr>
<tr>
<td>Private sector</td>
<td>24134</td>
<td>0.872</td>
<td>0.334</td>
<td>9419</td>
<td>0.826</td>
<td>0.379</td>
<td>14715</td>
<td>0.901</td>
<td>0.298</td>
</tr>
<tr>
<td>Public sector</td>
<td>24134</td>
<td>0.128</td>
<td>0.334</td>
<td>9419</td>
<td>0.174</td>
<td>0.379</td>
<td>14715</td>
<td>0.099</td>
<td>0.298</td>
</tr>
<tr>
<td>Health Insurance</td>
<td>24401</td>
<td>0.786</td>
<td>0.410</td>
<td>9519</td>
<td>0.857</td>
<td>0.350</td>
<td>14882</td>
<td>0.740</td>
<td>0.439</td>
</tr>
<tr>
<td>Union</td>
<td>18542</td>
<td>0.016</td>
<td>0.125</td>
<td>6815</td>
<td>0.024</td>
<td>0.152</td>
<td>11727</td>
<td>0.011</td>
<td>0.105</td>
</tr>
<tr>
<td>Employed</td>
<td>24401</td>
<td>0.970</td>
<td>0.171</td>
<td>9519</td>
<td>0.976</td>
<td>0.153</td>
<td>14882</td>
<td>0.966</td>
<td>0.181</td>
</tr>
<tr>
<td>Low English</td>
<td>8922</td>
<td>0.013</td>
<td>0.112</td>
<td>3052</td>
<td>0.003</td>
<td>0.058</td>
<td>5870</td>
<td>0.018</td>
<td>0.132</td>
</tr>
</tbody>
</table>

Source: SIPP 1996, 2001 and 2004. Notes: Sample includes all adults male between 22 to 65 years of age. The omitted category for length of training is less than 8 hours of training, age is 22 to 29 years, education is less than first grade and firm size is under 25 employees.
Table 3.2: Mean Values of Immigrants

<table>
<thead>
<tr>
<th>Summary</th>
<th>Obs Immigrant (All)</th>
<th>Standard Errors</th>
<th>Obs Training</th>
<th>Standard Errors</th>
<th>Obs No Training</th>
<th>Standard Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Hourly Wage</td>
<td>4486</td>
<td>2.319</td>
<td>0.435</td>
<td>991</td>
<td>2.457</td>
<td>0.456</td>
</tr>
<tr>
<td>Monthly Income</td>
<td>4486</td>
<td>1985</td>
<td>1579</td>
<td>991</td>
<td>2439</td>
<td>2118</td>
</tr>
<tr>
<td>Training last 10 year</td>
<td>4486</td>
<td>0.216</td>
<td>0.411</td>
<td>991</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Training last 1 year</td>
<td>4486</td>
<td>0.097</td>
<td>0.296</td>
<td>991</td>
<td>0.449</td>
<td>0.498</td>
</tr>
<tr>
<td>1 Day to 1 Week</td>
<td>117</td>
<td>0.322</td>
<td>0.469</td>
<td>117</td>
<td>0.322</td>
<td>0.469</td>
</tr>
<tr>
<td>More than 1 Week</td>
<td>117</td>
<td>0.345</td>
<td>0.477</td>
<td>117</td>
<td>0.345</td>
<td>0.477</td>
</tr>
<tr>
<td>Currently in Training</td>
<td>117</td>
<td>0.112</td>
<td>0.316</td>
<td>117</td>
<td>0.112</td>
<td>0.316</td>
</tr>
<tr>
<td>White</td>
<td>3972</td>
<td>0.834</td>
<td>0.372</td>
<td>829</td>
<td>0.779</td>
<td>0.415</td>
</tr>
<tr>
<td>Black</td>
<td>3972</td>
<td>0.109</td>
<td>0.312</td>
<td>829</td>
<td>0.144</td>
<td>0.351</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4486</td>
<td>0.600</td>
<td>0.490</td>
<td>991</td>
<td>0.417</td>
<td>0.493</td>
</tr>
<tr>
<td>Asian</td>
<td>3972</td>
<td>0.057</td>
<td>0.232</td>
<td>829</td>
<td>0.077</td>
<td>0.266</td>
</tr>
<tr>
<td>Age</td>
<td>4486</td>
<td>37.277</td>
<td>10.764</td>
<td>991</td>
<td>38.216</td>
<td>10.492</td>
</tr>
<tr>
<td>Age square</td>
<td>4486</td>
<td>15.054</td>
<td>8.744</td>
<td>991</td>
<td>15.705</td>
<td>8.556</td>
</tr>
<tr>
<td>Highest grade &lt; 9</td>
<td>4429</td>
<td>0.260</td>
<td>0.439</td>
<td>972</td>
<td>0.096</td>
<td>0.294</td>
</tr>
<tr>
<td>Highest grade &lt; 12</td>
<td>4429</td>
<td>0.132</td>
<td>0.338</td>
<td>972</td>
<td>0.097</td>
<td>0.296</td>
</tr>
<tr>
<td>High school diploma</td>
<td>4429</td>
<td>0.271</td>
<td>0.444</td>
<td>972</td>
<td>0.259</td>
<td>0.438</td>
</tr>
<tr>
<td>Some college</td>
<td>4429</td>
<td>0.201</td>
<td>0.401</td>
<td>972</td>
<td>0.336</td>
<td>0.473</td>
</tr>
<tr>
<td>Bachelor diploma</td>
<td>4429</td>
<td>0.090</td>
<td>0.286</td>
<td>972</td>
<td>0.156</td>
<td>0.363</td>
</tr>
<tr>
<td>Master or higher</td>
<td>4429</td>
<td>0.028</td>
<td>0.164</td>
<td>972</td>
<td>0.049</td>
<td>0.216</td>
</tr>
<tr>
<td>Married</td>
<td>4486</td>
<td>0.633</td>
<td>0.482</td>
<td>991</td>
<td>0.626</td>
<td>0.484</td>
</tr>
<tr>
<td>Female head</td>
<td>4486</td>
<td>0.041</td>
<td>0.198</td>
<td>991</td>
<td>0.040</td>
<td>0.196</td>
</tr>
<tr>
<td>Kids 18 years or less</td>
<td>4486</td>
<td>0.549</td>
<td>0.498</td>
<td>991</td>
<td>0.530</td>
<td>0.499</td>
</tr>
<tr>
<td>Metropolitan area</td>
<td>4393</td>
<td>0.903</td>
<td>0.296</td>
<td>967</td>
<td>0.906</td>
<td>0.291</td>
</tr>
<tr>
<td>25 to 99 employees</td>
<td>4438</td>
<td>0.261</td>
<td>0.439</td>
<td>980</td>
<td>0.270</td>
<td>0.444</td>
</tr>
<tr>
<td>100+ employees</td>
<td>4438</td>
<td>0.350</td>
<td>0.477</td>
<td>980</td>
<td>0.470</td>
<td>0.499</td>
</tr>
<tr>
<td>Private sector</td>
<td>4441</td>
<td>0.946</td>
<td>0.227</td>
<td>980</td>
<td>0.891</td>
<td>0.312</td>
</tr>
<tr>
<td>Public sector</td>
<td>4441</td>
<td>0.054</td>
<td>0.227</td>
<td>980</td>
<td>0.109</td>
<td>0.312</td>
</tr>
<tr>
<td>Health Insurance</td>
<td>4486</td>
<td>0.537</td>
<td>0.499</td>
<td>991</td>
<td>0.727</td>
<td>0.446</td>
</tr>
<tr>
<td>Union</td>
<td>3784</td>
<td>0.011</td>
<td>0.102</td>
<td>749</td>
<td>0.026</td>
<td>0.160</td>
</tr>
<tr>
<td>Employed</td>
<td>4486</td>
<td>0.973</td>
<td>0.162</td>
<td>991</td>
<td>0.967</td>
<td>0.179</td>
</tr>
<tr>
<td>Low English</td>
<td>1694</td>
<td>0.422</td>
<td>0.494</td>
<td>330</td>
<td>0.218</td>
<td>0.413</td>
</tr>
</tbody>
</table>

Source: SIPP 1996, 2001 and 2004. Notes: Sample includes all adults male between 22 to 65 years of age. The omitted category for length of training is less than 8 hours of training, age is 22 to 29 years, education is less than first grade and firm size is under 25 employees.
Table 3.3: Ordinary Least Squares Regression Results.

<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>Immigrant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training last 10 year</td>
<td>848.357</td>
<td>1480.882</td>
</tr>
<tr>
<td></td>
<td>(40.459)</td>
<td>(101.594)</td>
</tr>
<tr>
<td>Constant</td>
<td>3357.008</td>
<td>2671.511</td>
</tr>
<tr>
<td></td>
<td>(27.458)</td>
<td>(43.054)</td>
</tr>
<tr>
<td>N Observations</td>
<td>49642</td>
<td>8256</td>
</tr>
</tbody>
</table>


Table 3.4: OLS Model. Estimate Effect of Training on Earnings. Dependent Variable: Log of Hourly Earnings.

<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>Immigrant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Hourly Wage</td>
<td>Pooled OLS</td>
<td>Pooled OLS</td>
</tr>
<tr>
<td>Training last 10 year</td>
<td>0.177</td>
<td>0.197</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Year 2001</td>
<td>0.192</td>
<td>0.219</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Year 2004</td>
<td>0.238</td>
<td>0.249</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Constant</td>
<td>2.288</td>
<td>2.102</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>N Observations</td>
<td>24401</td>
<td>4486</td>
</tr>
</tbody>
</table>

Source: Data are from SIPP, and include individuals from year 1996, 2001 and 2004. Notes: Sample includes all adults male between 22 to 65 years of age. The omitted category for age is 22 to 29 years and education is less than first grade. Standard errors are in parentheses.
Table 3.5: OLS Model. Estimate Effect of Training on Earnings. Dependent Variable: Log of Hourly Earnings.

<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>Immigrant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pooled OLS</td>
<td>Pooled OLS</td>
</tr>
<tr>
<td>Training last 10 year</td>
<td>0.139 (0.006)</td>
<td>0.110 (0.018)</td>
</tr>
<tr>
<td>Year 2001</td>
<td>0.184 (0.007)</td>
<td>0.211 (0.018)</td>
</tr>
<tr>
<td>Year 2004</td>
<td>0.231 (0.007)</td>
<td>0.229 (0.017)</td>
</tr>
<tr>
<td>White</td>
<td>0.058 (0.030)</td>
<td>0.035 (0.031)</td>
</tr>
<tr>
<td>Black</td>
<td>-0.092 (0.031)</td>
<td>-0.046 (0.036)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.059 (0.012)</td>
<td>-0.117 (0.018)</td>
</tr>
<tr>
<td>Age: 30-39</td>
<td>0.221 (0.010)</td>
<td>0.148 (0.023)</td>
</tr>
<tr>
<td>Age: 40-49</td>
<td>0.301 (0.017)</td>
<td>0.172 (0.040)</td>
</tr>
<tr>
<td>Age: 50-65</td>
<td>0.249 (0.028)</td>
<td>0.124 (0.073)</td>
</tr>
<tr>
<td>Age square</td>
<td>0.002 (0.001)</td>
<td>0.005 (0.003)</td>
</tr>
<tr>
<td>Highest grade &lt; 9</td>
<td>0.243 (0.125)</td>
<td>0.194 (0.039)</td>
</tr>
<tr>
<td>Highest grade &lt; 12</td>
<td>0.334 (0.124)</td>
<td>0.243 (0.041)</td>
</tr>
<tr>
<td>High school diploma</td>
<td>0.434 (0.124)</td>
<td>0.293 (0.040)</td>
</tr>
<tr>
<td>Some college</td>
<td>0.476 (0.124)</td>
<td>0.347 (0.042)</td>
</tr>
<tr>
<td>Bachelor diploma</td>
<td>0.507 (0.124)</td>
<td>0.387 (0.049)</td>
</tr>
<tr>
<td>Master or higher</td>
<td>0.625 (0.126)</td>
<td>0.431 (0.064)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.604 (0.127)</td>
<td>1.719 (0.056)</td>
</tr>
<tr>
<td>N Observations</td>
<td>23896</td>
<td>3972</td>
</tr>
</tbody>
</table>

Source: Data are from SIPP, and include individuals from year 1996, 2001 and 2004. Notes: Sample includes all adults male between 22 to 65 years of age. The omitted category for age is 22 to 29 years and education is less than first grade. Standard errors are in parentheses.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Natives)</td>
<td>(Im.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ins.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.158</td>
<td>0.175</td>
<td>0.017</td>
<td>24401</td>
<td>4486</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.177</td>
<td>0.197</td>
<td>0.020</td>
<td>24401</td>
<td>4486</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.167</td>
<td>0.154</td>
<td>-0.013</td>
<td>23896</td>
<td>3972</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.139</td>
<td>0.110</td>
<td>-0.028*</td>
<td>23235</td>
<td>3929</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.126</td>
<td>0.111</td>
<td>-0.015</td>
<td>23235</td>
<td>3929</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.112</td>
<td>0.087</td>
<td>-0.025</td>
<td>22657</td>
<td>3803</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.097</td>
<td>0.056</td>
<td>-0.040**</td>
<td>22657</td>
<td>3803</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.096</td>
<td>0.053</td>
<td>-0.043**</td>
<td>17421</td>
<td>3258</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.091</td>
<td>0.050</td>
<td>-0.040**</td>
<td>17421</td>
<td>3258</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.089</td>
<td>0.058</td>
<td>-0.031*</td>
<td>17421</td>
<td>3258</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.076</td>
<td>0.039</td>
<td>-0.036**</td>
<td>17421</td>
<td>3258</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Data are from SIPP, and include individuals from year 1996, 2001 and 2004. Notes: Standard errors are in parentheses. Sample includes adults’ male between 22 to 65 years of age. Row 1 is the unconditional pooled OLS. Row 2 includes yearly dummies. Row 3 adds race (White, Black and Hispanic). Row 4 includes four brackets of seniority dummies, and seniority squared divided by 100 and seven brackets of years of education dummies. Row 5 adds marital status, dummies variable indicating whether the individual lives in the female head household and have children younger than 18 living in the family. Row 6 adds metropolitan, private firm and three brackets of firm size dummies. Row 7 includes dummy variable denoting possession of health insurance. Row 8 adds union dummies. Row 9 includes state dummies. Row 10 adds ten industry dummies. Row 11 includes ten occupation dummies. * indicates 90 percent statistically significant different between natives and immigrants. ** indicates 95 percent statistically significant different between natives and immigrants.
Table 3.7: Sensitivity Analysis. OLS Estimate Effect of Training on Earnings. Dependent Variable: Log of Hourly Earning

<table>
<thead>
<tr>
<th>Training</th>
<th>Native</th>
<th>Im.</th>
<th>Diff.</th>
<th>Firm Size</th>
<th>Health Insur.</th>
<th>Union</th>
<th>State</th>
<th>Ind.</th>
<th>Occ.</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.099)</td>
<td></td>
<td>-0.003</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>7194 / 838</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td></td>
<td>(0.023)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>1599 / 233</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td></td>
<td>-0.134</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>7853 / 1140</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td></td>
<td>(0.008)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>9568 / 2118</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td></td>
<td>(0.020)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>7853 / 1286</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td></td>
<td>(0.021)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>9568 / 2517</td>
</tr>
</tbody>
</table>

Source: Data are from SIPP, and include individuals from year 1996, 2001 and 2004. Sample includes adults’ male between 15 to 65 years of age. Notes: Standard errors are below coefficients. All observable covariates includes yearly dummies, race (White, Black and Hispanic) dummies, seniority dummies, seniority squared divide by 100, years of education dummies, marital status dummies, dummies variable indicating whether the individual lives in the female head household, having children younger than 18 living in the family dummies, metropolitan, private firm, firm size dummies, dummy variable denoting possession of health insurance, union dummies, state dummies, industry dummies and occupation dummies. Row 1 is the pooled OLS model conditioned on all observable characteristics with sample including individuals with High School diploma. Row 2 is the pooled OLS model conditioned on all observable characteristics with sample including individuals with bachelor diploma. Row 3 is the pooled OLS model conditioned on all observable characteristics with sample including not currently married individuals. Row 4 is the pooled OLS model conditioned on all observable characteristics with sample including currently married individuals. Row 5 is the pooled OLS model conditioned on all observable characteristics except health insurance, union, state, industry and occupation dummies with sample including not currently married individuals. Row 6 is the pooled OLS model conditioned on all observable characteristics except health insurance, union, state, industry and occupation dummies with sample including currently married individuals.
Table 3.8: OLS/Matching Estimate Effect of Training on Earnings

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Native</th>
<th>Immigrant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training N. Treat. N. Control</td>
<td>Training N. Treat. N. Control</td>
</tr>
<tr>
<td></td>
<td>Premium</td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>0.076*** (0.006)</td>
<td>0.039*** (0.016)</td>
</tr>
<tr>
<td>Nearest Neighbor Matching</td>
<td>0.108*** (0.007)</td>
<td>0.063*** (0.022)</td>
</tr>
<tr>
<td>Kernel Matching</td>
<td>0.229*** (0.005)</td>
<td>0.184*** (0.017)</td>
</tr>
<tr>
<td></td>
<td>1078 4450</td>
<td>304 2488</td>
</tr>
</tbody>
</table>

Note: Sample includes male workers between 15 to 65 years of age. Standard errors are in parentheses. * Statistically significant at 0.10 level; *** at the 0.05 level. For OLS and PSM, all observable covariates includes yearly dummies, race (White, Black and Hispanic) dummies, seniority dummies, seniority squared divide by 100, years of education dummies, marital status dummies, dummies variable indicating whether the individual lives in the female head household, having children younger than 18 living in the family dummies, metropolitan, private firm, firm size dummies, dummy variable denoting possession of health insurance, union dummies, state dummies, industry dummies and occupation dummies. The OLS observations for male native-born workers number 21,489 and those for male foreign-born workers 2,528.
Figure 3.1: Unconditional Effect of Job Training on Earnings for Natives and Immigrants. Dependent Variable: Log of Hourly Earnings.

Source: Data are from SIPP, and include individuals from year 1996, 2001 and 2004. Sample includes adults' male between 15 to 65 years of age.
Figure 3.2: Unconditional Effect of Job Training on Earnings for Natives and Immigrants. Dependent Variable: Log of Hourly Earnings.

Source: Data are from SIPP, and include individuals from year 1996, 2001 and 2004. Sample includes adults' male between 15 to 65 years of age.
Figure 3.3: Conditional Effect of Job Training on Earnings for Natives and Immigrants (Quantiles Regression). Dependent Variable: Log of Hourly Earnings.

Source: Data are from SIPP, and include individuals from year 1996, 2001 and 2004. Sample includes adults' male between 15 to 65 years of age. Note: Quantiles regression conditioning on yearly dummies, race (White, Black and Hispanic) dummies, seniority dummies, seniority squared divide by 100, years of education dummies.
Figure 3.4: Conditional Effect of Job Training on Earnings for Natives and Immigrants (Quantiles Regression). Dependent Variable: Log of Hourly Earnings.

Source: Data are from SIPP, and include individuals from year 1996, 2001 and 2004. Sample includes adults’ male between 15 to 65 years of age. Note: Quantiles regression conditioning on yearly dummies, race (White, Black and Hispanic) dummies, seniority dummies, seniority squared divide by 100, years of education dummies, marital status dummies, dummy variable indicating whether the individual lives in the female head household, having children younger than 18 living in the family dummies, metropolitan, private firm, firm size dummies, dummy variable denoting possession of health insurance, union dummies, state dummies, industry dummies and occupation dummies.
Figure 3.5: Unconditional and Counterfactual Effect of Job Training on Earnings for Natives and Immigrants (DFL Model). Dependent Variable: Log of Hourly Earnings.

Source: Data are from SIPP, and include individuals from year 1996, 2001 and 2004. Sample includes adults’ male between 15 to 65 years of age. Note: DiNardo, Fortin and Lemieux (DFL) model conditions on yearly dummies, race (White, Black and Hispanic) dummies, seniority dummies, seniority squared divide by 100, years of education dummies.
Figure 3.6: Unconditional and Counterfactual Effect of Job Training on Earnings for Natives and Immigrants (DFL Model). Dependent Variable: Log of Hourly Earnings.

Source: Data are from SIPP, and include individuals from year 1996, 2001 and 2004. Sample includes adults' male between 15 to 65 years of age. Note: DiNardo, Fortin and Lemieux (DFL) model conditions on yearly dummies, race (White, Black and Hispanic) dummies, seniority dummies, seniority squared divide by 100, years of education dummies, marital status dummies, dummies variable indicating whether the individual lives in the female head household, having children younger than 18 living in the family dummies, metropolitan, private firm, firm size dummies, dummy variable denoting possession of health insurance, union dummies, state dummies, industry dummies and occupation dummies.
Bibliography


all