UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THREE ESSAYS ON ECONOMICS OF LANGUAGE

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

CHEN XU
Norman, Oklahoma
2020

THREE ESSAYS ON ECONOMICS OF LANGUAGE

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF ECONOMICS

BY THE COMMITTEE CONSISTING OF

Dr. Le Wang, Chair

Dr. Deven Carlson

Dr. Daniel Hicks

Dr. Tyler Ransom

Dr. Chunbei Wang

Acknowledgments

I would like to express my sincere gratitude to my committee members: Dr. Le Wang, Dr. Deven Carlson, Dr. Daniel Hicks, Dr. Tyler Ransom, and Dr. Chunbei Wang, for their support and help through this long journey. This dissertation would not have been possible without their generous help. Special thanks to my advisor Dr. Le Wang for his mentorship, advice and guidance he provided while working on my doctorate degree. I learned from him what makes an excellent scholar.

I would also want to thank our department chair, Dr. Gary Hoover, for his continued support and encouragement. In addition, I want to thank other professors and colleges who have helped me during my Ph.D. study.

Finally, I want to thank my parents and my wife for their love and support. They have always believed in me. I also want to thank my baby boy Max, who has been a great motivation for me during this journey.

Contents

List of Figures

List of Tables

Abstract

My dissertation chapters examine the Economic returns to the proficiency on language use in different contexts. The first chapter challenges some of the standard assumptions imposed in the linear ordinary least model and quantile regression model in order to estimate the English Proficiency (EP) effect on Hispanic immigrants' wages. Instead, we impose several "weaker" assumptions to nonparametrically identify the returns. There are two contributions to the existing economics of language literature. One is to use the non-parametric bounds approach to partially identify the population average treatment effects over time. The second is to bound distributional treatment effect parameters. To the best of our knowledge, this is the first paper of the economic return to language literature that employs bounds on the distributional analysis, rather than solely focusing on average mean effect. By using 2000 census data and 2010 ACS data, we find that, on average, the overall impacts are positive in both years. Yet, the effects are heterogeneous on the entire distribution. The population in the middle and upper tails benefit more from having better language skills than those in the lower tail. Finally, we find that rates of English proficiency return increase over time for most parts of the earnings distribution.

By using the Pooled China General Social Survey (CGSS) data, the second chapter

examines whether proficiency in English leads to higher wages in the Chinese Labor Market. We employ a non-parametric bounding method to find out the lower bound and upper bound of the average treatment effect (ATE). The nonparametric bound approach reveals that the IV estimators of English proficiency (EP) suffer from an upward bias in the study of its return on wages. Our empirical findings show sizable returns to both English listening and speaking skills, and the return is unambiguously large when one's English skills improve from the lowest to the highest level. We also find significant heterogeneity in returns to English skills. Specifically, we find English skills matter more for females, residents with urban hukou, as well as prime-aged workers. Our results here corroborate the findings in this strand of the literature that English skills, as a form of human capital, can play an important role even if they are not the official language in the country.

To shed light on the effectiveness of promoting the use of Mandarin, the third chapter analyses the economic value of Mandarin in the Chinese labor market using the 2016 wave of China Labor-force Dynamics Survey(CLDS). To address the self selection issue, we use both instrument variable and non-parametric bounding approaches. Our results show that the overall return is positive, ranging from 10.5% to 49.9% for the whole sample. One driving mechanism we find is that workers with good Mandarin proficiency are more likely to find jobs that match their expertise, thereafter increase productivity at the workplace. We also find that the returns are heterogenous in different sub-groups. (1) Female benefit more than male, which also infers that the soft skill like language proficiency can help to reduce the gender wage gap; (2) The return is statistically significant and positive for young and mid-age workers, but shows no enough evidence for older workers; (3) The return is larger for the urban population than rural population. In addition, we find the IV method cannot extrapolate the marginal returns along with treatment groups, and the coefficient is incredibly large which overestimates the economics return to Mandarin proficiency.

Chapter 1

Non-parametric bounds on quantiles under monotonicity assumptions: an application on the return to language skills among Hispanics in U.S.

1.1 Introduction

Hispanics have been the primary driver of U.S. population growth, accounting for half of the national population growth since 2000 (Flores (2017)). Figure 1.1 visualizes the trend of growth of the Hispanics in the U.S. In 2000, there were approximately 35.66 million Hispanics in the U.S. while the number has rapidly evolved to 57.4 millions in year 2016. The numbers of both foreign-born and U.S. born Hispanics are growing, with a faster growth among U.S. born group (Flores, López and Radford (2017)). Much of the public debate has been expanded to discuss their integration and success to the host country.

One indication to show how successfully immigrants integrate to the host country is their performances in the labor market. Figure 1.2 shows the comparison of wages between the Hispanics and Whites over time from 2000 to 2016. The Hispanic workers earn significantly less than Non-Hispanic Whites overall. There is an apparent wage gap between these two groups. The wage gap is steady during this period, which is around 35-38% between male Hispanics and Whites. While for female Hispanics, the wage gap is even more wider, which is about 40-42%. In addition, one in four Hispanics was poor in 2009. Income inequality was higher among Hispanics than among non-Hispanic whites (Orrenius and Zavodny (2011)).

One channel has been widely discussed to explain the wage gap is language skills, for two dominant reasons. First, language skills would affect Hispanics' job search process and communication barriers can reduce their productivity at workplace (Kossoudji (1988)). Second, employers are more likely to discriminate against Hispanics with lower language proficiency (Grenier (1984)). As reported in Preston (2007): "Many Hispanics believe they will face discrimination if they speak Spanish and lack strong English skills. Poor English skills as the leading cause of discrimination against them, a far more significant cause than race or immigration status." According to Krogstad, Stepler and Lopez (2015), the proportion of those who speak English well is much smaller for the foreign-born than their US-born counterparts. What's more, the English proficiency of foreign-born Hispanics have hardly changed since 1980, shown in figure 1.3. However, on the other hand, the market structure has changed dramatically over time. "The decline in manufacturing employment, by about a third just since 1990. Meanwhile, employment in knowledge-intensive and service-oriented sectors, such as education, health, and professional and business services, has about doubled", (Center (2016)). The skill sets, for instance, communication skills and analytical skills are more demanded.

The language skills are indeed more important for foreign-born Hispanics. Although, the wage gap between the native-whites and the whole Hispanic group has not increased over time at first glance. This is very likely driven by the better performance of the native-born Hispanics. Among the whole Hispanic group, the growth of earning incomes is more stagnant for foreign-born Hispanics than their native-born counterparts (Massey and Gelatt (2010)). In this respect, if only comparing wages between whites and foreign-born Hispanics, the wage gap actually increases. Thus, it's important and necessary to examine the effects over time.

In terms of immigrant economic success and social well-being, the relationship between language and earnings has been the subject of considerable research. The ability to communicate smoothly with native English speakers is a critical factor contributing to immigrants' assimilation in the host country. Whether English education for the Hispanic immigrants is worthwhile depends largely on the market value of English skills. The questions are: (1) how important language acquisition affects Hispanics' labor market outcomes over time, for example their earnings; (2) if so, how large is the effect; (3) how do language skills affect different subgroups of people, and across the distribution.

Given the importance of better knowing the effect, researchers raised these questions back to 1980s, during the period while many immigrants were inflowing to U.S (Reimers (1983); McManus, Gould and Welch (1983); Grenier (1984); McManus (1985); Kossoudji (1988); Tainer (1988), and Chiswick (1991)). Among most of the existing literature, positive correlations are found between English proficiency (EP) and earnings. Yet, estimation of causal effects of English skills on earnings is, however, by no means a trivial task. The association may be a spurious one, capturing the effects of omitted variables, or reflect reverse causality.

The standard approach to estimate the effect of English skills on earnings is to include the index of EP in a linear model. Reimers (1983) uses a binary variable to measure EP, with 0 meaning bad English skill and 1 meaning good English skill, and finds significant effects (18-20%) of English skills on earnings for Hispanics. Grenier (1984) finds a smaller EP effect of around 15% among Hispanic males. McManus (1985) concludes that raising EP to perfect fluency would increase wages by 26% on average. All these early-stage studies treat selection as exogenous and treat the OLS estimates as the true causation between English skills and earnings. However, the sorting regime is not random. The simple OLS estimates do not take "ability bias" into consideration. If this is the case, then OLS estimates are contaminated.

Moreover, classical prescriptions for addressing the endogeneity issue in the OLS model, namely the linear instrumental variable model, may be untenable. The IV approach depends on the strength and validity of the IVs. In the context of our study, immigrant ages at arrival interact with a dummy for the non-English country, which was first proposed by Bleakley and Chin (2004). It is widely used as IV for English proficiency in the literature. However, immigrants who arrive at younger ages may face a lower cost of assimilation, which could weakly affect earnings; This questions the validity of the IV since it may not meet the requirement of mean-independence assumption.

The estimated mean effect, may not be informative about the effects for the rest of the population. Later on, with the development and popularity of Quantile Regression (QR), Wang and Wang (2011) adopts the instrument variable quantile regression(IVQR) approach to examine the heterogeneous returns to language skills across the earnings distribution among Hispanic immigrants. They also use immigrant ages at arrival as IV for English skills, and they find considerable heterogeneity in language premium across the earnings distribution. Likewise, the validity of IV is questionable. Furthermore, the both IV and IV quantile regressions impose homogeneity in the effects of English skills. We are not able to distinguish the different marginal effects with more than two groups. However, it is imperative to know the different marginal effects from the policy perspective.

Our discussions above are not meant to critique the current literature. Rather, we intend to point out the difficulties facing all the empirical analyses and call for alternative approaches. In this paper, we adopt a conservative approach, which is called Nonparametric bounds analysis. This approach is proposed in Manski (1990) and further developed in Manski (1997), Manski and Pepper (2000), and Manski and Pepper (2009). This approach has gained increasing popularity in empirical research across the fields in the past two decades [1]. It relies on transparent and rather weak assumptions, which are testable, to identify the causal effects. Also, it does not reply or any parametric form so that we can uncover the different marginal returns to English skills.

To perform our analysis dynamically, we utilize data from the 2000 census data and 2010 ACS (American Community Survey) data. Even when weak assumptions are imposed, we still reach several important findings. First, we find that the premium of being proficient in English is large among foreign-born Hispanics. Second, we find that the premium gains larger from 2000 to 2010. Third, we find significant heterogeneity in returns to English skills. Specifically, we find that English skills matter more for females, higher educated workers , residents in Metropolitan cities, as well as older workers. Finally, we also find there exists a

[1] For the applications, see Pepper (2000); Kang (2011) and Depalo (2018).

significant degree of heterogeneity in returns to English skills across the wage distribution. In particular, gains are more pronounced for individuals in the middle and upper tails of the distribution than those in the lower tail.

The rest of the paper is organized as follows: Section 2 discusses the data and descriptive statistics; Section 3 discusses the identification power on the mean and introduces different monotone assumptions on the mean analysis; Section 4 introduces monotone assumptions on the distribution analysis and discusses the findings, and section 5 concludes.

1.2 Data

Both 2000 Census of 1% population data and 2010 ACS Public Use Microdata Sample (PUMS) data of all 50 states in the United States are used in this study. We restrict the samples to those reported as having Hispanic origins and between 16 and 64 years old. We exclude self-employed and unpaid family workers and people who did not work in the previous year. We also exclude samples who report zero earnings and zero working hours. Here, we want to specifically focus on those born outside the USA because they seem to have more language barriers in the U.S compared to their native-born counterparts[2].

The treatment that each person receives is based on their English proficiency, which comes from a question that each respondent self-reported. There are five possible levels of English proficiency: t1 (do not speak English at all), t2 (speaks English not well), t3 (speaks English well), t4 (speaks English very well), and t5 (speaks only English at home). We group Hispanics in t4 and t5 groups as one group, because there is only a very little proportion

[2] For example, Gonzalez (2005) uses both native-born and foreign- born Hispanic in her study

of the population reported only speak English at home[3]. Though the subjective assessment on the levels of English proficiency has been widely used in the language-related literature, we admit its deficiency. Akbulut-Yuksel, Bleakley and Chin (2011) find that though self-reported answer may not be a perfect measure of language skills, it is "highly correlated with scores from tests designed to measure English language skills as well as functional measures of English language." Given that the limitation of information in our data, we adopt the same measure in our paper. The outcome variable is the logarithm form of the hourly wage. We calculate the hourly wage by dividing the total individual salary (Wages, salary, commissions, bonuses, or tips from all jobs) by the total weeks and hours per week worked in the prior year. For each person, we also observe and obtain a set of socioeconomic and demographic information, including age, working experience, education attainment, occupation, residency status. In our empirical analysis, we first test the average treatment effect with the full sample. Then we investigate the heterogeneous effect by introducing different categorical covariates (i.e., average treatment effect in the subsamples).

In a typical treatment - response analysis, our goal is to learn the "average treatment effect" (ATE). Here in our paper, our focus is on how much more the Hispanic immigrants could potentially earn if their English proficiency improves from one level to the next higher level, or even extremely, from the lowest proficiency level to the highest proficiency level, which is meant to be global effects hereafter in the paper. Thus, we seek bounds on differences between $\mathbb{E}[y(t_j)] - \mathbb{E}[y(t_i)]$ for any $t_i, t_j \in T$.

Before we impose any assumptions, we will start from the fundamentals and let the data

[3] In 2000 Census data, 4.5% of the sample was reported in group t5, and 4.16% of the sample was reported in group t5 in 2010 ACS data

and our economic sense to bound the possible range of $\mathbb{E}[y(t)]$. In 1999, the federal minimum hourly wage was \$ 5.15 (Bureau of Labor Statistics (BLS)), equivalent to the log hourly wage of 1.64. In our data, some people report their wages below this par. It is not our intention to lose any information by excluding those observations, so we set the lowest value of log hourly wage in our sample as \$1, which is noted as $\mathbb{K}_0$. We set the highest value of log hourly wage $\mathbb{K}_1$ equals to \$5, which corresponds to around \$150 per hourly wage. We choose \$5 as our supreme value based on the wage distribution in our data, as it is higher than 99% of the log hourly wage in our data, and the more extreme log hourly wages are outliers. Hence, we think \$5 is a reasonable value in our data[4].

According to the summary statistics shown in Table 1.1, the total number of observations in our sample are 54645 and 75653 in 2000 and 2010, respectively. From the table, we observe that most Hispanics report they have very well or above English proficiency. The distributions of the Hispanics in each group are barely changed in these two years, which is very consistent with the trend in Figure 1.3. Regarding the outcome variable, the average log wage in the lowest (highest) English proficiency group increases from 1.998 (2.415) in 2000 to 2.214 (2.712) in 2010. The average log wages rise with the levels of English proficiency, which indicates that Hispanic immigrants with better EP tend to gain higher wages in the U.S. labor market, on average[5]. Looking at education attainment, we observe that Hispanics with better English skills typically have more education, and those who have lower proficiency accumulate more working experience instead. Average years of education increase with

[4] Later we will show these arbitrary arbitrary values won't affect the final estimations.

[5] Here, we only emphasize the possible correlation between EP and wages, no causation is implied at this stage.

English skills, from 1.231 (1.311) to 2.257 (2.585) in 2000 (2010)[6], while average years of experience decrease with treatments, from 21.699 (22.095) to 16.550 (15.087). According to the table 1.1 , we find out better English proficiency, higher education attainment, or more working experience are all related to higher income wages among foreign-born Hispanics.

1.3 Empirical Specification

1.3.1 Identification Power on the Mean

The response function $y_j(\cdot) : T \to Y$ maps treatment into outcome. For each person, we have a response function , which maps treatment $t \in T$ into outcomes $y_i(t) \in Y$, where the treatment t is the level of English proficiency (EP), and Y is the log hourly wage. We are interested in the population average treatment effect of increasing EP from t_i to t_j, where $i \neq j$ and $t_i, t_j \in T$, on wage incomes. It is expressed as follows,

$$\triangle\,(t_i, t_j) = \mathbb{E}[y(t_j)] - \mathbb{E}[y(t_i)] \tag{1.1}$$

The average treatment effect consists of two parts: the mean wage we would observe if all Hispanic immigrants had EP level equals to t_j ($\mathbb{E}[y(t_j)]$), and the mean wage we would observe if they had EP level equals to t_i ($\mathbb{E}[y(t_i)]$). We will first focus on the mean wages we would observe if all Hispanics had the same level EP. Then we will be to look at the average treatment effect of increasing the EP level from one to another.

[6] Education has four possible values: 1. High school dropout, 2. High school and equivalent degree, 3. Some college experiences, 4. Bachelor degree and above.

According to the law of total expectation, we can decompose the expected value of the outcome in a certain group by two parts,

$$\mathbb{E}[y(t)] = \mathbb{E}[y(t)|z = t]\mathbb{P}(z = t) + \mathbb{E}[y(t)|z \neq t]\mathbb{P}(z \neq t) \tag{1.2}$$

The true population average wage for each treatment group is simply weighted by two groups, the first group is people who actually receive the treatment t, the second group is people who do not receive treatment t but what if they had received. According to the data we have, we could easily identify the values of $\mathbb{E}[y(t)|z = t]$, $\mathbb{P}(z = t)$ and $\mathbb{P}(z \neq t)$, but could not possibly figure out the part $\mathbb{E}[y(t)|z \neq t]$, which is also known as the counterfactual part. This unidentified part would bring us an issue on identification[7], and this is also the main focus that many researchers would like to impose assumptions on. Without further assumptions, we cannot know the population mean of $\mathbb{E}[y(t)]$, or to identify the average treatment effect neither.

From equation (1.2), we see that we are not able to get the average wages for the population. Instead, Manski (1989) shows that, it is possible to identify bounds on $\mathbb{E}[y(t)]$ without adding any assumption if the support of the outcome variable is bounded. We can obtain the worst case bounds for $\mathbb{E}[y(t)]$ by simply substituting the latent part $\mathbb{E}[y(t)|z \neq t]$ with the minimum value of y (thereafter, the notation is $\mathbb{K}_0$) and the maximum value of y (thereafter, the notation is $\mathbb{K}_1$), we can explicitly express the lower bound and the upper bound of $\mathbb{E}[y(t)]$

[7] Suppose we want to study the difference in expected outcomes if all persons are from certain treatment group A to another treatment group B, then $\mathbb{E}[y_B] - \mathbb{E}[y_A] = \mathbb{E}[y_B|z = B]\mathbb{P}(z = B) + \mathbb{E}[y_B|z = A]\mathbb{P}(z = A) - \mathbb{E}[Y_A|z = A]\mathbb{P}(z = A) + \mathbb{E}[y_A|z = B]\mathbb{P}(z = B)$, the sample process could not identify the latent outcomes,$\mathbb{E}[y_B|z = A]$ and $\mathbb{E}[y_A|z = B]$

accordingly. Then equation (1.2) would be bounded as follows,

$$\mathbb{E}[y(t_i)|z=t_i]\mathbb{P}(z=t_i)+\mathbb{K}_0\mathbb{P}(z\neq t)\leq\mathbb{E}[y(t_i)]\leq\mathbb{E}[y(t_i)|z=t_i]\mathbb{P}(z=t_i)+\mathbb{K}_1\mathbb{P}(z\neq t) \quad (1.3)$$

These worst-case bounds are not very informative because they are generally very wide. Still, we could gain information from these bounds since they should give us the overall range of $\mathbb{E}[y(t)]$ for each treatment level, so that the results that are based on different assumptions about $\mathbb{E}[y(t)|z\neq t]$ should lie within this region. In the following subsections, we will introduce monotone treatment response (MTR) assumption and monotone treatment selection (MTS), and add them into our analysis. While in the IV framework, we will introduce how the monotone instrument variable assumption and its combination with MTS and MTR make these bounds more informative. The MTR and MTS assumptions are introduced and derived by Manski (1997) and Manski and Pepper (2009). All these assumptions are used to find the plausible regions of the latent part in equation (1.2).

1.3.2 Assumption on Response Function (MTR)

The MTR assumption states that the outcome is a weakly increasing function of the treatment:

$$t_j\geq t_i\rightarrow y(t_j)\geq y(t_i) \quad (1.4)$$

equation (1.4) states that increasing one's English proficiency weakly increases his or her wage function, which imposes that improving Hispanic immigrants' language ability would never hurt their job performance, ceteris paribus. This assumption is originally derived from the human capital theory raised by Becker (1964), which states that language is a type of human capital, and people tend to be more productive with better language skills. There are lots of reasons to expect a positive impact of increasing EP on wages for immigrants. In the extant literature, the most convincing argument is that better EP indicates that immigrants not only suffer fewer assimilation problems but also have fewer barriers in communicating and working with native Americans. Note, it could, however, be the case that increasing EP does not necessarily improve wages. The positive associations that have been found may be due to positive selection to jobs and not causation. Therefore, it is important to point out that the MTR assumption does not rule out the possibility of EP being not productive at all.

The most common case for OLS is based on the "linear response assumption", which means that the hypothetical response of each individual on the treatment is assumed to be homogeneous. However, this "linear response assumption" is not applicable in this case, because the measure of EP is discrete. With the MTR assumption, we are able to loosen the OLS assumption.

After we combine the MTR assumption along with the worst-case bounds discussed earlier, equation (1.3) becomes:

$$\mathbb{E}[y(t_i)|z \leq t_i]\mathbb{P}(z \leq t_i) + \mathbb{K}_0\mathbb{P}(z > t_i) \leq \mathbb{E}[y(t_i)] \leq \mathbb{E}[y(t_i)|z \geq t_i]\mathbb{P}(z \geq t_i) + \mathbb{K}_1\mathbb{P}(z \leq t) \tag{1.5}$$

To show how the MTR assumption can be used to tighten the worst-case bounds, we can divide our sample into three parts: immigrants who have EP (1) lower than t($z < t$), (2) equals to t(z=t), and (3) higher than t($z > t$). Then, we are able to use the mean wages we observe for the first group to tighten the lower bound and use the mean wages we observe for the third group to tighten the upper bound. The bounds estimated from equation (1.5) are shown in Table 1.3. The 95% confidence intervals are drawn by the bootstrap method from 99 repetitions. Comparing with the worst-case bounds shown in table 1.2, we can see that the bounds are much narrower by imposing the MTR assumption along with the worst-case assumption, especially it helps to narrow down the lower bounds. According to Table 1.3, the log hourly wage for people who do not speak English at all is bounded between 1.153 and 2.245 in 2000 (between 1.162 and 2.497 in 2010), and the log hourly wage for people who have the best EP is bounded between 2.245 and 4.133 in 2000 (between 2.497 and 4.237 in 2010). The lower bound of the average treatment effect is restricted at zero, which happens when language does not have any effect on immigrants' wages, given the Monotone Treatment Response (MTR) assumption.

1.3.3 Assumption on Treatment Selections

The second assumption that we impose on the analysis is the Monotone Treatment Selection (MTS) assumption. Under this assumption, immigrants with better EP have weakly higher mean wage functions than those with lower EP. This can be mathematically shown as follows,

$$t_j > t_i \Longrightarrow \mathbb{E}[y(t_u)|z = t_j] \geq \mathbb{E}[y(t_u)|z = t_i] \tag{1.6}$$

This assumption relates to the sorting process, which gives us a better scenario on how the treatment groups are distributed to the different individuals in the population. As discussed earlier, language is one type of human capital. Immigrants who have the better EP should have higher ability and, therefore, higher wages, on average. Better EP also indicates less assimilation problems, such as the cost of communication and collaboration with natives, which can improve the quality of job match. Both arguments are consistent with the MTS assumption in our context, but they do not work appropriately with the Exogenous Treatment Selection (ETS) assumption, which is commonly assumed under an OLS framework. The underlying idea of ETS assumes that treatments are assigned to the population at random, but it does not seem to be the case in the context. The ETS estimates are equivalent to the coefficients obtained by running OLS on log earnings with one dummy variable for each treatment group without adding any covariates. The expected values for each treatment group under ETS are shown in Table 1.3.

To better understand and use the MTS assumption, we divide the sample into three parts - immigrants who have EP (1) lower than t($z < t$), (2) equals to t(z=t), and (3) higher than t($z > t$). Since the EP of the first group is below t, we can infer that the mean wage level in the first group would be weakly lower than the mean wage of immigrants with EP equal to t, from the assumption of MTS. Thus, we are able to use the mean wage level that we observe from the immigrants who have EP at level t as an upper bound for the first group. Applying the same logic, we use the mean wage from immigrants with EP at level t as the lower bound for the third group. Hence, we construct the new bound under the MTS assumption [8] as

[8] See Manski and Pepper (2000) for the full derivation

follows,

$$\mathbb{E}[y(t)|z=t]\mathbb{P}(z \geq t) + K_0\mathbb{P}(z \leq t) \leq \mathbb{E}[y(t)] \leq \mathbb{E}[y(t)|z \geq t]\mathbb{P}(z \leq t) + K_1\mathbb{P}(z \geq t) \quad (1.7)$$

and the bounds of expected wage for each treatment effect under MTS assumption is shown in table 1.3.

1.3.4 More Identification power: MTR + MTS Jointly

If we combine the assumptions of MTR and MTS, we basically assume that the response function is monotone and that selection into treatment is positive at the same time. Both assumptions are consistent with human capital theories as discussed above. By combining the two assumptions, we obtain the MTR-MTS bounds [9] as follows,

$$\begin{gathered}\mathbb{E}[y(t)|z<t]\mathbb{P}(z \leq t) + \mathbb{E}[y|z=t]\mathbb{P}(z>t) \\ \leq \mathbb{E}[y(t)] \leq \\ \mathbb{E}[y(t)|z=t]\mathbb{P}(z \leq t) + \mathbb{E}[y|z>t]\mathbb{P}(z>t)\end{gathered} \quad (1.8)$$

As we can see from equation (1.8), once we impose the assumptions together, K_0 and K_1 are eliminated. Here, it illustrates why the arbitrarily selected K_0 and K_1 will not affect the estimations once we impose more assumptions. Also, one advantage of imposing the MTR-MTS jointly is that it is a testable assumption. Under this assumption, the following

[9] For a full derivation of the MTR-MTS bounds, see Manski (1997) and Manski and Pepper (2000)

should hold[10] for $t_j > t_i$,

$$\mathbb{E}[y|z = t_j] = \mathbb{E}[y[t_j]|z = t_j] \geq \mathbb{E}[y[t_j]|z = t_i] \geq \mathbb{E}[y[t_j]|z = t_i] \\ \geq \mathbb{E}[y[t_i]|z = t_i] = \mathbb{E}[y|z = t_i] \tag{1.9}$$

In order to impose them simultaneously, the mean wages of immigrants should increase with the hypothetical increase in English proficiency. If this is not the case, the joint assumption of MTR and MTS should be rejected anyway.

1.3.5 Monotone Instrumental Variable Assumption

Endogenous variables are often encountered in empirical work. In order to identify these variables' causal effects on outcomes, scholars often rely on some forms of assumption. In the setting of the traditional Instrument Variable (IV) approach, suppose a variable v satisfies the IV assumption. Then, under the condition of mean independence, the following should hold for all treatments $t \in T$ and all values of the instrument $m \in M$,

$$\mathbb{E}[y(t)|v = m] = \mathbb{E}[y(t)] \tag{1.10}$$

This equation means that the immigrants' wage function should be mean independent of the variable v. Since it is not easy to find a valid IV which satisfies equation (1.10). We will use a weaker version: the monotone instrumental assumption. MIV is a relatively weaker version of IV proposed by Manski and Pepper (2000). A variable v is a MIV in the sense of

[10] The first inequality derives from the MTS assumption and the second inequality derives from the MTR assumption

mean monotonicity if it holds that, for $m_1 \leq m \leq m_2$,

$$\mathbb{E}[y(t)|v = m1] \leq \mathbb{E}[y(t)|v = m] \leq \mathbb{E}[y(t)|v = m2] \tag{1.11}$$

As opposed to assuming that mean responses are constant across an instrument, the MIV assumption allows for a weakly monotone relation between the variable v and the mean wage function of the immigrants. The idea of Bleakley and Chin (2004) IV is to use age at arrival interacted with a dummy for the non-English-speaking country as the identifying instrument. This IV has been popularly used in the immigrants' economics return to human capital related literature after then[11]. Although the credibility of this IV is not consistent in the literature, one argument is that one's age at arrival can affect her later on wages may not only through her language skill but also could through other channels such as ability to assimilate to the culture in the host country. However, this IV could still be a good option for our paper since it satisfies the setting of Monotone Instrument Variable (MIV). It ensures that immigrants' potential wages are weakly decreasing along with their ages at arrival. Hispanic Immigrants who arrive at the U.S. at younger ages have weakly higher wage functions than those who arrive at older ages, and this is attributed to a lower assimilation cost. The MTS assumption is, in fact, a special case of the MIV assumption. The significance of language for economic adaptation of immigrants derives from the fact that language is assumed to be more easily alterable than other forms of human capital, such as educational attainment, even though it generally takes 5 to 7 years to become fluent in English (Hakuta,

[11] This IV is highly adopted in the related literature. For examples, Ferrer, Green and Riddell (2006); Böhlmark (2008); Myers, Gao and Emeka (2009) and Blau, Kahn and Papps (2011) among others.

Butler and Witt (2000)). To use the MIV assumption, we can again divide the sample into sub-samples on the basis of v and obtain bounds for each subsample. From equation (1.11), it follows that $\mathbb{E}[y(t)|v = m]$ is no lower than the lower bound on $\mathbb{E}[y|v = m_1]$ and it is no higher than the upper bound on $\mathbb{E}[y|v = m_2]$. For the subsample with v equals to m, we can obtain the lower bound, which is the largest lower bound over all the subsamples in which v is lower than or equal to m. Similarly, we can obtain the upper bound over all subsamples with a value of v higher than or equal to m. Once we impose the MIV assumption along with MTR and MTS, the inequality equation for $E[y(t)]$ is as follows.

$$\sum_{v\in m} \mathbb{P}(v = m)\{sup_{m_1\leq m}(\sum_{k<t} \mathbb{E}[y|v = m_1, z = k]\mathbb{P}(z = k|v = m_1) + \mathbb{E}[y|v = m_1, z = t]\mathbb{P}(z \geq t|v = m_1))\} \leq$$

$$E[y(t)]$$

$$\leq \sum_{v\in m} \mathbb{P}(v = m)\{inf_{m_2\geq m}(\sum_{k>t} \mathbb{E}[y|v = m_2, z = k]\mathbb{P}(z = k|v = m_2) + \mathbb{E}[y|v = m_2, z = t]\mathbb{P}(z \leq t|v = m_2))\} \tag{1.12}$$

1.3.6 Average Treatment Effect

In any treatment-response analysis, we are always interested in knowing the average treatment effect (ATE), which is defined as $\mathbb{E}[y(t_j)] - \mathbb{E}[y(t_i)]$, where $j > i$. From the above section, we have discussed how to find the lower bound and upper bound of $\mathbb{E}[y(t)]$ under different sets of assumption. Now, we can find the lower bound and upper bound of the average treatment effect, accordingly. we subtract the lower bound of $\mathbb{E}[y(t_j)]$ from the upper bound of $\mathbb{E}(y(t_i)]$ to obtain the lower bound of the average effect, while we subtract the upper bound of $\mathbb{E}[y(t_j)]$ from the lower bound of $\mathbb{E}(y(t_i)]$ to obtain the upper bound of the average effect.

To better understand the global effect of language, we focus on both EP changing from "no English skill at all" to "well" and from "no English at all" to "speak English very well", additionally. In other words, we focus on the returns of improving EP from t_1 to t_3 and from t_1 to t_4. The estimated ATEs and the bootstrapped confidence intervals are shown in Table 1.2,1. 3, and 1.4 under different sets of assumptions. Since we already set the infimum and supreme of log wage at 1 and 5. The no-assumption bounds should belong to [-4, 4]. As we can see from the table 1.2, the worst-case bounds are also very wide and uninformative. The results show that the ATE(t_4-t_1) could be anywhere between -3.066 and 2.980 (between -3.058 and 3.075) in 2000 (2010). As the MTR assumption being imposed, the lower bounds from each treatment t_i to another treatment t_j (where $t_j > t_i$) are no longer smaller than zero [12]. The upper bound for ATE (t_3-t_1) is 2.342 (2.461), and 2.980 (3.075) for ATE (t_4-t_1) in 2000 (2010), respectively. Although the width of the bounds has reduced compared with the worst-case bounds, the MTR bounds are still too wide to be really informative. This is also the case of imposing MTS alone, which also reduces the width of the worst-case bounds, and particularly causes the shrinkage of the upper bounds of the worst-case. Yet, it remains to be not informative enough to make any inferences. By combining MTR and MTS, we can see from table 1.3 that the bounds are much tighter than the ones from imposing them separately. The worst-case bounds are narrowed down to [0, 0.342] ([0, 0.379]) for ATE(t_3-t_1), and to [0, 0.417] ([0, 0.498]) for ATE(t_4-t_1). So far, the global wage premium from no English at all to perfect English proficiency is up to 41.7% and 49.8% in two years, which seems very more likely to compare with the point estimations.

Since MTS + MTR is a testable hypothesis, we can conclude that all of the levels are not rejected at the 95% confidence interval, and the bounds under MTS + MTR are statis-

[12] Realize that assuming MTR will immediately implies that the lower bound is not smaller than zero.

tically reliable. When we bring the IV framework into consideration, we jointly impose MIV with MTR + MTS, and the results are shown in table 1.4. Thus, we can conclude that the bounds of ATE (t_3-t_1) and ATE (t_4-t_1) are narrowed to [0, 0.321] ([0, 0.351]) and [0, 0.399] ([0, 0.449]), respectively. Thus, monotone instrument variable provides identification power on the estimation of the upper bounds. Other than the global effect, it is also meaningful to analyze the average treatment effect from one treatment to the next level of treatment. Under MIV + MTR + MTS, we can see the pattern that the maximum wage gains vary along with the treatments. Although the lower bounds are always zero, the upper bounds could provide some fruitful insights. In terms of the upper bounds of the ATE, the marginal returns from lower ends (t1 to t2) are higher than the returns from the upper end (t3 to t4) in both years. In the year 2000, EP has a clear pattern of diminishing marginal return to wages along with the treatment groups. This somehow shows the different patterns from the previous literature, Gonzalez (2005) uses 1990 PUMS data and found increasing marginal returns regarding the upper bound, from 11.22% at the lower end to 32.17% at on the other hand. In addition, the overall penalty of being lack of English proficiency among Hispanic Immigrants has become more severe over time.

1.3.6.1 Introducing Covariates

So far, we have estimated the bounds effect over the whole sample without considering any covariates into the model. Up to this point, the bound estimations are homogeneous in the sample. In order to learn how EP may have different penalties or rewards in different subgroups, we consider adding covariates into our analysis.

First, we are interested in learning how English Proficiency plays a role in different educational groups. As many researchers have found that the wage penalty due to limited language proficiency increases with education attainment (McManus, Gould and Welch (1983); Carliner (1996); Mora and Davila (1998)). We divide our sample into three subgroups based on their reports on the highest education achievement. The three groups are Hispanics without a high school diploma (high-school dropouts), with just a high school diploma, and with some college experiences and above. Intuitively, jobs with the requirement of a high education usually require less physical effort, and jobs which do require a higher education usually need both physical effort and mental effort. Hence, this may require workers master some skills such as good communication skills. Later in the paper our results prove this hypothesis. Table 1.5 shows the bounds of the returns under different assumptions based on three levels of education. Under ETS, the returns are strictly increasing from gaining better English skills for groups of the high school graduates and above in both years. However, the marginal returns are not always positive in the lowest level group. As we discussed earlier, the MTR + MTS assumption is testable. The condition for MTR + MTS assumption is valid, only if $E[y(t)|z = t]$ is weakly increasing in t. In Table 1.5, we observe that this assumption is not valid for the high school dropout group, but it is valid for the other two groups in both years. Since the MTR + MTS hypothesis is rejected for the high school dropouts group, we want to figure out why this fails for this certain group, but not the other. Under the human capital theory, we can almost rule out the possibility that increasing one's language ability will lower his or her wage, ceteris paribus, which confirms that the MTR assumption is valid. However, the MTS assumption may not hold for lower education groups. The sorting to the treatment is not monotone for lower education groups, which implies that even the Hispanics have some good English skills, but they may still land in lower-paid job positions due to low educational achievement. In this sense, we can conclude that the negative ability bias causes the MTS assumption to fail. For the lower education group, Hispanic workers who speak poorer English do not necessarily have their wage function less than their counterparts who speak better English, because low-paid jobs do not require high English proficiency. Under MTR+MTS+MIV, the overall returns for high school graduates and with at least some college experiences and above are 28.3% (22.4%) and 42.5% (45.7%) in year 2000 (2010), implying that the role of education in English return is getting more important over time. We cannot rule out zero effects for all education levels under these conservative assumptions. Thus, the results are drawn from the previous studies may be due to higher ability bias in higher education levels.

We also show the language return on wages for different age groups of Hispanic Immigrants. We divide the whole sample into two groups based on their ages, one group is between age 16-45, and another group is between age 46 and 64. Under MTS+MTR assumptions, we cannot exclude the zero effects from our estimations for both age groups. Compare the upper bounds in particular, older cohorts have larger language return (up to 66% for global effect) than the younger cohorts (with the effect is 39.5%). What's more surprising finding is that at each margin, the effect is larger for the older age group than the younger age group corresponding to the upper bound effects. This finding can be explained by the conclusion that draw in Borjas (2015). He concludes that younger cohorts experience less economic assimilation than the older cohorts due to a concurrent decline in the rate at which the new immigrants add to their human capital stock, as measured by English language proficiency. Borjas also concludes that the rapid growth in the size of specific national origin groups in the United States is the big factor that causes the decline in both the rate of economic assimilation and the rate of English language acquisition. The larger the preexisting size of the national origin group in the United States, the less likely newer and young immigrants are less likely to assimilate and invest in human capital. While Hispanic is the third largest group after White and Black. This declining phenomenon should be more obvious. Borjas (2000), Xie and Gough (2011) and many others find that ethnic enclave participation actually harms economic outcomes.

Besides education and age groups, we also consider taking into account for other covariates. For example, gender and living location. The summary of bound estimations for different sub-groups under different assumptions is shown in table 1.7 and 1.8. Under ETS, we conclude that the returns to EP in terms of wages have a similar pattern for both genders. Female Hispanics experience a higher growth of return than their male counterparts from 2000 to 2010. Under MTR+MTS+MIV, the overall return is 48.2% for female, but only 42.6% for male in 2010, while in 2000, both male and female have the overall returns at the same level. Regarding the location of immigrants' residency, English skills are more important for those live in the metropolitan areas than non-metropolitan areas. In 2000, the global return is 41.9% in the metropolitan areas and increases to 47.2% in 2010. In comparison, the return is up to 36% in the non-metropolitan areas in 2000 and 37.2% in 2010.

1.4 Distribution Analysis

Thus far, we have focused solely on the mean effects of English proficiency on average individual earnings. These effects are potentially heterogeneous along with the sample distribution since individuals may have different marginal benefits from investments in language skills. Unless we are willing to assume constant effects across the earnings distribution, the estimated mean effect, while of interest in the earlier section, may not reflect the features of the other parts on the distribution.

In the following section, we take a step further to examine the heterogeneous returns to EP across immigrants earnings distribution. This allows us to explicitly assess who benefits most from improved language skills - immigrants in the lower tail, the upper tail, or the rest of the distribution. Analysis of heterogeneous language effects across the earnings distribution is of particular interest to policymakers. The parametric point identification on quantile regression, which was first proposed by Koenker and Bassett Jr (1978), has been widely used and developed in the past few decades. The quantile approach emphasizes the importance of the effect on the entire distribution rather than only the mean effect, and it also does a better job of taking outliers into account than OLS does. Moreover, this approach allows for more general behavior of regressors' effect on the outcome variable (Koenker (2000)).

Now, our focus is to identify the α-quantile of a distribution of interest $P[y(t)]$, denoted as $Q_\alpha[y(t)]$[13]. The major difficulty in predicting which distribution of outcomes P[y(t)] is that every individual were to receive the particular treatment t is the fact that one cannot observe the outcomes that a person would experience under treatments different from the received one (simply saying, the non-observability of counterfactual outcomes), by introducing the identifying properties of two main assumptions: $\alpha-$quantile monotone treatment selection (α-QMTS) and $\alpha-$quantile monotone instrument variables (α-QMIV). These two new assumptions apply the functions of MTS and MIV, respectively, to condition on the

[13] More Details, see Manski (1999)

distribution of the form $Q_\alpha[y(t)|z]$ and $Q_\alpha[y(t)|v]$. The main results are parallel to the results of mean estimates obtained from earlier sections. From now on, we will concentrate on achieving partial identification of the α-quantile of distribution of potential outcomes $Q_\alpha[y(t)]$ and α-quantile treatment effects (α-QTEs).

1.4.1 Identification of Quantiles under Empirical Evidence Alone

An upper (lower) bound on $Q_\alpha[y(t)]$ shares the same ideal with a quantile function. It is derived by first determining the lower (upper) bound on the corresponding distribution function $P[y(t) \leq \tilde{y}_\alpha]$, and then inverting it. Manski (1994) shows that, in the no-information or empirical-evidence-alone case, this procedure yields the bounds stated as follows [14],

Let $\alpha \in$ (0,1). Define $l^{EE}(\alpha, t)$ and $u^{EE}(\alpha, t)$ as,

$$l^{EE}(\alpha, t) = \begin{cases} Q_{[1-\frac{1-\alpha}{P(z=t)}]}(y|z=t) & \text{if P(z=t)} < \alpha < 1, \\ K_0 & \text{otherwise.} \end{cases} \tag{1.13}$$

$$u^{EE}(\alpha, t) = \begin{cases} Q_{[1-\frac{\alpha}{P(z=t)}]}(y|z=t) & \text{if } 0 < \alpha \leq \text{P(z=t)}, \\ K_1 & \text{otherwise.} \end{cases} \tag{1.14}$$

EE is an analogy analysis of the worst-case assumption but applied on different quantiles. Note that lower and upper bound on α-QTE of the type $Q_\alpha[y(t_j)] - Q_\alpha[y(t_i)]$,where j$\neq$ i, can be computed as LB= $l^{EE}(\alpha, t_j) - u^{EE}(\alpha, t_i)$ and UB=$u^{EE}(\alpha, t_j) - l^{EE}(\alpha, t_i)$.

[14] Proof: see Manski and Sims (1994)

1.4.2 Identification of Quantiles under MTR

From the previous mean level section, we know the assumption of MTR implies that the average wage function is weakly increasing in EP. In order to demonstrate how identification applies in the distributional case, for simplicity purposes, we assume that there are only two levels of English proficiency - good (t_2) and bad (t_1), and individuals can only receive either t_1 or t_2. In particular, we are interested in learning what the wage distribution would look like under the hypothetical uniform treatment rule assigning t_2 to everyone in the population. The difficulty is that without prior information, we only observe outcomes $y_i(z_i) = t_2$ contribute to identification of $P[y(t_2)]$. Once imposing MTR, we will be able to infer the counterfactual wages of individuals with poorer English proficiency t1. This, in turn gives us tighter bounds for $P[y(t_2) \leq \tilde{y}_\alpha]$ than under EE, since a logical upper bound for $P[y(t_2) \leq \tilde{y}_\alpha | z = t_1]$ is now given by $P[y(t_2) \leq \tilde{y}_\alpha | z = t_2] \leq 1$, while under EE, it assumes $P[y(t_2) \leq \tilde{y}_\alpha | z = t_1] = 0$. The general result under assumption MTR is as follows,

Let T be an ordered set of treatments. Let the response function $y_j(\cdot)$ with $j \in J$ be non-decreasing on T. Let $\alpha \in (0, 1)$,

$$l^{MTR}(\alpha, t) = \begin{cases} Q_{[1-\frac{1-\alpha}{P(z \leq t)}]}(y|z \leq t) & \text{if P(z>t)} < \alpha < 1, \\ K_0 & \text{otherwise.} \end{cases} \tag{1.15}$$

$$u^{MTR}(\alpha, t) = \begin{cases} Q_{[1-\frac{\alpha}{P(z \geq t)}]}(y|z \geq t) & \text{if } 0 < \alpha \leq \text{P(z} \geq \text{ t)}, \\ K_1 & \text{otherwise.} \end{cases} \tag{1.16}$$

It can be shown that for any value of α, a necessary but not sufficient condition for both

$l^{MTR}(\alpha, t)$ and $u^{MTR}(\alpha, t)$ to be informative is that there exists a positive fraction of the population receiving treatment t, $P(z = t) > 0$. Based on equations (1.15) and (1.16), we can see the sharp bounds $\forall j \in J$ and $t_1, t_2 \in T$ under MTR are given by $0 \leq Q_\alpha[y(t_2)] - Q_\alpha[(t_1)] \leq \mathbb{K}_1 - \mathbb{K}_0$, such that $y_j^{INF}(t) = y_j$ if $z_j \leq t$ and $y_j^{INF}(t) = \mathbb{K}_0$ otherwise; while $y_j^{SUP}(t) = y_j$ if $z_j \geq t$ and $y_j^{SUP}(t) = \mathbb{K}_1$ otherwise.

1.4.3 Introducing $\alpha-$Quantile Monotone Treatment Selection

By mimicking the idea of MTS, the α-quantile MTS assumes that: Suppose T is an ordered set and $\alpha \in$(0,1), then, for $\forall t_1, t_2 \in T$, $t_2 \geq t_1 \longrightarrow Q_\alpha[y(t)|z = t_2] \geq Q_\alpha[y(t)|z = t_1]$. The key to derive the identification result for $Q_\alpha[y(t)]$ under α-QMTS is to realize that $P[y(t)|z = t]$ represents a logical upper bound for all $P[y(t) \leq \tilde{y}_\alpha|z = \tilde{t}]$ such that $\tilde{t} \geq t$ and a lower bound for all $P[y(t) \leq \tilde{y}_\alpha|z = \tilde{t}]$ such that $\tilde{t} \leq t$, which yield the following identification region for $P[y(t) \leq \tilde{y}_\alpha]$

$$
\begin{aligned}
&P[y \leq \tilde{y}_\alpha|z = t]P(z \leq t) \leq \\
&\qquad\qquad P[y(t) \leq \tilde{y}_\alpha] \\
&P[y \leq \tilde{y}_\alpha|z = t]P(z \geq t) + P(z < t)
\end{aligned}
\tag{1.17}
$$

The corresponding bounds on $Q_\alpha[y(t)]$ are achieved by inverting equation (1.17); lower (upper) bound of $P[y(t) \leq \tilde{y}_\alpha]$ is the upper (lower) bound of $Q_\alpha[y(t)]$ for a certain treatment group t. They can be formally shown as follows, Let $\alpha \in (0,1)$, and define $l^{QMTS}(\alpha,t)$ and $u^{QMTS}(\alpha,t)$ as,

$$l^{QMTS}(\alpha,t) = \begin{cases} Q_{[1-\frac{1-\alpha}{P(z\geq t)}]}(y|z=t) & \text{if P(z<t)< } \alpha \text{ <1,} \\ K_0 & \text{otherwise.} \end{cases} \quad (1.18)$$

$$u^{QMTS}(\alpha,t) = \begin{cases} Q_{[1-\frac{\alpha}{P(z\leq t)}]}(y|z=t) & \text{if 0< } \alpha \leq \text{ P(z} \leq \text{ t),} \\ K_1 & \text{otherwise.} \end{cases} \quad (1.19)$$

Then for each t $\in$ T, the conditional quantile of outcome is bounded by,

$$l^{QMTS}(\alpha,t) \leq Q_\alpha[y(t)] \leq u^{QMTS}(\alpha,t)$$

1.4.4 Generalizing to $\alpha-$ Monotone Instrument Variable

Instead of the sense of mean monotonicity as MIV implies on the wage function, α-MIV somehow mimics the idea of MIV and assumes that the wage function is weakly increasing on the distribution for immigrants who arrived at the U.S. at younger ages than those who landed at later ages. Assumption (α-QMIV). Let V be an ordered set of instruments,

then $\forall v_1, v_2 \in$V,

$$v_2 \geq v_1 \longrightarrow Q_\alpha[y(t)|v=v_2] \geq Q_\alpha[y(t)|v=v_1]$$

The identifying power of α-MIV is established as follows,

If α-QMIV assumption holds, then for each v$\in$V, the identification region for $P[y(t) \leq \tilde{y}_\alpha]$ is

$$\sum_{v \in V} P(v = \boldsymbol{v}) P[y \leq \tilde{y}_\alpha | v = \hat{v}^\star(v), z = t] P(z = t | v = \hat{v}^\star(v)) \leq$$

$$P[y(t) \leq \tilde{y}_\alpha]$$

$$\leq \sum_{v \in V} P(v = \boldsymbol{v}) P[y \leq \tilde{y}_\alpha | v = \hat{v}^{\star\star}(v), z = t] P(z = t | v = \hat{v}^{\star\star}(v)) + P(z \neq t | v = \hat{v}^{\star\star}(v)) \tag{1.20}$$

where $\hat{v}^\star(v)$ is the supreme value of $P[y \leq \tilde{y}_\alpha | v = \hat{v}(v), z = t] P(z = t | v = \hat{v}(v))$, for $\forall \hat{v} \geq v$, and $\hat{v}^{\star\star}(v)$ is the infimum value of $P[y \leq \tilde{y}_\alpha | v = \hat{v}(v), z = t] P(z = t | v = \hat{v}(v)) + P(z \neq t | v = \hat{v}$, for $\forall \hat{v} \leq v)$ with $Q_\alpha[y(t)]=$ min $\tilde{y}_\alpha$ s.t. P$[y(t) \leq \tilde{y}_\alpha] \geq \alpha$[15]. Intuitively speaking, the ideals of the function of α-MIV are the following. First, given some fixed level of EP, an individual's wage is weakly increasing if he or she arrives at a younger age compared to when he or she arrives at an older age. Second, the distribution of language ability among immigrants who arrive at a younger age, weakly dominants the distribution of language ability among those who arrive at a relatively older age. On the other hand, under the setting of α-quantile, we may say that age at arrival has no impact or positive impact on an immigrant's wage. In other words, this IV (age at arrival) could not represent the unobservable confounding factors (such as the assimilation cost), but it is a weakly positive prediction of it.

1.4.5 Imposing Monotone Assumptions Jointly

In this section, we want to parallel to the work that we discussed in the mean-level analysis section. We impose MTR and QMTS assumptions jointly, which would help the identifi-

[15] More Details: see Giustinelli (2011)

cation region for the distribution P[y(t)$\leq \tilde{y}_\alpha$] shrink significantly compared to the bounds when we impose them separately. The findings are consistent with the results from the mean analysis discussed in earlier sections. Formally, by the law of total probability, we have the following inequality equation,

$$\begin{aligned} P[y(t) \leq \tilde{y}_\alpha | z \geq t]P(z \geq t) + \sum_{t'<t} P[y(t) \leq \tilde{y}_\alpha | z = t']P(Z = t') \\ \leq P[y(t) \leq \tilde{y}_\alpha] \leq \\ P[y(t) \leq \tilde{y}_\alpha | z \leq t]P(z \leq t) + \sum_{t''>t} P[y(t) \leq \tilde{y}_\alpha | z = t'']P(Z = t'') \end{aligned} \tag{1.21}$$

From QMTS, we know that $P[y \leq \tilde{y}_\alpha | z = t]$ is the lower bound for $P[y(t) \leq \tilde{y}_\alpha | z = s]$ for all $s = s' < t$, and an upper bound for it for all $s = s'' > t$. From MTR, we know that $P[y(t) \leq \tilde{y}_\alpha | z \geq t]$ is weakly greater than $P[y \leq \tilde{y}_\alpha | z \geq t]$, and $P[y(t) \leq \tilde{y}_\alpha | z \leq t]$ is weakly smaller than $P[y \leq \tilde{y}_\alpha | z \leq t]$. Then equation (21) becomes the form as follows[16],

$$\begin{aligned} P[y(t) \leq \tilde{y}_\alpha | z \geq t]P(z \geq t) + P[y \leq \tilde{y}_\alpha | z = t]P(z < t) \\ \leq P[y(t) \leq \tilde{y}_\alpha] \leq \\ P[y(t) \leq \tilde{y}_\alpha | z \leq t]P(z \leq t) + P[y(t) \leq \tilde{y}_\alpha | z = t]P(z > t) \end{aligned} \tag{1.22}$$

Noticeably, we mention earlier that joint MTR & MTS is a testable hypothesis at the mean level[17]. Giustinelli (2011)) concludes that the idea of testable hypothesis can also be applied

[16] Bounds of the identification region for $P[y(t) \leq \tilde{y}_\alpha]$ under α-QMTS and MTR are mixtures of empirical conditional distributions that cannot be inverted analytically, but can be inverted **numerically**.

[17] See Manski and Pepper (2000): they use education return as an example to illustrate the specification test on the validity of imposing MTR and MTS jointly.

to quantile regression [18].

$$t' \leq t'' \rightarrow Q_\alpha[y|z = t'] = Q_\alpha[y(t')|z = t'] \leq \\ Q_\alpha[y(t'')|z = t'] \leq Q_\alpha[y(t'')|z = t''] = Q_\alpha[y|z = t''] \tag{1.23}$$

The equation (1.23) is a necessary condition for the joint assumption to hold. In the next section, we will show that the joint assumption is not rejected at the 95% confidence interval for the entire ln(hourly wage) distribution. Once MTR+QMTS is not rejected from equation (1.23), we can add QMIV into the model to further shrink the identification region[19].

1.4.6 Distribution Analysis: Findings on the full sample

Our findings regarding bounds on α-quantile under EE, MTR, QMTS, QMIV and their combinations are consistent with the bounds obtained from the mean analysis. However, they are given by conditional quantile responses of the type $Q_\tau[y|z]$,with $\tau \in$ (0,1). In this section, we discuss the results for bounds on the α-quantile of the distribution of ln (hourlywage) under different assumptions and their combinations. Table 1.9 displays the conditional quantiles of the ln (wage) by treatment groups for $\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, the mean, and the distribution of English proficiency. We observe that the monotonicity condition for α-QMTS & MTR to hold is satisfied on the whole distribution in both years because the conditional quantile of ln(hourlywage) is increasing along with the

[18] The first inequality derives from the QMTS assumption and the second inequality derives from the MTR assumption

[19] We will report the results under MTR+QMTS+MTIV in the later sections, but we don't include the equations here, please see Giustinelli (2011) for the details.

distribution across all levels[20].

The estimated bounds on the α-quantile of the distribution of ln(hourly wage), $Q_\tau[\ln(\text{wage(t)})]$, and α-QTEs in 2000 and 2010, are reported in tables 1.10-1.27 for $\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, .0.8, 0.9\}$ (tables 1.10-1.18 show QTEs for 2000, and tables 1.19-1.27 show QTEs for 2010). We show the results under ETS assumption (in conditional quantile panels), worst case assumption (in EE panels), MTRQMTS and MTRQMTSMIV[21]. Note, that the 95% percentile bootstrap confidence intervals are also shown in the tables.

In tables 1.10-1.27, we look at the upper bounds of the global returns (t4-t1) along with the distribution under joint MTR, QMTS, and QMIV in particular, we find that the returns are higher in the upper tails than in the middle and lower tails, which implies that in general, Hispanic immigrants in the higher tails of the wage distribution are more likely to get a high return by acquiring more English skills. MTR and $\alpha-$QMTS assumption shrink the identification regions in the complementary fashion, which is parallel to the mean analysis, combining them substantially increase the identification power. Imposing them together would bring the very informative upper and lower bounds[22]. Comparing the returns in 2000 and 2010, we detect that for the lower percentiles, the returns are larger in 2000 than in 2010 (27.4% at the 10th percentile and 27.5% at the 20th percentile in 2000, but only 19.4% and 25.6% correspondingly in 2010), while the returns in upper percentiles are much larger in 2010 than in 2000. In general, the effects increase substantially over time, inferring that

[20] Note: In fact, the monotonicity assumptions do not have to hold for the entire distribution. This fact is different from the point-identification literature (Chernozhukov and Hansen (2006)).

[21] We don't include the results under MTR,QMTS,QMIV assumptions separately, each of them are much wider than impose them jointly. They are available upon request.

[22] In our case, imposing them together leads to quantile treatment effects between [0,1]

English ability is getting more critical for Hispanic immigrants to gain higher wage incomes and better standards of living in the U.S. In addition, the lack of English proficiency would increase inequality within the Hispanic immigrants, and the situation is getting worse over time.

1.5 Conclusion

In this paper, we examine the heterogeneous impact of English ability on Hispanic immigrants' earnings across the wage distribution over time. We impose some more "conservative" and weaker assumptions to obtain lower and upper bounds of the effects by following Manski's Partial Identification philosophy. We show that, overall, on average, better English skills lead to higher earnings. This positive effect shows a pattern of decreasing marginal return among foreign-born Hispanic immigrants in the U.S. Our results also show that, while the returns to English skills are positive everywhere, there exists a significant degree of heterogeneity in terms of the returns across the wage distribution. In particular, the returns are more pronounced for individuals in the middle and upper tails of the wage distribution than those in the lower tail. Moreover, The overall return increases over time from 2000 to 2010. It can imply that the increasing importance of English for Hispanic immigrants in the labor market as the labor market structures have changed dramatically over the years.

Chapter 2

Return to English Skills in China: Using a Non-Parametric Bound Approach

2.1 Introduction

Since 1978, when Reform and Opening-up started in China, the government has implemented a series of reforms to open the country to trade and foreign direct investment (FDI). As a result, China's foreign trade soared "from $21 billion in 1978, when China at best was a margin player in global trade, to more than $1.1 trillion in 2004, when China became the world's third largest trading economy", (Branstetter and Lardy (2006)). Meanwhile, China has now become the world's second largest recipient of FDI, with the amount of 137 billion dollars, (Reuters (2020)). Increases in both trade and FDI have significantly contributed to China's continuing deep involvement in the globalizing economy and spectacular economic growth over the past four decades (Berthélemy and Demurger (2000), Tsen (2006)).

Increased economic globalization in China has also led to an increasing demand for foreign language skills. The ability to communicate in English, a modern lingua franca, has been found to reduce trade barriers and promote trade especially between countries that do not share a common language in Ku and Zussman (2010). Also, "More and more multinational companies are mandating English as the common corporate language", (Neeley (2012)).

Acknowledging the importance of English skills, Chinese government has been actively promoting English education across the country. Starting from 1978, English has been one of the three core subjects that are tested in Gaokao (China's college entrance exam, which is almost the only path for Chinese students to receiving college education). The emphasis on English education continued to increase after China joined the World Trade Organization (WTO). In 2001, the country even mandated English education for primary school children starts from grade three, (Edwards (2017)). It is estimated that there are more than 440 million English learners and users in China Native Speaker Norms and China English: From the Perspective of Learners and Teachers in China English has become "virtually a second language for the educated Chinese population", (Overholt (2004)).

As noted in Edwards (2017), some are worried that continuing emphasis on English skills and their increasing ties to educational success may negatively impact the time and resources spent on Mandarin and other languages, as well as other subjects in China, and exacerbate educational disparity between urban and rural and ethnic minority remote areas as even access to Mandarin Chinese teachers is not always available in the latter. Mother tongue and bilingual minority education in China: International Journal of Bilingual Education and Bilingualism: Vol 12, No 5 There concerns call for reforms of the current curriculum to reduce the over-emphasis on English education.

Whether English education is worthwhile depends largely on the market value of English skills. The question is: do English skills pay off in the Chinese labor market? Despite the importance of answering the question, very few studies exist examining the impacts of English skills on earnings in the Chinese context. This paper fills the gap.

There are several reasons why English skills may increase one's earnings. First, English skills, often considered as a form of human capital, is particularly valuable at foreign-owned and multinational companies where English is a common language for communication, as mentioned above. Proficiency in English skills can increase access to the jobs at these firms that are usually more prestigious and better paying. Second, even at workplaces that do not mandate the use fo English, there are still many occasions in which English skills are needed, and these skills can also serve as an important positive signal of one's ability in a noisy labor market. Finally, there are also known beneficial effects on individual productivity of bilingualism. The knowledge of a second language can, for example, potentially enhance one's cognitive skills.

To the best of our knowledge, there are only two papers on the impacts of English skills on earnings in the mainland China. Guo and Sun (2014) and Wang, Smyth and Cheng (2017) both find a positive association between proficiency in English and earnings. The former focuses on the returns among college graduates, using the 2010 Chinese College Student Survey data of 19 colleges, whereas the latter examines the returns for the general population using the China Labor Force Dynamics Survey (CLDS). The latter further finds significant heterogeneity in returns to English skills across gender, regions, and urban vs rural areas.

Estimation of the causal effects of English skills on earnings is, however, by no means a trivial task. A positive association between English skills and earnings may simply capture the positive effects of unobservable determinants of earnings such as other productive skills, quality of education and individual ability; the so-called omitted variable bias. For example, as noted in Dovì (2019), there could be a trade-off between the amount of time allocated to improve language skills and that to other skills such as quantitative skills. As a result, when the latter skills are not measured nor included in the model, the estimates can be biased as well. There could also exist reverse causality; an individual with more resources may be able to afford more and better English education to improve her language skills. Guo and Sun (2014) attempt to mitigate the omitted variable bias by using the ordinary least square (OLS) approach and with possible control variables available in their dataset. Wang, Smyth and Cheng (2017) further address the empirical challenges by using the instrumental variable (IV) approach (lead English proficiency and smoking status at age 18).

These approaches rely on different identification assumptions. The OLS approach assumes conditional independence, in other words, all the unobservable determinants that can bias the analysis are included or captured in the model. The IV approach depends on the strength and validity of the IVs. Both approaches can potentially produce inconsistent estimates of the returns to English skills should the identification assumptions fail to hold. For example, Wang, Smyth and Cheng (2017) notes in their paper that smoking status at age 18 was also previously used as an IV for years of schooling in other contexts; this questions the validity the IV since it may affect the outcome through other channels, not meeting the exclusion restriction requirement. Furthermore, both studies use linear models that impose both linearity and homogeneity in the effects of English skills.

Using three waves of the Chinese General Social Survey (CGSS), we reach several important conclusions. First, we find that even when weak assumptions are imposed, we are able to identify a sizable returns to both English listening and speaking skills. More important, the return is unambiguously larger when one's English skills improve from the lowest to the highest level. Second, our results also indicate that the IV estimates can be substantially biased upward, overestimating the returns to English skills. Finally, we also find significant heterogeneity in returns to English skills. Specifically, we find English skills matter more for females, residents with urban hukou, as well as prime-aged workers. Our results here underscore the importance of English skills in the Chinese labor market, one of the largest labor market in the world. These results inform the current debate whether the Chinese government should reduce the emphasis on English education in their curriculum. Our results also contribute to the growing yet still small literature on the role of English skills in the labor markets of non-English-speaking countries. Our results here corroborate the findings in this strand of the literature that English skills, as a form of human capital, can play an important role even if they are not the official language in the country.

Our discussions above are not meant to critique the current literature. Rather, we intend to point out the difficulties facing all the empirical analyses and call for alternative approaches. In this paper, we adopt a conservative approach, called nonparametric bounds analysis, developed in Manski and Pepper (2000). This approach has gained increasing popularity in empirical research for two reasons. First, it relies on transparent and rather weak assumptions, which are in part testable, to identify the causal effects. Second, it does not impose any parametric functional forms in estimation, thereby allowing us to uncover potentially heterogenous English skills with respect to varying proficiency levels. Our results can therefore provide more convincing estimates and potentially help assess the robustness of the existing results. Further, this approach would be particularly important when marginal returns to English skills may only be incremental and small, but the impacts of a larger improvement in English skills may lead to a greater reward.

2.2 Background and Literature Review

In the last quarter-century, In the past few decades, English education has become more and more important in China, which is reflected in both the national and individual levels (Cortazzi and Jin (1996); Adamson (2001); Hu (2005); Jin and Cortazzi (2002)). On the national level, Ross (1992) and Adamson and Morris (1997) summarize that the foreign language education is the key for facilitating economic development and national modernization for the past few decades.

Ms Xiao Yan, the public relations manager of the Wall Street English language school chain, who gave her explanation for the current popularity of English in the following terms:

"[More] and more importance has been given to English after China carried out the policy of reform and opening up to the outside world in the late 1970s. And accompanying China's rise on the world stage in recent years is growing connections of commerce and culture with other countries, especially those developed English speaking countries [. . .] The entire Chinese society attaches high importance to the English study as sometimes it even plays a vital role for a person who plans to pursue further education and seek a better career. There is no doubt that people who have a good command of English are more competitive than their peers." (Daily (2010))

In terms of the generally accepted status of English as a lingua franca, the ***Harvard Business Review*** citing English as the fastest-spreading tongue, the influence of this communication form shows no sign of slowing down. It is unquestioned that English serves as an international language and that its role as such is growing in the world, so it is in China. There are approximately 300–350 million people who have studied English in China. That is more than the entire population of the United States.

Most studies have focused on economic returns to speaking a foreign language in developed countries, with a specific target at the language acquisition and assimilation among the inflowing immigrants. Since the 1980s, the literature started to grow and develop due to the massive age of migration. McManus, Gould and Welch (1983), and Grenier (1984) find the correlation between language skills and wages return for the Hispanic men in the United States. Later on, Tainer (1988) extend the literature by including all foreign-born men into consideration. To deal with the endogenous issue, Bleakley and Chin (2004) find a novel instrument variable, which is immigrants' ages at arrival in the host country intersect with their origin countries for English proficiency. The IV approach is the mainstream for the related literature since then. Usually, the effects estimated from IV estimation is much higher than that from OLS ones.

Putting English skills into the foreign country context, as English has become an international lingua franca, it would be essential to know whether similar patterns are found for a non-English speaking country. Chiswick (1998) is one of the first papers to examine returns to English in a non-English speaking Country. He examines the determinants of, and the returns to, Hebrew fluency for male immigrants. He find that Hebrew speakers who speak English as a second language have the highest earnings of all people in his analysis. The "wage premiums" on English skills are also found in European countries (Ginsburgh and Prieto-Rodriguez (2011)), in India (Azam, Chin and Prakash (2013)), in South Africa (Casale and Posel (2011)) among others. Overall, the average returns are in the range of 10% to 50% among these studies.

2.3 Empirical Specification

2.3.1 Standard Approach

The standard approach is followed by the idea of Mincer (1974) earnings function,modified by the study of workers' earning and augmented with the variable of the measure of language proficiency. The regression model looks as follows

$$lnwage_i = \alpha + \beta \mathbf{language}_i + \gamma X_i + \epsilon \tag{2.1}$$

where the language is the main interest independent variable measures of the English proficiency for an individual, **X** includes all the other control variables. The main estimation regresses the logarithm of wage income on *language* and a set of control variables. One of the most underline assumptions for OLS estimator to be consistent is the ignorable selection (also known as exogenous treatment selection(ETS) i.e., $D \perp\!\!\!\perp (Y_0, Y_1)$), which implicitly implies that the self-selection to language proficiency is random. For simplicity, let us assume *language* is a binary variable, with zero means the individual has bad skills, and one means one has good skills. Therefore, $\mathbb{E}[y_d \mid language = 0] = \mathbb{E}[y_d \mid language = 1]$.

One Dimension of English Proficiency

Exploring the effect of language on wages is similar to other treatment evaluation literature, such as economic returns to education. Equation (2.1) imposes a linear (and homogeneous) relationship between English proficiency and outcome variable of interests, such that each additional level increase of English proficiency has the same marginal impact on the outcome. Let T denote the treatment space with J + 1 ordered values. Provided that there is no correlation between English proficiency and the error term ϵ, correctly specified ordinary linear regression produces an unbiased estimate of β. β is the same on the margins. The ATE of treatment t_3 (highest level) compared to treatment t_1 (lowest level) can be directly obtained from the estimate of β, as follows,

$$ATE(t_3, t_1) = E[y(t_3)] - E[y(t_1)] = \beta \cdot (t_3 - t_1) \tag{2.2}$$

OLS produces a biased and inconsistent estimator of the language effect β if there is a nonzero correlation between indicator *language* and the error term. Such biased may caused by omitted-variable bias, measurement error bias, or reverse effect (Blundell et al.,Richard, Dearden and Sianesi (2005)). The parametric estimation results for the one dimension case can be found in the even columns in tables 2.2 and 2.3. In the odd columns, each level of English Proficiency is treated as a dummy variable, so that one can figure out the marginal return for each treatment group. In the even columns, EP is implicitly treated as a continuous variable.

If individuals sort based on their unobservable preferences, such that the selection is nonignorable (Little,1995), the OLS is an inconsistent estimator for the wage premium of having better language skills. First, language ability may represent many personal characteristics that generate economic returns. For instance, people from richer families or higher wages may have more incentive to invest money and time on pursing a foreign language. Other than this, innate ability, attitudes, and other immeasurable personal identities are correlated with language ability. Solving this drawback has become the central task in the literature. Standard approaches involve an IV estimator, based on an instrument variable correlated with wage only through the channel of *language* and is not correlated with any other variables at the same time. For the literature studying the return to language skill in English speaking countries among Immigrants, origins interact with ages at arrival is widely been used as an instrument variable for English skill (Bleakley and Chin (2004)). There are some other IVs been used in the literature, such as father's education (Dustmann and Soest (2001)), number of children in the household (Chiswick and Miller (1995)), numbers of university in one's living province (Lingguang and Yuanyuan (2017)) and so forth. However, the IV estimator only identifies the local average treatment effect even if the IV is valid. It can not be extrapolated to achieve the population average treatment effect. Also, the exclusion restriction assumption in the IV framework is not easily detectable, and it could fail apart from the mean independent of the potential outcome. Also, by using both OLS and IV regression model, we are only able to quantify the homogeneous marginal effect[1], which is β in equation (2.1). In practice, the constant marginal return is questionable.

[1] *language* is implicitly treated as a continuous variable

Multilevel modeling of English Proficiency

The one-dimension OLS model captures a linear relationship between EP and wage earnings. This assumption may not be very realistic in the real world scenario. In the social capital literature, the size of the effect of a marginal year of schooling seems to have different marginal return to outcomes (Huang, Van den Brink and Groot (2009)). Language skills, as another form of human capital, should be expected to have similar features to education. Thus,the multilevel model seems to be more realistic in the context than the one-dimensional model.

In a homogeneous setting, the outcome is specified as a function of a set of mutually exclusive and exhaustive *language* variable $language_{ij}$ and some observed covariates X, as follows,

$$lnwage_i = \sum_{j=1}^{J} \beta_j language_{ij} + \theta X_i + \epsilon_i \tag{2.3}$$

Provided that if there is no correlation between *language* and ϵ, will lead to an unbiased estimate of β_j. However, when $language_j$ is endogenous in the model, the OLS is an inconsistent estimator for the wage premium of having better language skills. The model (2.3) contains a set of J exclusive binary language level indicators; one may have to handle J binary endogenous variables. Conventional IV approach may not handle the issue thoroughly, due to (1) Difficult to find J instruments, (2) Difficult to interpret the findings from the IV estimation with multiple endogenous regressors. (3) Conventional IV estimation relies on some strong function form assumptions, such as homogeneous treatment effect, conditional independent assumption (CIA), which may difficult to make a consensus on using it.

2.3.2 A different approach:Bounds

Our study employs a nonparametric technique to obtain the bounds on the causal effect of multilevel of English proficiency on wage earnings. Non-parametric bounding methods were first introduced by Manski (1990) and further developed in Manski (1997), Manski and Pepper (2000), Manski and Pepper (2009).

This method is widely used in many different topics later on. Some examples of using this method include Blundell et al. (2007),Gerfin and Schellhorn (2006), Hotz, Mullin and Sanders (1997), Kreider et al. (2012), Gonzalez (2005), De Haan (2011), Mariotti and Meinecke (2015), and many other works.

In this section, we formally define the empirical question and the selection problem and introduce the monotone assumptions. For each individual, we have a response function $y_i(\cdot)$: T $\rightarrow$ Y, which maps treatment t $\in$ T into outcomes $y_i(t) \in$ Y, where the treatment t is the level of individual's English proficiency and y is the person's wages. For each individual in our sample, we observe the realized level of English proficiency and his realized wage $y_i \equiv y_i(Z_i)$, the empirical problem is that we would never observe the outcome $y_i(t)$ for $t \neq z_i$.Empirically, we are interested in the average treatment effect of increasing English skills from s to t on the wage return, and it can be expressed as follows,

$$\triangle\,(s,t) = \mathbb{E}[y(t)] - \mathbb{E}[y(s)] \tag{2.4}$$

The average treatment effect is the difference between two parts: the mean wage we would observe if everyone had English skill level t ($E[y(t)]$) and the wage we would observe if everyone had English level s ($E[y(s)]$). By using the law of total expectations, we can decompose these $E[y(t)]$ and $E[y(s)]$. Let take $E[y(t)]$ as an example,

$$\mathbb{E}[y(t)] = \mathbb{E}[y(t)|z = t]\mathbb{P}(z = t) + \mathbb{E}[y(t)|z \neq t]\mathbb{P}(z \neq t) \tag{2.5}$$

With a data set in which we observe the levels of English proficiency and wages, we can observe the average wage for the group with EP level t, we can know the portion of both receiving t ($P(z = t)$) and also not receiving t ($P(z \neq t)$) in the sample. However, for those who have English skills that are different from t, we would never observe their average wage if they had EP level t. That is, we could not obtain the value of $\mathbb{E}[y(t)|z \neq t]$ no matter any type of data.

2.3.3 Identification Power on the Mean

As discussed earlier, to identify the treatment effect, we need to impose assumptions on the latent part in equation (2.5). The mean-independent assumption[2] is the one being widely used in the literature, and it is also the assumption of OLS regression. This assumption is expressed as: $E[y(t|z)] = E[y(t)]$

We start by estimating the worst-case bounds, and then we will introduce other assumptions to increase the identifying power further afterward.

Worst-Case

Manski (1989) shows that it is possible to identify bounds on E[y(t)] without adding any assumption if the support of the outcome variable is bounded. We can obtain the worst-case

[2] It is also commonly known as exogenous treatment selection (ETS) which assumes that treatments are assigned to the population at random.

bounds for $\mathbb{E}[y(t)]$ by substituting the latent part $E[y(t)|z \neq t]$ with arbitrary $\mathbb{K}_0$ as the lower bound, and arbitrary $\mathbb{K}_1$ as the upper bound. Then equation (2.5) would be bounded as follows,

$$E[y(t_i)|z = t_i]P(z = t_i) + K_0 P(z \neq t) \leq E[y(t_i)] \leq E[y(t_i)|z = t_i]P(z = t_i) + K_1 P(z \neq t) \tag{2.6}$$

These no-assumption bounds are not very informative because they are generally very wide. However, these no-assumption bounds are interesting because all results that are based on different assumptions on the counterfactual part $E[y(t)|z \neq t]$ should lie within this region. In the following subsections, we will introduce monotone treatment response (MTR) assumption and monotone treatment selection (MTS) to our analysis. While in the IV framework, we will introduce how monotone instrument variable assumption and its combination with MTS, MTR make these bounds more informative. The MTR and MTS assumptions are introduced and derived by Manski (1997) and Manski and Pepper (2009). All these assumptions are used to find the plausible regions of $\mathbb{E}[y(t)]$ in equation (2.5).

Assumption on Response Function MTR

The MTR assumption states that the outcome is a weakly increasing function of the treatment:

$$t_j \geq t_i \rightarrow y(t_j) \geq y(t_i) \tag{2.7}$$

equation (2.7) states that increasing one's English proficiency weakly increases his or her wage function, implying that improving one's language ability would never hurt his or her job performance, ceteris paribus. According to the human capital accumulation theory, if we think language skills are one type of human capital, people with better language skills are more productive, at least will not hurt his/her productivity. There are many reasons to expect a positive impact of increasing EP on workers' wages. For example, people who are more willing to invest in a language other than their mother tongue may be more motivated in jobs and who may have some certain expectations when searching for jobs. Note, it could, however, be the case that increasing EP does not necessarily improve wages. MTR takes the possibility into consideration. Therefore, it is important to point out that the MTR assumption does not rule out the possibility of EP being not productive at all and that the positive associations have been found in the literature are due to selection, and therefore, not causation.

The most common case for OLS is based on the "linear response assumption", which means that each individual's hypothetical response on the treatment is assumed to be the same, captured by the constant estimated value $\hat{\beta}$. The linear model implicitly treats the ordinal discrete variable as a continuous variable to estimate the marginal effect. However, this "linear response assumption" is too strong and not always reflect real-world scenarios. After we combine the MTR assumption along with the worst-case bounds discussed earlier, equation (2.6) becomes as follows,

$$\mathbb{E}[y(t_i)|z \leq t_i]P(z \leq t_i) + \mathbb{K}_0 P(z > t_i) \leq \mathbb{E}[y(t_i)] \leq \mathbb{E}[y(t_i)|z \geq t_i]P(z \geq t_i) + \mathbb{K}_1 P(z \leq t) \tag{2.8}$$

To show how equation (2.8) is obtained from equation (2.6) after imposing MTR, we can divide the sample into three groups for illustration: (1) workers with EP less than t, (2) workers with EP equals to t, (3) and workers with EP greater than t. For the second group, we observe the effect on mean wages of having EP with level t. For the first group, we know that their observed mean wage is less than or equal to their mean wage if they had EP level t. So we can use the mean wage we observe to replace $\mathbb{K}_0$ for the first group in equation (2.6). In addition, on the other hand we can use the observed mean wage of the third group to replace $\mathbb{K}_1$ to tighten the upper bound of no-assumption bounds. The bounds estimated from equation (2.8) are shown in tables 5 and 6. The 95% confidence intervals are drawn by the bootstrap method. Compared with the no-assumption bounds, we can see that the bounds are much narrower by imposing the MTR assumption along with the worst-case assumption. According to tables 5 and 6, the log hourly wage for people who do not have listening (speaking) ability at all is bounded between 6.776 and 7.777 (6.855 and 7.777), and the log monthly wage for people who have the best English ability is bounded between 7.777 and 9.589 (7.777 and 9.671). The lower bound of the average treatment effect is restricted at zero, which happens when English ability does not affect workers' wages, given the Monotone Treatment Response (MTR) assumption.

Assumption on Treatment Selections

The second assumption that we impose to the analysis is the Monotone Treatment Selection (MTS) assumption. Under this assumption, workers with better EP have weakly higher mean wage functions than those with lower EP. This can be mathematically shown as follows,

$$t_j > t_i \Longrightarrow E[y(t_u)|z = t_j] \geq E[y(t_u)|z = t_i] \tag{2.9}$$

This assumption relates to the sorting process, which gives us a better scenario of how the treatment groups are distributed to the different individuals in the population, commonly known as "ability bias". As discussed earlier, language is one type of human capital. Workers who have better EP could be different from workers with lower EP, better EP workers should also have higher ability and, therefore higher wages, on average. For some particular occupations, workers who master in English are more able to be productive and efficient. Both arguments are consistent with the MTS assumption in our context. However, they do not work appropriately with the Exogenous Treatment Selection (ETS) assumption, which is commonly assumed under an OLS framework. ETS assumes that treatments are assigned to the population at random, but it does not seem to be the case. The ETS estimates are equivalent to the coefficients obtained by running OLS on log earnings with one dummy variable for each treatment group without adding any covariates. The expected values for each treatment group under ETS are shown in tables 5 and 6[3].

In order to illustrate the usage of MTS assumption, we again divide the sample into three parts - workers who have EP (1) lower than t($z < t$), (2) equals to t(z=t), and (3) higher than t($z > t$). Since the EP of the first group is below t, we can infer that the mean wage level in the first group would be weakly lower than the mean wage of workers with EP equal to t, from the assumption of MTS. Thus, we are able to use the mean wage level

[3] ETS results are the same as multilevel case without any other controls in the linear parametric model.

that we observe from the individuals who have EP at level t as an upper bound for the first group. Applying the same logic, we use the mean wage from individuals with EP at level t as the lower bound for the third group. Hence, we construct the new bound under the MTS assumption as follows,

$$E[y(t)|z=t]P(z \geq t)+K_0P(z<t) \leq E[y(t)] \leq E[y(t)|z=t]P(z \leq t)+K_1P(z>t) \quad (2.10)$$

MTR + MTS

Since MTR and MTS tighten the no-assumption bounds from different perspectives and angles, if we combine MTR and MTS assumptions, we basically assume that the response function is monotone and that selection into treatment is positive simultaneously. Both assumptions are consistent with human capital theories. Thus the combined assumptions should have more identification power on narrowing the bounds. By combining the two assumptions, we obtain the MTR-MTS bounds[4].

$$\begin{gathered} E[y(t)|z<t]P(z \leq t)+E[y|z=t]P(z>t) \\ \leq E[y(t)] \leq \\ E[y(t)|z=t]P(z \leq t)+E[y|z>t]P(z>t) \end{gathered} \quad (2.11)$$

As we can see from equation (2.11), once we impose the assumptions together, $\mathbb{K}_0$ and $\mathbb{K}_1$ are eliminated. Also, the joint MTR-MTS assumption is testable. Under this assumption,

[4] For a full derivation of the MTR-MTS bounds, see Manski (1997) and Manski and Pepper (2000)

the following should hold[5], for $t_j > t_i$,

$$\begin{aligned} E[y|z = t_j] = E[y[t_j]|z = t_j] &\geq E[y[t_j]|z = t_i] \\ &\geq E[y[t_i]|z = t_i] = E[y|z = t_i] \end{aligned} \tag{2.12}$$

Under the joint assumption, the mean wage of people should increase with the hypothetical increase in EP. If this is not the case, the joint assumption of MTR and MTS should be rejected, implying any assumptions fail in the context.

Monotone Instrumental Variable Assumption

Due to "ability bias", exogenous treatment assumption is failed. Endogenous variables are often encountered in empirical work. Hence, the method of instrumental variables is widely used in the evaluation of the treatment effect. In the setting of the traditional Instrument Variable (IV) approach, suppose now we observe not only the workers' EP and their wages but also a variable v, where v is related with EP, and only affect wages through the channel of EP. Though standard IV assumptions can greatly aid in identification, the credibility of it is often a matter of no consensus. Under the condition of mean independence form of the standard IV condition, the following should hold for all treatments $t \in T$ and all values of the instrument $m \in M$,

$$\mathbb{E}[y(t)|v = m] = \mathbb{E}[y(t)] \tag{2.13}$$

Equation (2.13) means that the wage of the workers should be mean-independent of the

[5] The first inequality derives from the MTS assumption and the second inequality derives from the MTR assumption

instrument variable v. If v satisfies the IV assumption, we can obtain an IV lower bound on $\mathbb{E}[y(t)]$ by taking the maximum lower bound and an IV upper bound by taking the minimum upper bound, which can be shown as follows,

$$max(LB_{E[y(t)|v=m]}) \leq E[y(t)] \leq min(UB_{E[y(t)|v=m]}) \tag{2.14}$$

In reality, it is usually complicated to find a valid IV, which satisfies the mean independent assumption. Neither there is a consensus way to prove the validity of IV. Here we use a weaker version on IV assumption: the monotone instrumental variable assumption (MIV). A monotone instrumental variable satisfies the following condition,

$$\begin{aligned} m_1 < m_2 < m_3 \rightarrow \mathbb{E}[y(t)|v = m_1] &\leq \mathbb{E}[y(t)|v = m_2] \\ &\leq \mathbb{E}[y(t)|v = m_3] \end{aligned} \tag{2.15}$$

Instead of assuming mean independence, a variable v is a MIV in the sense of mean monotonicity, which allows for a weakly monotone relation between the variable v and the mean wages function of the workers. To use the MIV, we can still divide the sample into subsample based on v and obtain bounds of each sub-group, and then weighted the bounds of each of them to find the population's bounds. Under three assumption jointly imposed, the bound for $\mathbb{E}[y(t)]$ is as follows,

$$\begin{gathered} \sum_{v \in m} \mathbb{P}(v = m)\{sup_{m_1 \leq m}(\sum_{k<t} \mathbb{E}[y|v = m_1, z = k]\mathbb{P}(z = k|v = m_1) + \mathbb{E}[y|v = m_1, z = t]\mathbb{P}(z \geq t|v = m_1))\} \leq \\ E[y(t)] \\ \leq \sum_{v \in m} \mathbb{P}(v = m)\{inf_{m_2 \geq m}(\sum_{k>t} \mathbb{E}[y|v = m_2, z = k]\mathbb{P}(z = k|v = m_2) + \mathbb{E}[y|v = m_2, z = t]\mathbb{P}(z \leq t|v = m_2))\} \end{gathered} \tag{2.16}$$

2.4 Data and Descriptive Statistics

Our sample was constructed by pooling together the 2012, 2013, 2015 waves of the Chinese General Social Survey (CGSS). We choose to use these three years of data for this study because the questions related to English proficiency are only asked in these years. At the same time, a larger sample size is obtained by pooling the cross-sectional data. The CGSS is an annual cross-sectional survey of urban and rural Chinese households. All CGSS datasets are stratified with more than 10,000 households per survey that are selected at random from more than 100 districts and counties in mainland China. An eligible household member is randomly selected to be the survey respondent from each sampled household. The CGSS system collects data from multiple elements of the society, including community, family, and individuals, which provide essential information on demographic, economic, and other personal and family characteristics. In the survey, there are two questions firmly close to language ability: (1) How do you evaluate your English listening ability? (2) How do you evaluate your English Speaking ability? Thus, one is able to investigate the effect of speaking English and listening English separately. Obviously, both values for these two questions are self-reported in our data. Though the subjective report on the levels of English proficiency has been widely used in the language-related literature, and we admit the deficiency of it. Given that the self-reported English proficiency variable is commonly used in the literature, we adopt the same measure in our paper. While self-reported answers may not be a perfect measure of language skills, they are highly correlated with scores from tests designed to measure English language skills and functional measures of English language skills (Akbulut-Yuksel, Bleakley and Chin (2011)). The English proficiency has three levels : 0 means one has no English skills at all; 1 means one has little English skills, and 2 means one has good English skills. Our dependent variable is the logarithm of monthly wages. Due to the lack of information on monthly working hours, we are not able to impute the accurate hourly wages[6].

In order to efficiently estimate the wage return to English skills in the Chinese labor market, we impose some restrictions on the original pooled data. As Labor Law explicitly states the work-age to be 16 or over, we excluded the age group under 16. Male individuals aged above 60 and female individuals above 55 are also excluded because most of them have retired due to the Chinese pension system, and fewer of those above 60 have ever had a chance to learn English. To consider labor supply as a factor affecting the effect, we only keep individuals firmly attached to the labor market, individuals whose working hours per week less than 35 hours are dropped from the sample. Also, to ensure that our sample includes only those workers with a stronger attachment to the labor market, we exclude those individuals with less than 600 RMB (equivalent to 85 U.S. dollars) per month. In our specification, y is the natural log of monthly income wages that may vary from negative infinity to positive infinity in principle. According to the data, the observed y is in the range from 6.38 to 11.06. Thus, we arbitrarily set lower bound ($\mathbb{K}_0$) of y as 6 and 10 as its upper bound ($\mathbb{K}_1$)[7]. Benefit from the demographic information covered in the CGSS data, we also study on the heterogeneous effect into different sub-group, the characteristics we study are such as, education attainment,age, occupation and location.

Table 2.1 provides the summary statistics of English proficiency and monthly wage under

[6] In China, wages are paid to the workers on a monthly basis

[7] The arbitrary setting won't affect our final estimations.

different categories and demographic groups. The average English Proficiency among the full sample is 0.725 on a scale of 0-2, shows that the overall English proficiency in China is still at a very low level. The self-reported listening proficiency seems always to be higher than speaking proficiency in terms of either the whole samples or different subgroups, which shows the pattern that non-native speaker mostly use English for job tasks rather than oral communication. From 2012 to 2015, the average English proficiency in China is in a increasing trend. Females tends to have better English skills than males on average, in terms of both speaking and listening ability. One's English Proficiency is highly correlated with her education attainments. Those who have at least some college education or above have reported much better proficiency. English as a subject is one of the most important curriculum in the Chinese educational system. For some Universities, for those high quality and well-known ones in particular, college students need to pass CET-4 and CET-6 tests to get the bachelor diploma. Guo and Sun (2014) examined returns to English proficiency, measured by CET-4, for college graduate in China. In the bottom panel of table 1, one can observe is that English proficiency is negatively correlated with ages, younger cohorts have significant better English skills than older cohorts. In China, "since 2001, it has been the official policy that English should be learnt from the age of 8 or 9 onwards, in Grade 3 of the national education system (Bolton and Graddol (2012) page2)". Hence, the younger generations experience early exposure to learning English. The average speaking proficiency is 1.038, and average listening proficiency is 1.134 for people under age 30.

2.5 Estimation

2.5.1 Parametric Estimation

Before we turn to our main estimation results on the ATE, we present OLS estimates of the return to English skills, for both one dimension and multilevel dimension cases. Tables 2.2 and 2.3 present preliminary findings from the OLS regression. For each odd column, *language* is treated as a multilevel variable, and it is treated as a one-dimension variable for each even column. One can observe that we can find different marginal effects between each level in the odd column. While the marginal return is constant in each even column.

The OLS estimations indicate a robust positive effect of *language* on wages no matter in which dimensions and if other covariates are controlled. The homogeneous effect can be up to 29.7% with no other variables are controlled in the model. While the global effect can be up to 59.1% without controlling for other covariates in the model. When we control for different sets of control variables, the coefficients vary with the model specifications.

The findings from the multilevel model suggest a strictly monotone relationship between English Proficiency and wage earnings, and estimates from the multilevel model are moderately derived from the estimates produced by the one-dimension model. The marginal effects from the multilevel model are generally larger than the constant marginal effect produced by the one dimension model. As we add more control variables into the model, there would be a higher chance that the *error term* is not correlated with the main interest variable EP. The magnitudes of the effect decrease from 28.4% to 6.75% after controlling for personal and location characteristics variables, which implies that omitted variables upward bias the estimation.

Results estimated from the 2SLS model are reported in tables 2.4 and 2.5. Following the instrument variables used in the literature (Shields and Price (2002), Lingguang and Yuanyuan (2017)), we use both the numbers of kid in the family and the numbers of university in each province as IVs for language ability. The partial F statistics of the excluded instrument are above 10 in each model specification. Thus the proposed IV sets do not suffer from the weak instrument problem.In the first stage results, the numbers of university is positively associated with the English proficiency, which infers that the surrounding environments have impact on exposure to the language study. Provinces with more universities must have better vibes and more opportunities that people can be engaged to learn foreign language. On the other hand, the number of kids is negatively associated with adults' language proficiency. This may conflict with the hypothesis proposed in the developed countries, which says that younger children can be their parents translators and help them improve the pronunciations. However, the richer family in China is more willing to have less children than the poorer ones, which can be reflected from the comparison between urban and rural areas. According to the second stage results, one unit increase in English listening ability and speaking ability could increase her monthly wage by 67.3% and 70.1%, respectively. In terms of the magnitudes of effect, IV estimates are much higher than the OLS estimates, which are consistent with the sake of measurement error being more important than omitted variable bias and are consistent with the previous results, such as the arguments in Bleakley and Chin (2004), Chiswick and Miller (1995), Saiz and Zoido (2005). Chiswick and Miller (2010) finds that the ratio of IV to OLS is something in between 2.4 and 9.1. Economic returns to fluency in English among non-immigrant populations in the range 40-50% or higher have been found in the Baltic states (Toomet (2011)), Europe (Ginsburgh and Prieto-Rodriguez

(2011)) and South Africa (Casale and Posel (2011)).

2.5.2 Non-parametric Bounds Estimation

Feasibility of nonparametric assumptions

The OLS model above may produce a biased and inconsistent estimate of the EP effect on wages when some unobservable heterogeneity simultaneously influences the choice of investing in EP and wages specific to the individuals.

Nonparametric bounds analysis requires a couple of relatively weak nonparametric assumptions to identify the upper and lower bounds of the causal effects of English Proficiency. In this study, the monotone treatment assumption(MTR) implies that improving someone's EP cannot lower her wage, which means that each person's wage function is weakly increasing in conjectured EP. The human capital accumulation theory can explain this assumption that language skills can boost one's productivity in the labor market; at least will not hurt one's productivity. The monotone treatment selection(MTS) assumption implies that workers with higher EP have a weakly higher mean's wage function than workers with lower EP.(Positive selection on job types). To some extent, the MTS assumption is a natural way to deal with the selection issue, before putting IV into consideration.

Manski and Pepper (2009) suggest that MTR-MTS is a testable hypothesis jointly. Under this hypothesis, the average outcome for the realized treatment must be a weakly increasing function of the realized treatment. Otherwise, we should reject this hypothesis and can not impose them jointly. Figures 2.1 and 2.2 demonstrate that monthly wage increases with the level of treatment in a solid way for both listening and speaking ability. Thus, it makes sense to impose MTR-MTS assumptions jointly in our context in terms of the whole sample.

Considering IV, along with MTR-MTS assumptions, can further tighten the bounds. The credibility of mean independence condition has often been subject to disagreement in empirical research. Most of the time, there is no consensus conclusion on the validation of the instrument variables. Instead of mean-independence conditions, the MIV assumption explicitly implies a weakly monotone relation between the instrumental variable and the mean outcome function. In this paper, we adopt the numbers of universities in each province as the MIV. Because the numbers of the university should not directly affect individuals' wages, while persons who were born in the provinces having more universities may have more opportunities to learn English and start to learn English at younger ages. From the view of linguistic literature, young children tend to learn languages much faster and easier than adults and older children, because learning a language is a part of their brain chemistry. Young children absorb new information typically in an unconscious state of mind. On the other hand, adults learn new information in a more conscious way that is more likely to get lost. Thus, we should expect that one's English proficiency is supposed to be positively correlated with the general educational environment. Luckily, the first stage results in table 4 further complement the conjecture[8].

Average Treatment Effect

Our goal is to find the average treatment effect of EP on income wages, which is defined as $\mathbb{E}[y(t_{\mathrm{j}})-y(t_{\mathrm{i}})]$. We are interested in gains in expected wages when moving from one treatment group to the next, which measures the gains from improving English acquisition. Figure 2.3

[8] We do not adopt the numbers of kid here because its effectiveness is skeptical in the Chinese context.

and 2.4 graphically show the bounds of the ATE($\mathbb{E}[y(t_3) - y(t_1)]$) estimated under varying individual and joint assumptions for English listening and English speaking, respectively. The bounds of ATE under each of WC, MTR, MTS assumptions solely are broad, and therefore are less informative. We are not able to make any solid conclusions based on these wider bounds. In order to further tighten the bounds, we impose these assumptions jointly. When impose them together, they can have substantial identifying power. From the figures, the bound is much narrower after combining MTR and MTS assumptions, and become more informative after adding MIV assumption. The estimated bounds of E[y(t)] and ATEs based on these assumptions and joint assumptions are reported in table 2.5, 2.6, 2.7. The worst-case bounds imply that ATE(t_3-t_1) could be anywhere between -2.328 and 2.813 for listening ability, and between -2.305 and 2.816 for speaking ability. When we assume MTR+MTS, the worst-case bounds are narrowed down to [0, 0.564] and [0,0.591], which provide the largest possible ranges of the overall effect of English listening ability on wage incomes is anywhere between 0% and 56.4%. For English speaking ability, the overall return must be anywhere between 0% and 59.1%. When we add the MIV to the MTR+MTS assumptions, both lower bound and upper bound for (t_3-t_1) are shrunk. Now the overall effect for listening ability and speaking ability are [0.056,0.498] and [0.056,0.521],respectively.

Other than the global effect, it is also informative to see the ATEs from one treatment group to the next, because the marginal returns could very likely be heterogeneous. In our case, we have also shown ATE from t_1 to t_2 and t_2 to t_3, which interpreting the average wage returns from no English ability to some English ability, and from having some English ability to well English ability. Based on the results, the maximum wage gains vary considerably along with the treatment, even though the lower bounds are always zero from t_1 to t_2 and

t_2 to t_3, but the upper bounds still provide us the informative findings. For listening ability, the upper bounds corresponding to moving from t_1 to t_2 and t_2 to t_3 are 33% and 34.5%. For speaking ability, the corresponding upper bounds are 32.1% and 37.8%. The margin returns for both treatments(listening and speaking) are increasing along with treatment groups. At the lower level (from t_1 to t_2), the return from gaining better listening ability is a little larger than the return from gaining better speaking ability. In comparison, at a higher level (from t_2-t_3), the marginal return is higher from accumulating more speaking skills.

Introducing Covariates

So far, we only show the population average treatment effect without considering any subgroup analysis. The overall effect for the population is pretty high, and it turns out not to be so informative to understand the effect. Also, in terms of the effect could be heterogeneous among groups with different characteristics and identities. We want to explore the how effect may vary across individuals in the following section.

English ability is a complementary skill to educational attainment. In the previous study, the wage penalty may increase with the level of education(Carliner (1996); Mora and Davila (1998)). Carliner pointed out that education is associated with both lower costs and more enormous benefits to learning English. His argument fits the Chinese context well. The subject of English is an essential part of the Chinese curriculum. In order to get admission to high-ranking schools, students have to achieve a good score on English exam.

Table 2.8 and 2.9 show that the values for the bounds on log wages for each of the treatments under different assumptions, conditioning on two different education levels. (Less than a high school degree and at least a high school and above degree). As we discussed in the earlier part, the joint MTR+MTS assumption is a testable hypothesis, which would be rejected if average outcomes are not weakly increasing along with the treatments. This hypothesis is valid for listening ability among all educational groups. Nevertheless, we would be more likely to reject this joint assumption on returning to the speaking ability for lower-level education group, which including high school dropout workers in China, but not for the other group. Comparing the effects of listening ability and speaking ability in high educated group, the overall returns by having a well-speaking ability is larger than having a well-listening ability. However, based on upper bound estimations, the listening ability can boost more returns than speaking ability from no English skill at all (t_1) to mild English skill (t_2). One possible way to explain this interesting finding is that employees can finish the work if they understand the tasks, which is getting from reading the instruction or listening to their managers. However, in order to finish more difficult tasks that may need certain level communication skills, workers need to have a greater speaking ability. In China, most people especially recent young generations are able to read English and have good ability to understand English via listening, but the real lack is to use English for daily communication. Gonzalez (2005) finds a similar pattern in the U.S. labor market among Hispanic immigrants, where English proficiency could have significant and positive on wages only for people with at least a high school diploma. Mora and Davila (1998) finds that language proficiency is a complement of education, which can be explained in our case as well. If MTR+MTS assumption is rejected, which infers that at least one of these two assumptions are failed to impose in the context. MTR assumption simply implies that if someone accumulates more human capital, at least it will not hurt her job performance, which should not consider as a lousy assumption in our case. However, we may challenge the validity of the MTS assumption, which infers positive ability selection. Thus the results suggest the presence of negative ability bias in the group of low-level educated workers. For this specific group, it is not necessary to conclude that workers who speak English very well have higher mean wage functions than workers who do not know how to speak English well.In terms of both the returns to listening and speaking ability, we can rule out the large effects for workers belong to lower education groups, especially in terms of speaking ability. When we apply MIV along with MTR+MTS for high school and above group, the overall effect of listening ability on wages is [8.9%, 30.7%] and the overall effect of speaking ability is [8.9%, 32.8%].

Typically, the distribution of female and male workers are not random in different occupations. In China, female workers are more likely to work in service types, and males may be more likely to work in technology and manager positions. The returns may vary among the gender groups. Hence we divide our sample by different gender and study the returns to male and female workers separately. The returns can also be different in other dimensions, such as geographical location, hukou types, ages, and occupation types.

Male workers tend to have higher returns than female workers under any assumptions, regardless of listening and speaking ability. In terms of listening, under ETS, the return from no listening to a little listening ability is 28% for females and 30.6% for males, and the overall return is 57.8% for female workers and 63.4% for male workers. When applying non-parametric bound approach, under MTR+MTS, all the estimates under ETS assumption are between LB and UB. However, when MIV is imposed jointly, the overall returns to listening ability is between 10.8% and 44.6% for female workers, while 4.2% and 56.4%

for male workers. The lower bounds are both statistically different from zero, which indeed shows the returns are positive for both male and female workers. Also, we notice that for female workers, the return from having little listening ability to well-listening ability is between 1.5% and 25.7%. However, the lower bound is zero for male workers. One way to illustrate the finding is that female workers are more easily to gain benefits, but male workers are more likely to gain more substantial benefits from having better listening skills.

Age reflects different stages of workers' career. We classify the whole sample into three age groups: 17-29, 30-45, 46-60, which correspond to early career, mid-career, and late-career. By doing so, we can find out if the effects are heterogeneous among different stages. The results for returns to proficiency in English listening and English speaking skills are reported in tables 2.12 and 2.13. Assuming exogenous selection would lead us to conclude that the return is high in a mid and late-career than an early career. When applying non-parametric approach, the lower bound for early career group and late career group can not be excluded from zero, and this is the case for both English listening and English speaking skills. One story behind the results is that most workers have their most promotion opportunities during their mid-career stage. If human capital is one assessment criterion for promotion, then workers with better foreign language skills may have a better chance of getting a higher salary. Here English skill is served as a signal for unobservable ability in the noisy labor market. Azam, Chin and Prakash (2013) find a similar pattern in India, in which middle-aged workers hold the highest return from English skills. Overall, the returns to English listening skill at the mid-career stage are between 7.3% and 53.6%, and the rate of return to English speaking skill is between 7.3% and 58.1%. Still, the return is high from having better speaking ability than listening ability.

2.6 Conclusion

For the past several decades, the Chinese economy's ongoing development has fostered the relevance of the demand for foreign language competences in the labor market. English, which serves as a common tongue among international business, is the most important foreign language that we want to test its value. Most existing studies have focused on returns to English proficiency in developed countries, but few studies have explored for developing countries. To estimate the economics return to English skills in the labor market, regressing one's English ability on his/her wage incomes generally gives positive estimates. Though, the literature applies several strategies to identify the causal effects, most of which are OLS and IV methods; nevertheless, the empirical evidence on the causal relationship between EP and wages is moderately convincing, because these methods identify the effect for a limited population, and the different methods give inconsistent results. This paper fills these two gaps.

In order to find different marginal effects due to English proficiency as a categorical variable, and to deal with the selection problem caused by people's selections into works, we adopt non-parametric bound method to quantify the lower bound and upper bound of the population average treatment effects. We find there is a convinced "English Premium" in the Chinese labor market. On average, the overall monthly wage premiums are between 5.6% and 49.8% in terms of English listening ability, and the overall premiums for having better English speaking ability is higher, which is between 5.6% and 52.1%. The estimates obtained by assuming exogenous sorting (OLS model) are 56.4% and 59.1%, while the estimates obtained from the IV model are 67.3% and 70.1%, which all turn out to overestimate the

English premiums in China.

The returns are heterogeneous in different subgroups. Regarding education attainment, the English premium is much larger for workers having more than a high school diploma compared to those who only have lower educational attainment. English speaking premium does not necessarily exist for these low educated workers.

Several policy implications can be drawn from the results. Policymakers should emphasize the teaching of English at schools in order to improve the overall English proficiency of future generations in China. Oral English ability improvement should be more emphasized in English education. This is because people in China are usually more confident of listening English rather than speaking it out, but on the other hand, having the ability to communicate in English is able to boost productivity and gain efficiency in the workplace, particularly for some specific tasks which need intensive foreign language skills. Our results also contribute to the growing yet still small literature on the role of English skills in the labor markets of non-English-speaking countries. Our results here corroborate the findings in this strand of the literature that English skills, as a form of human capital, can play an important role even if they are not the official language in the country.

Chapter 3

Economic Value of Language in China: How important is Mandarin Proficiency in China Labor Market - From A Different Approach

3.1 Introduction

To many, Mandarin is often perceived as the only language in China synonymous with "Chinese". This is actually not true. China is a linguistically diverse country with hundreds of local languages or dialects that are not necessarily mutually intelligible. More important, the differences among these local languages are far greater than many would have thought. In fact, as noted in Dovì (2019) and Tang and Van Heuven (2009), "the mutual intelligibility of these dialects can be as low as 2% (measured on a sample of words and sentences), and on average does not surpass 50%."

Not only did such great differences create communication barriers among people from different regions in China, but they were also considered by the Chinese government to have adversely affected economic progress and cultural exchange. As a result, the Chinese government has been promoting the standardization of the spoken Chinese language since 1950s and gradually came to settle upon Mandarin," a common speech with pronunciation based on the Beijing dialect", as the standard spoken language. In 1956, the government issued The State Council Directive on the Promotion of Putonghua, which aimed to popularize Mandarin and promote its use in schools and in many aspects of daily life and government official activities. In 2000, the use of Mandarin even became law (Jiang (2000)), and was further established as the official national language in 2010 by China's National Language Law.

Because Mandarin is essentially developed from a dialect, it is not necessarily well understood to many in China. Despite the great promotion effort, there were still 30% of the population who is unable to communicate in Mandarin (Dovì (2019)). "[F]or most Chinese people, Mandarin is somewhere between being a native language, an intelligible dialect, and an almost 'foreign' language." (Dovì (2019)). In this sense, Mandarin should be considered as more of lingua franca in the Chinese labor market, and its role should also be similar to that of English in the global economy or those countries with English being the dominant or official language such as India. Proficiency in the official or dominant language has been shown to be positively related to one's earnings. For example, Azam, Chin and Prakash (2013) find that in India, where English is the national language but the proficiency levels remain low, there exists a sizable effect of English skills on wages in India, and more important, that the returns to English proficiency are even higher than the returns from education.

A natural question follows: do Mandarin skills pay off in the Chinese labor market, especially for migrants, as do English skills in many countries? This is what we attempt answer in this paper.

According to the literature on the economics of language, there are a few reasons why a higher level of language skills can be associated with higher earnings. First, use of a common language can certainly reduce communication barrier and increase the quality of job matches, as suggested above, and increase one's productivity at workplace (Grogger (2011) and Yao and van Ours (2019)). Reduced communication barrier may also expand one's social network which is an important channel for job search in China. Second, when one's other productive traits are unobservable, language skills also often serve as a signal of one's ability and increase her likelihood of employment (Dovì (2019)). Dovì (2019) shows that one standard-deviation increase in Mandarin proficiency increases employment probabilities by 4% for the whole sample and up to roughly 5% in urban China.

Both reasons above suggest a causal impact of language skills on earnings. However, the positive association between language skills and earnings may also be a spurious one, capturing the effects of omitted variables, or reflect reverse causality. First, better language skills may be related to unobservable variables such as quality of education that are positively related to earnings. Effort and time spent on improving one's language skills can also crowd out resources spent on other productive skills that are not necessarily measured in the data (e.g., quantitative ability in Dovì (2019)). Second, communication may be more prevalent for high-skilled or -paying jobs, and language skills can improve through more practice at such places (Dovì (2019)).

The key challenge facing the studies of economic returns to Mandarin skills is indeed to isolate the causal impacts of language skills. Common approaches to address the challenge in this context include the ordinary least squares (OLS), propensity score (PS) (Saiz, Zoido et al. (2002)), as well as instrumental variable (Wang, Smyth and Cheng (2017), Ginsburgh and Prieto-Rodriguez (2011)) approaches. Both the OLS and PS approaches rely primarily on the conditional independence assumption that requires the exogeneity of language skills once the control variables are included. The IV approach relies on existence of an IV variable that is not only strongly correlated with the language skills, but also uncorrelated with the unobservable determinants of earnings and does not affect earnings either directly or indirectly through channels other than language skills. For example, Dovì (2019) uses language use outside of the workplace and migration variables as IVs.

Independence or exogeneity assumptions, however, are often too strong and cannot be directly tested in practice. Moreover, both the OLS and IV approaches typically impose linearity in the relationship between language skills and earnings, and hence homogeneity in returns to language skills. While these challenges are well known, convincing solutions addressing them together have proven elusive, thereby calling for alternative approaches. In this paper we employ nonparametric bounds analysis that relies on transparent and weaker assumptions and are in part testable. Since our approach is nonparametric, it also allows us to uncover varying returns to language skills with respect to the levels of language skills.

To perform our analysis, we utilize data from the 2016 wave of the national- wide China Labor-force Dynamics Survey (CLDS). This is the most recent available CLDS data, which specifically focuses on the labor market and demographic characteristics of the respondents.

One unique advantage of CLDS data is that each respondent's Mandarin proficiency was objectively assessed by the interviewers, which will reduce the subjective measurement error. We indeed find a Mandarin premium in the Chinese labor market. The positive effects are a result of better job matches and higher levels of productivity. Second, we show that while the returns on wages are positive for both male and female workers, proficiency in Mandarin skills benefit females more than males, which can potentially help narrow the gender gap. Finally, our results show that the IV results are usually high and sometimes are overestimated.

The rest of the paper is organized as follows. Section 2 briefly reviews the literature; Section 3 presents and discusses the empirical methods employed; Section 4 describes the data; Section 5 and 6 discuss the results; Section 7 discusses the potential mechanisms, section 8 contrasts the nonparametric results with traditional parametric results, and Section 9 concludes.

3.2 Literature Review

Most literature that relate to the Economics of language in China has evolved after 2010. Gao and Smyth (2011) found significant and positive returns to Mandarin proficiency in China, but they only targeted at the rural-urban migrants. They found that: (1) the average return is 42% for the whole sample; (2) the return for female migrants is 50.8%; and (3) no significant effect is found for male migrants in China. They also pointed out economics return to English proficiency is the future direction for the topic in China, because the demanding for English speakers has keep increase accompanied by the stunning economic growth. Wang, Smyth and

Cheng (2017) found positive earnings returns to proficiency in English in China, in which the average return to hourly wage is at 16%. Later on, outcomes other than wages have been studied in China as well. Dovì (2019) investigated the association between Mandarin proficiency and employment probabilities, and positive 'employment premium' was found. Wang, Cheng and Smyth (2019) found that be proficiency in Mandarin can improve health. Moreover, they measured that about 2% to 20% health inequality in China can be explained by Mandarin proficiency.

3.3 Empirical Specification

3.3.1 General Approach

The conventional way to explore the effect of language on wages is similar to other treatment evaluation literature, such as economic returns to education. It is followed by the idea of Mincer (1974) earnings function,modified by the study of workers' earning and augmented with the variable of the measure of Mandarin proficiency. The regression model looks as follows

$$lnwage_i = \alpha + \beta \mathbf{MP}_i + \gamma X_i + \epsilon_i \tag{3.1}$$

where MP is the main interest independent variable measures the Mandarin proficiency for the individuals, $\mathbf{X}$ includes all the other control variables. The main estimation regress the logarithm of wage income on MP and a set of control variables[1]. One of the most underline assumption for OLS estimator to be consistent is ignorable selection (also known

[1] Note: Here, β is a scalar, γ is a vector.

as exogenous treatment selection(ETS) i.e.,$D \perp\!\!\!\perp (Y_0, Y_1)$), which implicitly imply that the self-selection to Mandarin proficiency is random. For simplicity, let's assume MP is a binary variable, with zero means individual cannot speak Standard Mandarin and one means can speak Standard Mandarin. Therefore, $\mathbb{E}[y_d \mid MP = 0] = \mathbb{E}[y_d \mid MP = 1]$. If individuals sort on the basis of their unobservable preferences, such that the selection is nonignorable (Little,1995), the OLS is inconsistent estimator for the wage premium of having better language skills. First, language ability may reflected by many personal characteristics that generate economic returns, such as people from richer family or with higher wages may have more incentive to invest money and time on pursing a foreign language. Other than this, innate ability, attitudes and other some immeasurable personal identities are correlated with language ability. Solving this drawback has become the central task in the literature.

As causal effect has been more popular and important in the social science research, instrument variable approach becomes more dominated to deal with endogeneity issue of language variable. The idea is that to find a new variable v, which satisfies two conditions: (1) v is strongly correlate to the language variable; (2) v cannot directly have impact on wages, if and only if through the channel of language. The second condition also infers that v is mean-independent of the outcome variable ($v \perp\!\!\!\perp Y|X$). However, the exclusion restriction assumption is not easily detected. The mean-independent assumption is hard to be convinced. In addition, we are only able to quantify the constant return in sub-population groups from using IV approach[2]. In addition, the constant marginal return is questionable in practice.

[2] *language* is implicitly treated as a continuous variable

3.3.2 A different approach:Bounds

Non-parametric bounding methods were first introduced by Manski (1990) and further developed in the sequential works of Manski (1997), Manski and Pepper (2000), Manski and Pepper (2009). The idea of this method is to calculate lower and upper bounds of the treatment effect based on a fewer weaker assumptions.

This methods are widely used in many different topics later on. Some examples of using this method include Blundell et al. (2007),Gerfin and Schellhorn (2006), Hotz, Mullin and Sanders (1997), Kreider et al. (2012), Gonzalez (2005), De Haan (2011), Mariotti and Meinecke (2015) and many among others.

In this section, we formally define the empirical question and the selection problem and introduce the monotone assumptions. For each individual we have a response function $y_i(\cdot)$: T $\rightarrow$ Y, which maps treatment t $\in$ T into outcomes $y_i(t) \in$ Y, where the treatment t is the level of individual's Mandarin proficiency and y is the wages of that person. For each individual in our sample, we observe the realized level of Mandarin proficiency and his realized wage $y_i \equiv y_i(Z_i)$, the empirical problem is that we would never observe the outcome $y_i(t)$ for $t \neq z_i$.Empirically, we are interested in the average treatment effect of increasing Mandarin skills from s to t on the return to wage, it can be expressed as follows,

$$\triangle\,(s,t) = \mathbb{E}[y(t)] - \mathbb{E}[y(s)] \tag{3.2}$$

The average treatment effect is the difference between two parts: the mean wage we would observe if everyone had Mandarin skill level t ($E[y(t)]$) and the wage we would observe if everyone had Mandarin level s ($E[y(s)]$). By using the law of total expectations, we can decompose these $E[y(t)]$ and $E[y(s)]$, let's take $E[y(t)]$ as an example,

$$\mathbb{E}[y(t)] = \mathbb{E}[y(t)|z = t]\mathbb{P}(z = t) + \mathbb{E}[y(t)|z \neq t]\mathbb{P}(z \neq t) \tag{3.3}$$

With a data set in which we observe the MP level and wage, we can observe the average wage for group with MP level t, we can know the portion of both receiving t ($P(z = t)$) or not receiving t ($P(z \neq t)$) in the sample. However, for those who have Mandarin skills that are different from t, we would never observe what their average wage would have been if they have MP level t. That is , we could not obtain the value of $\mathbb{E}[y(t)|z \neq t]$ no matter any types of data we could get. This is the part we have to impose some assumptions on.

3.3.3 Identification Power on the Mean

As discussed earlier, in order to identify the treatment effect, we need to impose assumptions on the latent part in equation (1). The mean-independent assumption [3]is the one being widely used in the literature and it is also the assumption of OLS regression. This assumption is expressed as: $E[y(t|z)] = E[y(t)]$

We start with estimating the worst-case bounds, and then we will introduce other assumptions to further increase the identifying power afterwards.

Worst-Case

Manski (1989) shows that, it is possible to identify bounds on E[y(t)] without adding any assumption if the support of the outcome variable is bounded. We can obtain the worst case

[3] It is also commonly known as exogenous treatment selection (ETS) which assumes that treatments are assigned to the population at random.

bounds for $\mathbb{E}[y(t)]$ by simply substituting the latent part $E[y(t)|z \neq t]$ with $\mathbb{K}_0$ as the lower bound, and $\mathbb{K}_1$ as the upper bound. Then equation (3) would be bounded as follows,

$$E[y(t_i)|z = t_i]P(z = t_i) + K_0 P(z \neq t) \leq E[y(t_i)] \leq E[y(t_i)|z = t_i]P(z = t_i) + K_1 P(z \neq t) \tag{3.4}$$

These worst-case bounds are not very informative because they are generally very wide, but these no-assumption bounds are interesting because all results that are based on different assumptions on the counterfactual part $E[y(t)|z \neq t]$ should lie within this region. In the following subsections, we will introduce monotone treatment response (MTR) assumption and monotone treatment selection (MTS), and add them into our analysis. While in the IV framework, we will introduce how monotone instrument variable assumption and its combination with MTS, MTR make these bounds more informative. The MTR and MTS assumptions are introduced and derived by Manski (1997) and Manski and Pepper (2009). All these assumptions are used to find the plausible regions of the latent part in equation (3).

Assumption on Response Function MTR

The MTR assumption states that the outcome is a weakly increasing function of the treatment:

$$t_j \geq t_i \rightarrow y(t_j) \geq y(t_i) \tag{3.5}$$

equation (3.4) states that increasing one's Mandarin proficiency weakly increases his or her wage function, which imposes that improving one's language ability would never hurt his or her job performance, ceteris paribus. This assumption is originally derived from the human capital theory. If we think language skill as one type of human capital, people with better language skill are more productive, at least will not hurting his/her productivity. Note, it could, however, be the case that increasing MP does not necessarily improve wages. MTR takes the possibility into consideration. Therefore, it is important to point out that the MTR assumption does not rule out the possibility of MP being not productive at all and that the positive associations have been found in the literature are due to selection and therefore not causation.

The most common case for OLS is based on the "linear response assumption", which means that the hypothetical response of each individual on the treatment is assumed to be the same, captured by the constant estimated value $\hat{\beta}$. The linear model implicitly treats ordinal discrete variable as a continuous variable to estimate the marginal effect. However, this "linear response assumption" is too strong and not always reflecting the real-world scenarios, but with the MTR assumption, we are able to loosen the "linear response assumption" and reflect the real world as close as possible.

After we combine the MTR assumption along with the no-assumption bounds discussed earlier, equation (3.4) becomes:

$$\mathbb{E}[y(t_i)|z \leq t_i]P(z \leq t_i) + \mathbb{K}_0 P(z > t_i) \leq \mathbb{E}[y(t_i)] \leq \mathbb{E}[y(t_i)|z \geq t_i]P(z \geq t_i) + \mathbb{K}_1 P(z \leq t) \tag{3.6}$$

To show how equation 3.6 is obtained from equation 3.4 after imposing MTR, we can divide the sample into three groups for the purpose of illustration: (1) workers with MP less than t, (2) workers with MP equals to t, (3) and workers with MP greater than t. For the second group we observe the effect on mean wages of having MP with level t. For the first group, we know that their observed mean wage is less than or equal to what their mean wage would have been if they had MP level t. So we can use the mean wage we observe for the first group to replace $\mathbb{K}_0$ in equation (3.4). In addition, on the other side we can use the observed mean wage of the third group to replace $\mathbb{K}_1$ in order to tightened the upper bound of no-assumption bounds.

Assumption on Treatment Selections

The second assumption that we impose to the analysis is the Monotone Treatment Selection (MTS) assumption. Under this assumption, workers with better MP have weakly higher mean wage functions than those with lower MP. This can be mathematically shown as follows,

$$t_j > t_i \Longrightarrow E[y(t_u)|z = t_j] \geq E[y(t_u)|z = t_i] \tag{3.7}$$

This assumption relates to the sorting process, which gives us a better scenario on how the treatment groups are distributed to the different individuals in the population, which is commonly known as "ability bias". As discussed earlier, language is one type of human capital. Workers who have better MP could be different from workers with lower MP, such as better MP workers should also have higher ability and better family background, therefore higher wage, on average. For some particular occupations, workers are more productive with better communication ability. Both arguments are consistent with the MTS assumption in our context, but they do not work appropriately with Exogenous Treatment Selection (ETS) assumption, which is commonly assumed under an OLS framework. ETS assumes that treatments are assigned to the population at random, but it does not seem to be the case in our data. The ETS estimates are equivalent to the coefficients obtained by running OLS on log earnings with one dummy variable for each treatment group without adding any covariates.

In order to illustrate the usage of MTS assumption, we again divide the sample into three parts - workers who have MP (1) lower than t($z < t$), (2) equals to t(z=t), and (3) higher than t($z > t$). Since the MP of the first group is below t, we can infer that the mean wage level in the first group would be weakly lower than the mean wage of workers with MP equal to t, from the assumption of MTS. Thus, we are able to use the mean wage level that we observe from the individuals who have MP at level t as an upper bound for the first group. Applying the same logic, we use the mean wage from individuals with MP at level t as the lower bound for the third group. Hence, we construct the new bound under the MTS assumption as follows,

$$E[y(t)|z = t]P(z \geq t) + K_0 P(z < t) \leq E[y(t)] \leq E[y(t)|z = t]P(z \leq t) + K_1 P(z > t) \quad (3.8)$$

MTR + MTS

Since MTR and MTS tighten the no-assumption bounds from different perspectives and angles, if we combine the assumptions of MTR and MTS, we basically assume that the response function is monotone and that selection into treatment is positive simultaneously.

Both assumptions are consistent with human capital theories, thus the combined assumptions should have more identification power on narrowing the bounds. By combining the two assumptions, we obtain the MTR-MTS bounds[4].

$$
\begin{gathered}
E[y(t)|z < t]P(z \leq t) + E[y|z = t]P(z > t) \\
\leq E[y(t)] \leq \\
E[y(t)|z = t]P(z \leq t) + E[y|z > t]P(z > t)
\end{gathered} \tag{3.9}
$$

As we can see from equation (3.9), once we impose the assumptions together, $\mathbb{K}_0$ and $\mathbb{K}_1$ are eliminated. Also, the joint MTR-MTS assumption is testable. Under this assumption, the following should hold[5], for $t_j > t_i$,

$$
\begin{aligned}
E[y|z = t_j] = E[y[t_j]|z = t_j] \geq E[y[t_j]|z = t_i] \\
\geq E[y[t_i]|z = t_i] = E[y|z = t_i]
\end{aligned} \tag{3.10}
$$

Under the joint assumption, the mean wage of people should increase with the hypothetical increase in MP. If this is not the case, the joint assumption of MTR and MTS should be rejected anyway, either or neither assumptions fail in the context.

Monotone Instrumental Variable Assumption

Due to "ability bias", exogenous treatment assumption is likely failed. Endogenous variables are often encountered in empirical work. Hence, the method of instrument variables is widely

[4] For a full derivation of the MTR-MTS bounds, see Manski (1997) and Manski and Pepper (2000)

[5] The first inequality derives from the MTS assumption and the second inequality derives from the MTR assumption

used in the evaluation of treatment effect. In the setting of the traditional Instrument Variable (IV) approach, suppose now we observe not only the workers' MP and their wages but also a variable v, where v is related to MP, and only affect wages through the channel of MP. Though standard IV assumption can aid greatly in identification, the credibility of it is often a matter of no consensus. Under the condition of *mean independence* form of the standard IV condition, the following should hold for all treatments $t \in T$ and all values of the instrument $m \in M$,

$$\mathbb{E}[y(t)|v = m] = \mathbb{E}[y(t)] \tag{3.11}$$

Equation 3.11 means that the wage of the workers should be mean-independent of the instrument variable v. If v satisfies the IV assumption, we can obtain an IV lower bound on $\mathbb{E}[y(t)]$ by taking the maximum lower bound and an IV upper bound by taking the minimum upper bound, which can be shown as follows,

$$max(LB_{E[y(t)|v=m]}) \leq E[y(t)] \leq min(UB_{E[y(t)|v=m]}) \tag{3.12}$$

In reality, it is usually very difficult to find a valid IV which satisfies the mean independent assumption shown in above equation 3.11. Neither there is a consensus way to prove the validity of IV. Here we use a weaker version on IV assumption: the monotone instrumental variable assumption (MIV). A monotone instrumental variable satisfies the following condition,

$$\begin{aligned} m_1 < m_2 < m_3 \rightarrow \mathbb{E}[y(t)|v = m_1] &\leq \mathbb{E}[y(t)|v = m_2] \\ &\leq \mathbb{E}[y(t)|v = m_3] \end{aligned} \tag{3.13}$$

So instead of assuming mean independence, a variable v is an MIV in the sense of mean monotonicity, which allows for a weakly monotone relation between the variable v and the mean wages function of the workers. To use the MIV, we can still divide the sample into subsample based on v and obtain bounds of each sub-group, and then weighted the bounds of each of them to find the bounds of the population. Under three assumption jointly imposed, the bound for $\mathbb{E}[y(t)]$ is as follows,

$$\sum_{v \in m} \mathbb{P}(v = m)\{sup_{m_1 \leq m}(\sum_{k<t} \mathbb{E}[y|v = m_1, z = k]\mathbb{P}(z = k|v = m_1) + \mathbb{E}[y|v = m_1, z = t]\mathbb{P}(z \geq t|v = m_1))\} \leq$$
$$E[y(t)]$$
$$\leq \sum_{v \in m} \mathbb{P}(v = m)\{inf_{m_2 \geq m}(\sum_{k>t} \mathbb{E}[y|v = m_2, z = k]\mathbb{P}(z = k|v = m_2) + \mathbb{E}[y|v = m_2, z = t]\mathbb{P}(z \leq t|v = m_2))\} \tag{3.14}$$

3.4 Data

We utilize data from the 2016 wave of the national- wide China Labor-force Dynamics Survey to empirically investigate the effects. The China Labor-force Dynamics Survey(CLDS) is lunched by Sun Yat-Sen University. It is the first longitudinal social survey targets at China labor force, which covers 29 provinces, and the major focus is on the labor market characteristics.

The key treatment variable is the Mandarin proficiency based on a four-point Likert scale. One question in the survey was about "How do you assess the respondent's putonghua ability". The five possible responses are: (1)Cannot speak Mandarin; (2)Can understand but cannot speak; (3) Speak not very well; (4) Cam speak but with accent and (5) Can speak very well. We integrate (1) and (2) as one group meaning people who cannot speak Mandarin.

Comparing to most other datasets with language measures in China, the key independent variable, Mandarin proficiency level, is objectively reported by the interviewers rather than subjectively self-reported in CLDS data. People who conduct the interviews are required for a tainting beforehand. Thus, the measure in CLDS should be more accurate than subjective responses that based on preferences.

The main outcome variable is the logarithm of monthly wage, including all wage payments,bonuses and allowances each month [6]. In order to measure the average percentage change on wages, we take natural logarithm form of monthly wage for estimations.

The legal working age is China is 16, the Mandatory retirement age are either 50 or 55 for females, and 65 for males, depending on types of job positions. Thus we restrict the samples to individuals aged from 16 to 65. Also, to ensure that our sample only targets at those workers with stronger attachment to the labor market, we exclude those individuals with less than 600 RMB (equivalent to 85 U.S. Dollar)per month from the sample. To further study on the heterogeneous effects, we also gather information , such as age, education, gender, *hukou*, location among others.

Table 3.1 provides descriptive statistics of the Mandarin proficiency (MP). Overall, the Mandarin proficiency for the whole sample is 2.991 out of a 4 scale, which means on average people can speak Mandarin but with accent. However, the big disparities appear when we break down the sample into different sub-groups. People with urban hukou and living in urban cities have much higher proficiency level than people in the rural areas. What's more important is that, it also relates to education. The group with college degree can almost

[6] Due to the incomplete information on working weeks per month and working hours per day, we choose to use monthly wage to avoid measurement error.

speak perfect Mandarin while the lowest education group has poor MP. Mandarin proficiency also matters with ages, younger cohort is significantly better than the older cohorts. The monthly wage is higher for males than females, although proficiency in Mandarin is similar across gender. Table 3.2 illustrates monthly wage for each proficiency group. On average, the Monthly wage positively increases with MP, for both males and females across all levels.

3.5 Results

3.5.1 Parametric Estimation

Before jumping into the main estimation, let's present the results from OLS and IV model first. Table 3.3 shows the OLS estimates for the full sample with different model specifications. For each even column, MP is treated as multilevel variable, and it is treated as a one-dimension variable for each odd column. The first column shows the pure correlation between MP and wage. The marginal return is 21.26% cross all levels. The second column shows the conditional mean for each proficiency group, where we can see an increasing marginal return across levels, ranging from 8% (from t1 to t2) to 59.54% (from t1 to t4). After we control for a set of other control variables (in column 3 and column 4), the overall effect is still positive but smaller magnitude. As MP is endogenous in the model, it may be biased towards zero. For other control variables, OLS indicate that an additional year of schooling is associated with wages being 8.4% higher. Males earn 31% more than females, and urban workers earn 9% higher than rural counterparts monthly. The marriage premium is around 11.2% found in the sample.

As discussed above, the OLS estimates of *MP* is biased because of selection problem. Hence, we also reports IV estimates. We adopt IV following Paolo and Raymond (2012) and Dovì (2019) who use language outside the workplace as instrument for *MP*. The idea behind the IV is that it provides an assessment for one's real Mandarin proficiency but should not directly affect her/his wage. The 2SLS results are shown in table 3.4. First column is the result for the whole sample, the coefficient of MP suggests that the return to Mandarin proficiency is around 32%. Hence, the OLS estimate appears to be downward biased. The 32% return is large but comparable to the previous literature using IV approach. Chiswick and Miller (1995) found economic return to English skills were 41.3% for immigrants in Canada and 57.1% in U.S. Some other papers also found large return from IV approach in different countries, such as Chiswick and Miller (2003), Godoy et al. (2007), ranging from 26.4% to 46.9%.

Disparities in economics return across gender are also found here. Overall, the return for males is 27.24% and 38% for females. Females benefit 10.76% more from returns to *MP*. In all three models, the F-statistics are all above 10, reflecting the IV used here does not suffer from weaker IV problem.

One drawback of using IV approach is that we get a constant effect even *MP* is indeed more than 2 groups. In practice, it's easy to think that the return should be varied on the margins. This part cannot be reflected from the parametric model.

3.5.2 Nonparametric bound Estimation

As we mention above, Our goal is to find the average treatment effect of MP on income wages, which is defined as $\mathbb{E}[y(t_{\mathrm{j}}) - y(t_{\mathrm{i}})]$, where $i \neq j$. We are interested in gains in expected wage when moving from one treatment group to the next, measuring the gains from improving Mandarin proficiency. The estimated bounds of E[y(t)] and average treatment effects (ATE) for the full sample are presented in table 5. The 0.95 bootstrap confidence intervals from 99 repetitions are also attached in the table.

Our dependent variable is a natural log of monthly wage which may vary from 6.4 to 11.7. Thus, to make bounds analysis more feasible, we arbitrarily impose lower and upper bounds of y. We set K_0 equals to 6 and K_1 equals to 11 [7].

First, the worse-case bounds don't rely on any statistical assumptions on $\mathbb{E}[y(t)]$. Under this assumption, the overall effect (from t1 to t4) is from -4.212 to 4.949. And for other marginal effects, the bounds are all somewhere between a negative number to a positive number. These bounds imply that the return can be either negative or positive. Even though these bounds are not informative, but it has already narrower than the no assumption bounds, ranging from [-5,5][8]. The MTR assumption, which assumes that being more proficient in Mandarin would not hurt his/her wage shrinks the bounds further. Especially it rules out the negative effect by design. Instead, MTS mainly tightens the upper bounds of the treatment effects. The upper bound of the global effect is now shrunk to 0.595, which is much narrower than under worst-case alone.

[7] K_0 and K_1 are unnecessary for the MTR+MTS and MIV+MTR+MTS bounds because they are not a function of K_0 and K_1. The final findings of the paper are qualitatively unaffected.

[8] The lower bound of No-assumption bound equals to K_0-K_1, and upper bound equals to K_1-K_0.

In above, we compare the bounds under MTR and MTS separately with the worst-case bound. We find that each of them provides identification power in its own way. Each of them gives narrower bounds than WC bounds. If we impose MTR and MTS jointly, the bounds are tightened than any of them imposed separately. Now all ATEs are between zero and one, which make the estimations having more economics sense. Without any further information, the bounds under MTR+MTS is sharp. The upper bound of the global effects from t1 to t4 and t1 to t3 are 59.5% and 43.6% accordingly, while the lower bounds for both are not able to exclude zero. Finally, the joint MIV+MTS+MTR assumptions yields the tightest bounds of ATEs. The more important finding is that the lower bound of the global effect is statistically excluded from zero turns to be positive. The overall premium on wages from Mandarin proficiency is between 10.5% and 49.9%. Namely, if one accepts to impose MIV,MTR and MTS simultaneously, she/he can infer that the population average treatment effect on wages from no MP to have decent MP is at least 10.5% and not more than 49.9%. At other margins, except from t1 to t2, the lower bounds are all above zero but less than 10.5%. Nonetheless, they are not statistically different from zero. Across all margins, the marginal effects display a diminishing trend in terms of the upper bound estimations.

Feasibility of nonparametric assumptions

The OLS model above may produce biased and inconsistent estimate of the EP effect on wages when the choice of investing on MP and wages are simultaneously influenced by some unobservable heterogeneity specific to the individuals.

Nonparametric bound analysis requires a couple of relatively weak nonparametric assumption to identify the upper and lower bounds of the causal effects of MP. In this study, the monotone treatment assumption(MTR) implies that improving someone's MP cannot lower her wage, which means that each worker's wage function is weakly increasing in conjectured MP. This assumption can be explained by the human capital accumulation theory, language skill as a human capital, can boost one's productivity in the labor market, at least will not hurt one's productivity. The monotone treatment selection(MTS) assumption implies that workers with higher MP have weakly higher mean's wage function than workers with lower MP. (Exclude negative selection on job types). To some extent, MTS assumption is a natural way to deal with the selection issue, before taking IV into consideration.

Manski and Pepper (2009) suggest that MTR-MTS is a testable hypothesis jointly. Under this hypothesis, the average outcome for the realized treatment must be a weakly increasing function of the realized treatment. Otherwise we should reject this hypothesis, and cannot impose them jointly. From table two, we see the conditional means are increasing along the treatment groups. Thus, it makes sense to jointly impose MTR-MTS assumptions in our context in terms of the whole sample.

Considering IV along with MTR-MTS assumptions can further tighten the bounds. The credibility of mean independence conditions has often been subject to disagreement in empirical research. Most of time, there is no consensus conclusion on the validation of the instrument variables. Instead of mean-independence conditions, the MIV assumption does not strictly require fully exogenous in the model. In this paper, the MIV we use is the language used out of the work space. It has three categories: (1) speak one's hometown dialects; (2) speak local dialects; (3) speak Mandarin. We refer group 2 is monotonically better than group 1 because people who speak their hometown are more likely to be rural-urban migrants and stay in the enclaves with people from the same village[9]. Of course, the Mandarin proficiency of third group should be positive monotonic to group two.

3.6 Disaggregated Results of the Bounds Analysis

Above sections mainly discuss the population average treatment effect. To assess the potential heterogenous effects, we now turn to the subgroup analysis.

By Gender

To investigate the MP premiums by gender, we divide full sample to male and female. The bounds results are shown in table 3.6[10]. Overall, both males and females can earn more monthly wages from having better MP in the Chinese labor market. For males, the global effect is between 11.2% and 51%, inferring an average marginal effect between 3.7% and 17%. While females benefit more from language skills than males in China. The global return for female is between 10.5% and 61.6%, inferring an average marginal return from 3.5% to 20.5%. From table 3.1, we see that the average wage for male workers is indeed higher than female workers, but the returns to MP is higher for females rather than males. First possible explanation is that language proficiency has different gender effects on wages through occupational choice found in Mora and Davila (1998),Williams (2011), and Wang,

[9] In order to test this conjecture, we calculate the percentages of agricultural hukou owned who usually speak Mandarin less proficiently. in the three groups. 80% of the first group own a agricultural hukou, 68% of the second group own a agricultural hukou, and only 33% of the third group own a agricultural hukou.

[10] We only show the estimation results under ETS, MTR+MTS and MTR+MTS+MIV. Other estimation results are available upon requested.

Smyth and Cheng (2017). Female workers are more likely to work in service occupation and occupation that needed a lot communication opportunities, while males are more likely to work in technician, production types of occupation which require less communication skills. Fan (2000) showed evidence on that rural-urban female migrants have intention of marrying local men in the host city. If migrants female married to a local urban male, then she will have more opportunities to practice Mandarin at home and gain more proficiency and less accent gradually. "Chiswick and Miller (1998) argued that one reason for low returns to those whose English was their second language in the United States was that they tended to have fewer labour market opportunities because they were concentrated in urban enclaves reflecting their language deficiencies."

By Age Group

We split the sample to two age groups, one is age from 16 to 45 and another is from 46-64. The first group represents the younger cohorts who were in their early and mid careers, and the second group represents the older cohorts who were in their late careers. We compare how do the returns vary across the two age groups. The results are shown in table 3.7. The conditional mean (ETS) tells us the correlation between MP and wages are positive in both groups as $\mathbb{E}[y(t)]$ increases along the treatment groups. The overall effects are 57.9% and 69.2% for these two groups under ETS accordingly. MTR+MTS provides sharp bounds without any information rather than the distribution of MP and wages. Under MTR+MTS, all the lower bounds are excluded from negative numbers by the design of MTR. After impose MIV jointly with MTR+MTS, the results are more fruitful and informative for inference.

For young and mid-age workers, the overall return to MP is from 10.8% to 28.6%, and it is also statistically different from zero. On the other hand, the upper bound of the global effect for the older group is larger, at 47.7%, but the lower bound is not excluded from zero. Thus, whether MP benefits on wages for older workers in China is inconclusive.

Urban VS Rural

We also find positive return in both Rural and Urban areas. The evidence is shown in table 3.8. The conditional mean under exogenous treatment effect (ETS) show that wages are monotonically increasing along the treatments, which secure us to impose MTR+MTS jointly. Under MTR+MTS+MIV, the bounds are much informative than MTR+MTS bounds, particular regarding the lower bounds. The global effect in rural China is between 9.8% and 29.7%, while it is larger in urban China, where the global effect can be up to 57.1%. Zero lower bounds are statistically excluded from the estimations for both groups.

In order to gain Economics growth for the entire country, and at the same time, to help the poor rural areas get rid of poverty, China government has realized the importance of spreading the effectiveness of using putonghua in rural China back to years ago. The popularization of putonghua is always one of the ultimate goals that the government care and wants to achieve. Based on our estimation results, the economic values of Mandarin in rural China indeed exist. From the perspective of long run development, we believe language integration is a very important and effective way for economic growth. Due to the large return that is found in urban China, rural-urban migrants can be more competitive in the labor market if with better communication capacity in Mandarin.

3.7 Mechanism analysis

Based on the estimations from both parametric and non-parametric approach, we have found MP can affect wages in a positive way. The culture and language are diversity in China, the relationship between Mandarin and the variety of dialects nationwise is somehow alike the context of the relationship between English and other language among immigrants who are not native born English speakers. Owning a good proficiency in the core language can bring many good signals in the labor market. As we discussed in the earlier section, Yao and van Ours (2019) and Grogger (2011) found that the most two potential channels that language proficiency can affect wages are through productivity and through that employers discriminate against poor speakers. With the help of the information in the dataset, we can test if this two channels can explain the positive MP premium in China. We use one question from the survey to implicitly test one's job match, which asks the respondent "Do you think you can showcase your ability and talent in the current job position". Job match can also indirectly relate to one's productivity. For the other channel, we use promotion opportunity as proxy for discrimination to test if discrimination against less proficiency workers commonly happen. One questions asks each respondent "Do you think you have promotion opportunities?" For both questions, the answers are subjectively reported, which is based on a 5 Likert scale. We are going to use ordered Logit model for this section. The results are shown in table 3.8 and 3.9, accordingly. Both male and female workers seem have the same pattern on the satisfaction on job matched and promotion opportunity. Workers with higher MP are more likely to feel satified on job match. Also, we can see that education plays more important role in determining the satisfaction than Mandarin proficiency does. on the contrary, MP does not affect workers' satisfaction on promotion opportunity given they already have jobs, for both genders. With this in mind, Dovì (2019) found that a one-standard-deviation increase in Mandarin proficiency decreases unemployment probabilities by 5%, which infers that discrimination against poor Mandarin proficiency may happen during the job search rather than having a job already.

3.8 Comparing the results

As Chiswick and Miller (1995)(p. 277) describe it: "The coefficient in the instrumental variables analysis shows considerable instability. It is positive... but the magnitudes are surprisingly large." We also find the a substantial return of 31.59% from our IV estimation. Given the fact that the usual education return generally ranges from 12% to 16% concluded in Ashenfelter and Krueger (1994). 31.59% seems too big here as for language return. The overall bounding effect is big and wide as well. Yet, if we assume the returns are constant across all levels. The bounds for the average marginal effect are actually between 3.5% and 16.6% , which is very likely comparable with the return to education.

From Figure 3.1, we visually compare the estimation results between the IV model and the bounding model. We can observe that the effect at each margin varies greatly, where IV estimation can not capture the variations. In addition, IV estimation overestimates the return from no Mandarin skills (t1) to have little Mandarin Skills (t2).

3.9 Conclusion

Ever since its economic reforms began in 1978, China has enjoyed remarkable economic growth. but it is accompanied by both relatively low standards of living and high inequality. One strategy the China government has been implemented since 1958 is to promote language integration. In 1987, the National Conference on Language and Script decided that Mandarin was to become the language used in instruction at school, the working language of government at all levels, the language used in media, and the lingua franca among speakers of different local dialects. To shed light on the effectiveness of promoting the use of Mandarin, this paper analyses its economic value in China labor market. we use both instrument variable and non-parametric bounding approaches for estimations. Our results show that the overall return is positive, ranging from 10.5% to 49.9% for the whole sample. One driving mechanism we find is that workers with good Mandarin proficiency are more likely to find jobs matched to their expertise, thereafter increasing productivity at the workplace. We also find that the returns are heterogeneous in different sub-groups. (1) Female benefit more than male, which also infers that the soft skill like language proficiency can help to reduce the gender wage gap; (2)The return is statistically significant and positive for young and mid-age workers, but no enough evidence for older workers; (3) The return is larger for urban population than rural population. In addition, we find the IV method cannot extrapolate the marginal gains along with treatment groups and the coefficient is incredibly large, which overestimates the economics return to Mandarin proficiency.

In terms of policy implication, our results suggest that language integration could be an effective way to promote both efficiency and equality in China. From the government side, it should not only promote the use of Mandarin from the angles of national language integration and macro-level economic growth, but also emphasize individual gains from learning Mandarin, such as wage benefits and better standards of living.

References

Adamson, Bob. 2001. "English with Chinese characteristics: China's new curriculum." *Asia Pacific Journal of Education*, 21(2): 19–33.

Adamson, Bob, and Paul Morris. 1997. "The English curriculum in the People's Republic of China." *Comparative Education Review*, 41(1): 3–26.

Akbulut-Yuksel, Mevlude, Hoyt Bleakley, and Aimee Chin. 2011. "The effects of English proficiency among childhood immigrants: are Hispanics different?" In *Latinos and the Economy*. 255–283. Springer.

Ashenfelter, Orley, and Alan Krueger. 1994. "Estimating the returns schooling using a nex sample of twins." *American Economic Review*, 84(5): 1157–73.

Azam, Mehtabul, Aimee Chin, and Nishith Prakash. 2013. "The returns to English-language skills in India." *Economic Development and Cultural Change*, 61(2): 335–367.

Becker, Gary S. 1964. "Human capital theory." *Columbia, New York*, 1964.

Berthélemy, Jean-Claude, and Sylvie Demurger. 2000. "Foreign direct investment and economic growth: theory and application to China." *Review of development economics*, 4(2): 140–155.

Blau, Francine D, Lawrence M Kahn, and Kerry L Papps. 2011. "Gender, source country characteristics, and labor market assimilation among immigrants." *The Review of Economics and Statistics*, 93(1): 43–58.

Bleakley, Hoyt, and Aimee Chin. 2004. "Language skills and earnings: Evidence from childhood immigrants." *Review of Economics and statistics*, 86(2): 481–496.

Blundell, Richard, Amanda Gosling, Hidehiko Ichimura, and Costas Meghir. 2007. "Changes in the distribution of male and female wages accounting for employment composition using bounds." *Econometrica*, 75(2): 323–363.

Böhlmark, Anders. 2008. "Age at immigration and school performance: A siblings analysis using Swedish register data." *Labour Economics*, 15(6): 1366–1387.

Bolton, Kingsley, and David Graddol. 2012. "English in China today."

Borjas, George J. 2000. "Ethnic enclaves and assimilation." *Swedish Economic policy review*, 7(2): 89–122.

Borjas, George J. 2015. "The slowdown in the economic assimilation of immigrants: Aging and cohort effects revisited again." *Journal of Human Capital*, 9(4): 483–517.

Branstetter, Lee, and Nicholas Lardy. 2006. "China's embrace of globalization." National Bureau of Economic Research.

Carliner, Geoffrey. 1996. "The wages and language skills of US immigrants." National Bureau of Economic Research.

Casale, Daniela, and Dorrit Posel. 2011. "English language proficiency and earnings in a developing country: The case of South Africa." *The Journal of Socio-Economics*, 40(4): 385–393.

Center, Pew Research. 2016. "The state of American jobs."

Chernozhukov, Victor, and Christian Hansen. 2006. "Instrumental quantile regression inference for structural and treatment effect models." *Journal of Econometrics*, 132(2): 491–525.

Chiswick, Barry R. 1991. "Speaking, reading, and earnings among low-skilled immigrants." *Journal of labor economics*, 9(2): 149–170.

Chiswick, Barry R. 1998. "Hebrew language usage: Determinants and effects on earnings among immigrants in Israel." *Journal of Population Economics*, 11(2): 253–271.

Chiswick, Barry R, and Paul W Miller. 1995. "The endogeneity between language and earnings: International analyses." *Journal of labor economics*, 13(2): 246–288.

Chiswick, Barry R, and Paul W Miller. 1998. "The economic cost to native-born Americans of limited English language proficiency." *Report Prepared for the Center for Equal Opportunity, available www. ceousa. org/earnings. html.*

Chiswick, Barry R, and Paul W Miller. 2003. "The complementarity of language and other human capital: immigrant earnings in Canada." *Economics of Education review*, 22(5): 469–480.

Chiswick, Barry R, and Paul W Miller. 2010. "Occupational language requirements and the value of English in the US labor market." *Journal of Population Economics*, 23(1): 353–372.

Cortazzi, Martin, and Lixian Jin. 1996. "English teaching and learning in China." *Language teaching*, 29(2): 61–80.

Daily, China. 2010. "Rush to learn English fuels quality issues." *Online at¡ http://www. chinadaily. com. cn/china/2010-08/05/content_11098499. htm¿(Accessed January 10, 2014).*

De Haan, Monique. 2011. "The effect of parents? schooling on child?s schooling: a nonparametric bounds analysis." *Journal of Labor Economics*, 29(4): 859–892.

Depalo, Domenico. 2018. "Identification issues in the public/private wage gap, with an application to Italy." *Journal of Applied Econometrics*, 33(3): 435–456.

Dovì, Max-Sebastian. 2019. "Does higher language proficiency decrease the probability of unemployment? Evidence from China." *China Economic Review*, 54: 1–11.

Dustmann, Christian, and Arthur van Soest. 2001. "Language fluency and earnings: Estimation with misclassified language indicators." *Review of Economics and Statistics*, 83(4): 663–674.

Edwards, Jette G. Hansen. 2017. "The Rise of English in China: A Threat to China?s National Unity?" *Georgetown Journal of International Affairs.*

Fan, C Cindy. 2000. "Migration and gender in China." *China review*, 423–454.

Ferrer, Ana, David A Green, and W Craig Riddell. 2006. "The effect of literacy on immigrant earnings." *Journal of Human Resources*, 41(2): 380–410.

Flores, Antonio. 2017. "How the US Hispanic population is changing." *Pew Research Center*, 18.

Flores, Antonio, Gustavo López, and Jynnah Radford. 2017. "2015, Hispanic population in the United States statistical portrait." *Washington, DC: Pew Research Center. Retrieved October*, 30: 2019.

Gao, Wenshu, and Russell Smyth. 2011. "Economic returns to speaking ?standard Mandarin?among migrants in China's urban labour market." *Economics of Education Review*, 30(2): 342–352.

Gerfin, Michael, and Martin Schellhorn. 2006. "Nonparametric bounds on the effect of deductibles in health care insurance on doctor visits–Swiss evidence." *Health economics*, 15(9): 1011–1020.

Ginsburgh, Victor A, and Juan Prieto-Rodriguez. 2011. "Returns to foreign languages of native workers in the European Union." *ILR Review*, 64(3): 599–618.

Giustinelli, Pamela. 2011. "Non-parametric bounds on quantiles under monotonicity assumptions: with an application to the Italian education returns." *Journal of Applied Econometrics*, 26(5): 783–824.

Godoy, Ricardo, Victoria Reyes-García, Craig Seyfried, Tomás Huanca, William R Leonard, Thomas McDade, Susan Tanner, and Vincent Vadez. 2007.

"Language skills and earnings: Evidence from a pre-industrial economy in the Bolivian Amazon." *Economics of Education Review*, 26(3): 349–360.

Gonzalez, Libertad. 2005. "Nonparametric bounds on the returns to language skills." *Journal of Applied Econometrics*, 20(6): 771–795.

Grenier, Gilles. 1984. "The effects of language characteristics on the wages of Hispanic-American males." *Journal of Human Resources*, 35–52.

Grogger, Jeffrey. 2011. "Speech patterns and racial wage inequality." *Journal of Human Resources*, 46(1): 1–25.

Guo, Qian, and Wenkai Sun. 2014. "Economic returns to English proficiency for college graduates in mainland China." *China Economic Review*, 30: 290–300.

Hakuta, Kenji, Yuko Goto Butler, and Daria Witt. 2000. "How Long Does It Take English Learners To Attain Proficiency?."

Hotz, V Joseph, Charles H Mullin, and Seth G Sanders. 1997. "Bounding causal effects using data from a contaminated natural experiment: Analysing the effects of teenage childbearing." *The Review of Economic Studies*, 64(4): 575–603.

Huang, Jian, Henriette Maassen Van den Brink, and Wim Groot. 2009. "A meta-analysis of the effect of education on social capital." *Economics of education review*, 28(4): 454–464.

Hu, Guangwei. 2005. "English language education in China: Policies, progress, and problems." *Language policy*, 4(1): 5–24.

Jiang, Zemin. 2000. "Law of the People?s Republic of China on the Standard Spoken and Written Chinese Language (Order of the President No. 37)."

Jin, Lixian, and Martin Cortazzi. 2002. "English language teaching in China: A bridge to the future." *Asia Pacific Journal of Education*, 22(2): 53–64.

Kang, Changhui. 2011. "Family size and educational investments in children: Evidence from private tutoring expenditures in South Korea." *Oxford Bulletin of Economics and Statistics*, 73(1): 59–78.

Koenker, Roger. 2000. "Galton, Edgeworth, Frisch, and prospects for quantile regression in econometrics." *Journal of Econometrics*, 95(2): 347–374.

Koenker, Roger, and Gilbert Bassett Jr. 1978. "Regression quantiles." *Econometrica: journal of the Econometric Society*, 33–50.

Kossoudji, Sherrie A. 1988. "English language ability and the labor market opportunities of Hispanic and East Asian immigrant men." *Journal of Labor Economics*, 6(2): 205–228.

Kreider, Brent, John V Pepper, Craig Gundersen, and Dean Jolliffe. 2012. "Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported." *Journal of the American Statistical Association*, 107(499): 958–975.

Krogstad, Jens Manuel, Renee Stepler, and Mark Hugo Lopez. 2015. "English proficiency on the rise among Latinos."

Ku, Hyejin, and Asaf Zussman. 2010. "Lingua franca: The role of English in international trade." *Journal of Economic Behavior & Organization*, 75(2): 250–260.

Lingguang, Huo, and Chen Yuanyuan. 2017. "The Estimation of Economic Returns to Proficiency in English in China." *South China Journal of Economics*, , (12): 5.

Manski, C, and J Pepper. 2000. "Monotone treatment response, with an application to the returns to schooling." *Econometrica*, 68: 997–1012.

Manski, Charles F. 1989. "Anatomy of the selection problem." *Journal of Human resources*, 343–360.

Manski, Charles F. 1990. "The use of intentions data to predict behavior: A best-case analysis." *Journal of the American Statistical Association*, 85(412): 934–940.

Manski, Charles F. 1997. "Monotone treatment response." *Econometrica: Journal of the Econometric Society*, 1311–1334.

Manski, Charles F. 1999. *Identification problems in the social sciences.* Harvard University Press.

Manski, Charles F, and C Sims. 1994. "The selection problem." Vol. 1, 143–70.

Manski, Charles F, and John V Pepper. 2009. "More on monotone instrumental variables." *The Econometrics Journal*, 12(suppl_1): S200–S216.

Mariotti, Martine, and Juergen Meinecke. 2015. "Partial identification and bound estimation of the average treatment effect of education on earnings for South Africa." *Oxford Bulletin of Economics and Statistics*, 77(2): 210–233.

Massey, Douglas S, and Julia Gelatt. 2010. "What happened to the wages of Mexican immigrants? Trends and interpretations." *Latino Studies*, 8(3): 328–354.

McManus, Walter S. 1985. "Labor market costs of language disparity: An interpretation of Hispanic earnings differences." *The American Economic Review*, 75(4): 818–827.

McManus, Walter, William Gould, and Finis Welch. 1983. "Earnings of Hispanic men: The role of English language proficiency." *Journal of Labor Economics*, 1(2): 101–130.

Mincer, Jacob. 1974. "Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2."

Mora, Marie T, and Alberto Davila. 1998. "Gender, earnings, and the English skill acquisition of Hispanic workers in the United States." *Economic Inquiry*, 36(4): 631–644.

Myers, Dowell, Xin Gao, and Amon Emeka. 2009. "The gradient of immigrant age-at-arrival effects on socioeconomic outcomes in the US." *International Migration Review*, 43(1): 205–229.

Neeley, Tsedal. 2012. "Global business speaks English." *Harvard business review*, 90(5): 116–124.

Orrenius, Pia, and Madeline Zavodny. 2011. "Trends in poverty and inequality among Hispanics."

Overholt, William H. 2004. *China and globalization.* RAND.

Paolo, Antonio Di, and Josep Lluís Raymond. 2012. "Language knowledge and earnings in Catalonia." *Journal of Applied Economics*, 15(1): 89–118.

Pepper, John V. 2000. "The intergenerational transmission of welfare receipt: A nonparametric bounds analysis." *Review of Economics and Statistics*, 82(3): 472–488.

Preston, Julia. 2007. "Latino immigrants? children found grasping English." *New York Times*, 30.

Reimers, Cordelia W. 1983. "Labor market discrimination against Hispanic and black men." *The review of economics and statistics*, 570–579.

Reuters. 2020. "China?s 2019 foreign direct investment grew most in two years, but outbound investment fell 6 per cent." *South China Morning Post*.

Richard, Blundell, Lorraine Dearden, and Barbara Sianesi. 2005. "Evaluating the impact of education on earnings in the UK: models, methods and results from the NCDS." *Journal of the Royal Statistical Society: Series, A (Statistics in Society)*, 168(3): 473–512.

Ross, Heidi. 1992. "Foreign language education as a barometer of modernization." *Education and modernization: The Chinese experience*, 239–254.

Saiz, Albert, and Elena Zoido. 2005. "Listening to what the world says: Bilingualism and earnings in the United States." *Review of Economics and Statistics*, 87(3): 523–538.

Saiz, Albert, Elena Zoido, et al. 2002. "The returns to speaking a second language." Federal Reserve Bank of Philadelphia.

Shields, Michael A, and Stephen Wheatley Price. 2002. "The English language fluency and occupational success of ethnic minority immigrant men living in English metropolitan areas." *Journal of population Economics*, 15(1): 137–160.

Tainer, Evelina. 1988. "English language proficiency and the determination of earnings among foreign-born men." *Journal of Human Resources*, 108–122.

Tang, Chaoju, and Vincent J Van Heuven. 2009. "Mutual intelligibility of Chinese dialects experimentally tested." *Lingua*, 119(5): 709–732.

Toomet, Ott. 2011. "Learn English, not the local language! Ethnic Russians in the Baltic States." *American Economic Review*, 101(3): 526–31.

Tsen, Wong Hock. 2006. "Granger causality tests among openness to international trade, human capital accumulation and economic growth in China: 1952–1999." *International Economic Journal*, 20(3): 285–302.

Wang, Chunbei, and Le Wang. 2011. "Language skills and the earnings distribution among child immigrants." *Industrial Relations: A Journal of Economy and Society*, 50(2): 297–322.

Wang, Haining, Russell Smyth, and Zhiming Cheng. 2017. "The economic returns to proficiency in English in China." *China Economic Review*, 43: 91–104.

Wang, Haining, Zhiming Cheng, and Russell Smyth. 2019. "Health outcomes, health inequality and Mandarin proficiency in urban China." *China Economic Review*, 56: 101305.

Williams, Donald R. 2011. "Multiple language usage and earnings in Western Europe." *International Journal of Manpower*.

Xie, Yu, and Margaret Gough. 2011. "Ethnic enclaves and the earnings of immigrants." *Demography*, 48(4): 1293–1315.

Yao, Yuxin, and Jan C van Ours. 2019. "Daily dialect-speaking and wages among native Dutch speakers." *Empirica*, 46(4): 653–668.

Figures

Chapter 1 Figures

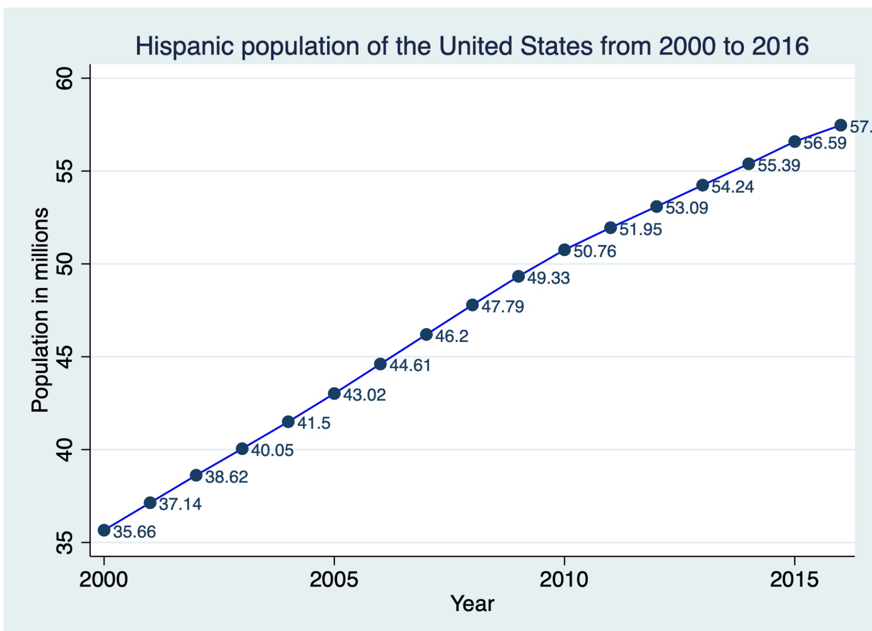


Figure 1.1: The Trend of Hispanic population in US

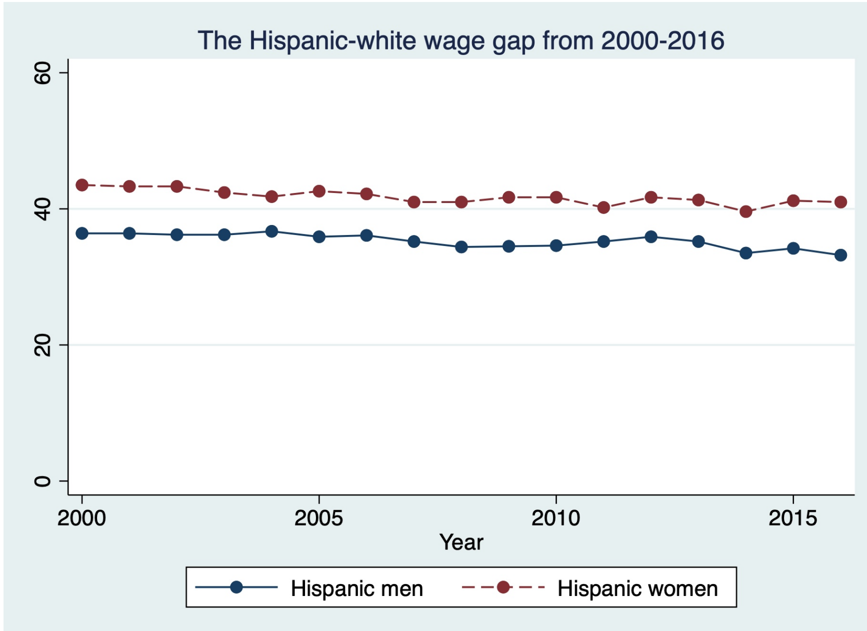
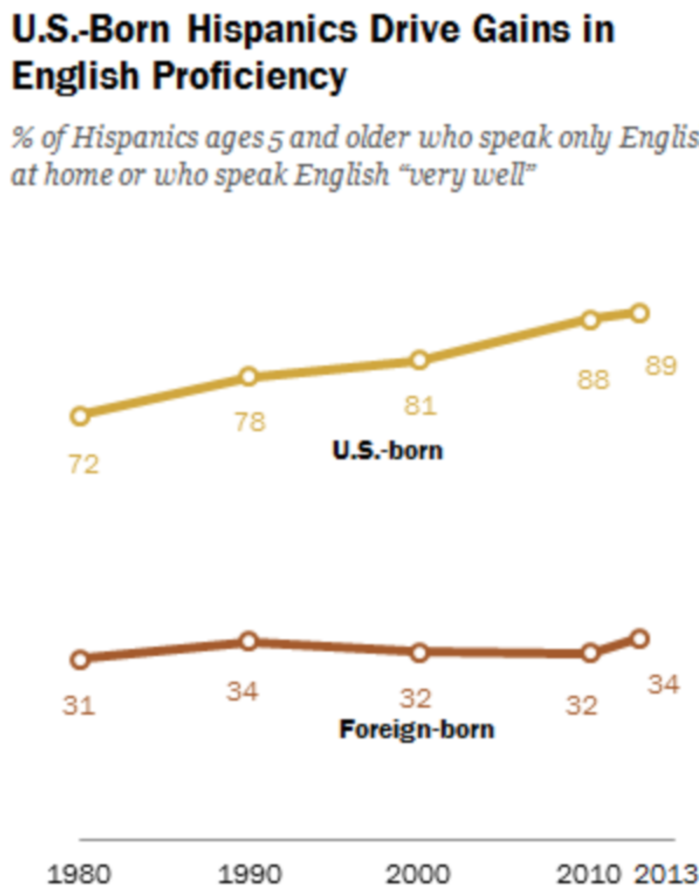


Figure 1.2: Comparison on wages between Whites and Hispanics
Note: The wages compared are average hourly wages and the population is full-time workers ages 18–64. The wage gap is how much less, in percent terms, the average member of each identified subgroup makes less than the average non-Hispanic white man



Source: Pew Research Center

Figure 1.3: The English Proficiency trend of Hispanics in US

Chapter 2 Figures

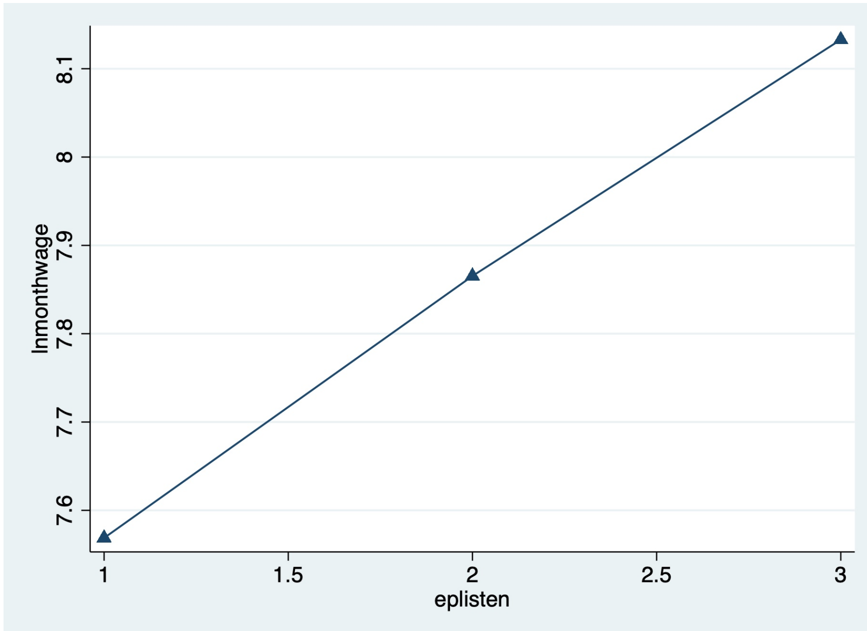


Figure 2.1: Mean of (ln)monthly wage by different English listening abilities

Notes: The horizontal axis represents the proficiency level of English listening. The vertical axis represents the logarithmic form of the monthly wage.

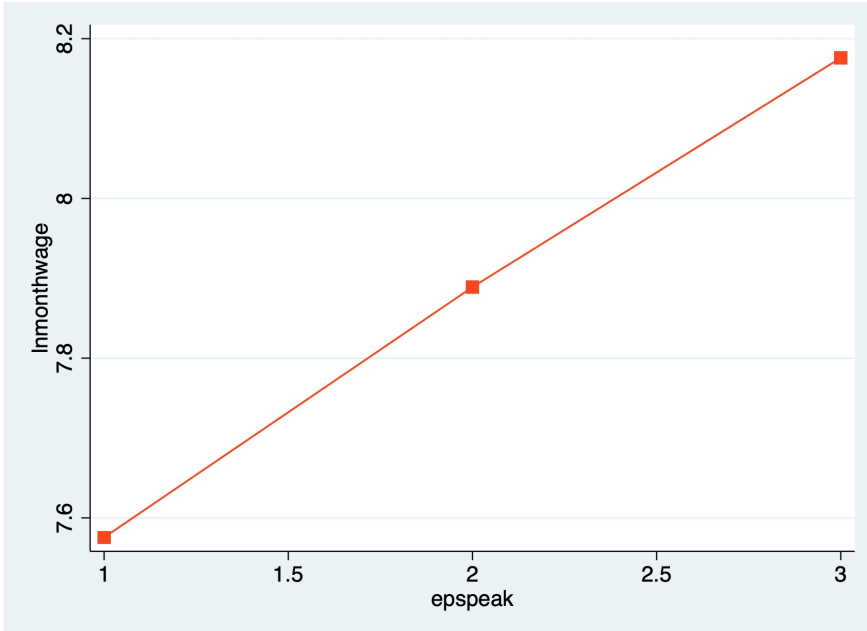


Figure 2.2: Mean of (ln)monthly wage by different English speaking abilities

Notes: The horizontal axis represents the proficiency level of English speaking. The vertical axis represents the logarithmic form of the monthly wage.

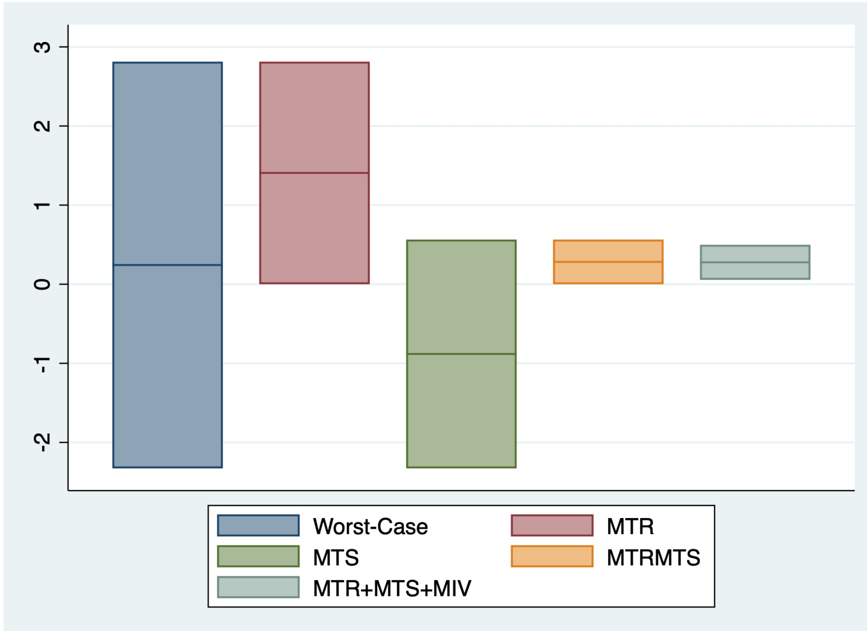


Figure 2.3: Average treatment effect of English listening ability on wages(t3-t1)

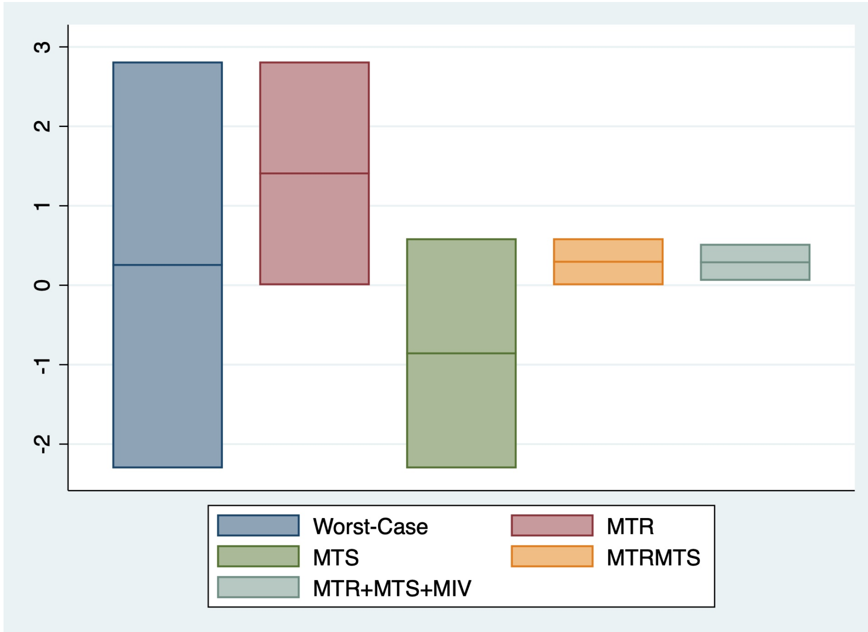


Figure 2.4: Average treatment effects of English speaking ability on wages (t3-t1)

Chapter 3 Figures

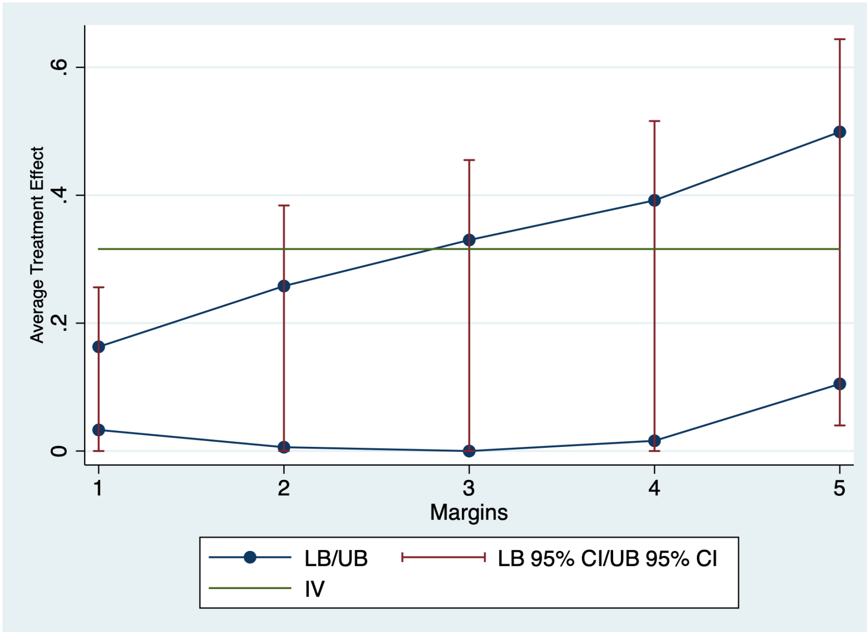


Figure 3.1: Comparison of results: IV VS. Non-parametric bound
Notes: Margin 1 represents the marginal effect from "cannot speak Mandarin" to "cannot speak very fluently" (t_1 to t_2); Margin 2 represents the marginal effect from "cannot speak very fluently" to "can speak fluently" (t_2 to t_3); Margin 3 represents the marginal effect from "can speak fluently" to "can speak very fluently" (t_3 to t_4). Marginal 4 and 5 represent the global effects, from "cannot speak Mandarin" to "can speak fluently" (from t_1 to t_3) and "cannot speak Mandarin" to "can speak very fluently" (from t_1 to t_4), respectively.

Tables

Chapter 1 Tables

Table 1.1: Summary Statistics: English proficiency and average earnings

Treatment	1	2	3	4	Totals
Year 2000					
Observations	8,370	15,024	12,919	18,332	54645
% observations	0.153	0.275	0.236	0.335	1
Average hourly wage	9.923	10.897	12.907	14.708	12.501
Average logwage	1.998	2.127	2.303	2.415	2.245
Average Education	1.231	1.419	1.782	2.257	1.757
Average Experience	21.699	21.134	19.973	16.550	19.408
Year 2010					
Observations	10,075	21,521	18,825	25,232	75653
% observations	0.133	0.284	0.249	0.334	1
Average hourly wage	11.040	12.611	15.791	19.901	15.624
Average logwage	2.214	2.346	2.534	2.712	2.497
Average Education	1.311	1.534	1.964	2.585	1.962
Average Experience	22.095	21.363	19.551	15.087	18.916

Notes: Education has four possible values: 1. High school dropout, 2. High school and equivalent degree, 3. Some college experiences, 4. Bachelor degree and above.

Table 1.2: Bounds of ln(hourly wage) and average treatment effect under ETS & Worst-Case Assumptions

Year	Assumption	ETS	95% conf.in.		Worst-Case		95% Conf.in.	
2000			LB	UB	LB	UB	LB	UB
	E[y(t1)]	1.998	1.985	2.011	1.153	4.540	1.149	4.548
	E[y(t2)]	2.127	2.115	2.139	1.310	4.210	1.306	4.225
	E[y(t3)]	2.303	2.292	2.313	1.308	4.362	1.305	4.369
	E[y(t4)]	2.415	2.403	2.427	1.475	4.133	1.468	4.142
	E[y(t4)]-E[y(t3)]	0.112	0.090	0.135	-2.888	2.825	-2.900	2.837
	E[y(t3)]-E[y(t2)]	0.176	0.152	0.198	-2.902	3.052	-2.920	3.063
	E[y(t2)]-E[y(t1)]	0.129	0.104	0.155	-3.230	3.057	-3.243	3.076
	E[y(t3)]-E[y(t1)]	0.305	0.280	0.329	-3.232	3.209	-3.244	3.220
	E[y(t4)]-E[y(t1)]	0.417	0.391	0.442	-3.066	2.980	-3.080	2.993
	Assumption	ETS	95% conf.in.		Worst-Case		95% Conf.in.	
2010			LB	UB	LB	UB	LB	UB
	E[y(t1)]	2.214	2.204	2.224	1.162	4.629	1.159	4.634
	E[y(t2)]	2.346	2.340	2.353	1.383	4.245	1.380	4.252
	E[y(t3)]	2.534	2.527	2.541	1.382	4.386	1.377	4.393
	E[y(t4)]	2.712	2.704	2.720	1.571	4.237	1.564	4.248
	E[y(t4)]-E[y(t3)]	0.178	0.163	0.193	-2.815	2.855	-2.829	2.871
	E[y(t3)]-E[y(t2)]	0.187	0.174	0.201	-2.863	3.003	-2.876	3.013
	E[y(t2)]-E[y(t1)]	0.132	0.116	0.149	-3.246	3.083	-3.254	3.094
	E[y(t3)]-E[y(t1)]	0.320	0.303	0.337	-3.247	3.225	-3.257	3.234
	E[y(t4)]-E[y(t1)]	0.498	0.480	0.516	-3.058	3.075	-3.070	3.089

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included. (99 repetitions)

Table 1.3: Bounds of ln(hourly wage) and average treatment effect under MTR,MTS and joint MTR & MTS Assumptions

Year	Assumption	ETS		95% conf.in.			MTR		95% Conf. in.	
2000				LB	UB		LB	UB	LB	UB
	E[y(t1)]	1.998		1.985	2.011	E[y(t1)]	1.153	2.245	1.149	2.252
	E[y(t2)]	2.127		2.115	2.139	E[y(t2)]	1.463	2.705	1.458	2.713
	E[y(t3)]	2.303		2.292	2.313	E[y(t3)]	1.771	3.495	1.764	3.511
	E[y(t4)]	2.415		2.403	2.427	E[y(t4)]	2.245	4.133	2.239	4.145
	E[y(t4)]-E[y(t3)]	0.112		0.090	0.135	E[y(t4)]-E[y(t3)]	0.000	2.362	0.000	2.381
	E[y(t3)]-E[y(t2)]	0.175		0.152	0.198	E[y(t3)]-E[y(t2)]	0.000	2.032	0.000	2.053
	E[y(t2)]-E[y(t1)]	0.129		0.104	0.155	E[y(t2)]-E[y(t1)]	0.000	1.552	0.000	1.564
	E[y(t3)]-E[y(t1)]	0.305		0.280	0.329	E[y(t3)]-E[y(t1)]	0.000	2.342	0.000	2.362
	E[y(t4)]-E[y(t1)]	0.417		0.391	0.442	E[y(t4)]-E[y(t1)]	0.000	2.980	0.000	2.996
	Assumption	MTS		95% Confi.in.			MTR+MTS		95% conf.in.	
		LB	UB	LB	UB		LB	UB	LB	UB
	E[y(t1)]	1.998	4.540	1.983	4.549	E[y(t1)]	1.998	2.245	1.983	2.250
	E[y(t2)]	1.955	3.770	1.942	3.782	E[y(t2)]	2.108	2.265	2.097	2.271
	E[y(t3)]	1.745	3.207	1.734	3.224	E[y(t3)]	2.208	2.340	2.198	2.347
	E[y(t4)]	1.475	2.415	1.471	2.423	E[y(t4)]	2.245	2.415	2.239	2.423
	E[y(t4)]-E[y(t3)]	-1.733	0.670	-1.753	0.689	E[y(t4)]-E[y(t3)]	0.000	0.207	0.000	0.224
	E[y(t3)]-E[y(t2)]	-2.025	1.253	-2.048	1.281	E[y(t3)]-E[y(t2)]	0.000	0.233	0.000	0.250
	E[y(t2)]-E[y(t1)]	-2.586	1.772	-2.607	1.799	E[y(t2)]-E[y(t1)]	0.000	0.267	0.000	0.289
	E[y(t3)]-E[y(t1)]	-2.795	1.209	-2.816	1.241	E[y(t3)]-E[y(t1)]	0.000	0.342	0.000	0.364
	E[y(t4)]-E[y(t1)]	-3.066	0.417	-3.079	0.440	E[y(t4)]-E[y(t1)]	0.000	0.417	0.000	0.440
	Assumption	ETS		95% conf.in.			MTR		95% Conf. in.	
2010				LB	UB		LB	UB	LB	UB
	E[y(t1)]	2.214		2.204	2.224	E[y(t1)]	1.162	2.497	1.159	2.501
	E[y(t2)]	2.346		2.340	2.353	E[y(t2)]	1.545	2.868	1.540	2.875
	E[y(t3)]	2.534		2.527	2.541	E[y(t3)]	1.926	3.623	1.919	3.630
	E[y(t4)]	2.712		2.704	2.720	E[y(t4)]	2.497	4.237	2.493	4.246
	E[y(t4)]-E[y(t3)]	0.178		0.163	0.193	E[y(t4)]-E[y(t3)]	0.000	2.310	0.000	2.327
	E[y(t3)]-E[y(t2)]	0.187		0.174	0.201	E[y(t3)]-E[y(t2)]	0.000	2.078	0.000	2.090
	E[y(t2)]-E[y(t1)]	0.132		0.116	0.149	E[y(t2)]-E[y(t1)]	0.000	1.707	0.000	1.716
	E[y(t3)]-E[y(t1)]	0.320		0.303	0.337	E[y(t3)]-E[y(t1)]	0.000	2.461	0.000	2.471
	E[y(t4)]-E[y(t1)]	0.498		0.480	0.516	E[y(t4)]-E[y(t1)]	0.000	3.075	0.000	3.087
	Assumption	MTS		95% Confi.in.			MTR+MTS		95% conf.in.	
		LB	UB	LB	UB		LB	UB	LB	UB
	E[y(t1)]	2.214	4.629	2.201	4.638	E[y(t1)]	2.214	2.497	2.206	2.503
	E[y(t2)]	2.167	3.892	2.161	3.902	E[y(t2)]	2.329	2.515	2.322	2.519
	E[y(t3)]	1.893	3.356	1.886	3.368	E[y(t3)]	2.438	2.593	2.432	2.602
	E[y(t4)]	1.571	2.712	1.564	2.719	E[y(t4)]	2.497	2.712	2.493	2.721
	E[y(t4)]-E[y(t3)]	-1.785	0.818	-1.803	0.833	E[y(t4)]-E[y(t3)]	0.000	0.274	0.000	0.289
	E[y(t3)]-E[y(t2)]	-1.998	1.189	-2.016	1.206	E[y(t3)]-E[y(t2)]	0.000	0.264	0.000	0.280
	E[y(t2)]-E[y(t1)]	-2.462	1.678	-2.477	1.700	E[y(t2)]-E[y(t1)]	0.000	0.301	0.000	0.313
	E[y(t3)]-E[y(t1)]	-2.736	1.142	-2.752	1.166	E[y(t3)]-E[y(t1)]	0.000	0.379	0.000	0.395
	E[y(t4)]-E[y(t1)]	-3.058	0.498	-3.074	0.517	E[y(t4)]-E[y(t1)]	0.000	0.498	0.000	0.515

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included. (99 repetitions)

Table 1.4: Bounds of ln(hourly wage) and average treatment effect under MIV with MTR & MTS assumptions

Year	Assumption	MTR+MTS+MIV		95% Conf. in.	
2000		LB	UB	LB	UB
	E[y(t1)]	2.006	2.245	1.990	2.250
	E[y(t2)]	2.113	2.265	2.105	2.270
	E[y(t3)]	2.203	2.326	2.197	2.341
	E[y(t4)]	2.245	2.405	2.240	2.418
	E[y(t4)]-E[y(t3)]	0.000	0.201	0.000	0.221
	E[y(t3)]-E[y(t2)]	0.000	0.213	0.000	0.236
	E[y(t2)]-E[y(t1)]	0.000	0.260	0.000	0.280
	E[y(t3)]-E[y(t1)]	0.000	0.321	0.000	0.351
	E[y(t4)]-E[y(t1)]	0.000	0.399	0.000	0.428
	Assumption	MTR+MTS+MIV		95% Conf. in.	
2010		LB	UB	LB	UB
	E[y(t1)]	2.230	2.497	2.218	2.501
	E[y(t2)]	2.333	2.515	2.322	2.520
	E[y(t3)]	2.429	2.581	2.423	2.589
	E[y(t4)]	2.497	2.679	2.493	2.692
	E[y(t4)]-E[y(t3)]	0.000	0.251	0.000	0.269
	E[y(t3)]-E[y(t2)]	0.000	0.248	0.000	0.268
	E[y(t2)]-E[y(t1)]	0.000	0.285	0.000	0.303
	E[y(t3)]-E[y(t1)]	0.000	0.351	0.000	0.372
	E[y(t4)]-E[y(t1)]	0.000	0.449	0.000	0.475

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included. (99 repetitions)

Table 1.5: Bounds of ln(hourly wage) and average treatment effect by different education levels

2000	Less than a high school degree	ETS	MTR+MTS		MTR+MTS+MIV		2010	Less than a high school degree	ETS	MTR+MTS		MTR+MTS+MIV	
			LB	UB	LB	UB				LB	UB	LB	UB
	E[y(t4)]-E[y(t3)]	-0.083	0.000	0.023	0.000	-0.046		E[y(t4)]-E[y(t3)]	-0.048	0.000	0.075	0.000	-0.011
	E[y(t3)]-E[y(t2)]	0.134	0.000	0.146	0.000	-0.008		E[y(t3)]-E[y(t2)]	0.149	0.000	0.172	0.000	0.028
	E[y(t2)]-E[y(t1)]	0.122	0.000	0.161	0.000	0.074		E[y(t2)]-E[y(t1)]	0.131	0.000	0.179	0.000	0.136
	E[y(t4)]-E[y(t1)]	0.173	0.000	0.173	0.000	0.100		E[y(t4)]-E[y(t1)]	0.231	0.000	0.231	0.000	0.134
	Just a high school degree	ETS	MTR+MTS		MTR+MTS+MIV			Just a high school degree	ETS	MTR+MTS		MTR+MTS+MIV	
			LB	UB	LB	UB				LB	UB	LB	UB
	E[y(t4)]-E[y(t3)]	0.045	0.000	0.102	0.000	0.095		E[y(t4)]-E[y(t3)]	0.059	0.000	0.114	0.000	0.056
	E[y(t3)]-E[y(t2)]	0.150	0.000	0.177	0.000	0.163		E[y(t3)]-E[y(t2)]	0.126	0.000	0.156	0.000	0.114
	E[y(t2)]-E[y(t1)]	0.116	0.000	0.236	0.000	0.220		E[y(t2)]-E[y(t1)]	0.109	0.000	0.210	0.000	0.195
	E[y(t4)]-E[y(t1)]	0.311	0.000	0.311	0.000	0.283		E[y(t4)]-E[y(t1)]	0.294	0.000	0.294	0.004	0.224
	College	ETS	MTR+MTS		MTR+MTS+MIV			College	ETS	MTR+MTS		MTR+MTS+MIV	
			LB	UB	LB	UB				LB	UB	LB	UB
	E[y(t4)]-E[y(t3)]	0.201	0.000	0.235	0.000	0.220		E[y(t4)]-E[y(t3)]	0.215	0.000	0.253	0.000	0.213
	E[y(t3)]-E[y(t2)]	0.201	0.000	0.326	0.000	0.264		E[y(t3)]-E[y(t2)]	0.247	0.000	0.382	0.000	0.336
	E[y(t2)]-E[y(t1)]	0.083	0.000	0.375	0.000	0.335		E[y(t2)]-E[y(t1)]	0.071	0.000	0.415	0.000	0.384
	E[y(t4)]-E[y(t1)]	0.485	0.000	0.485	0.006	0.425		E[y(t4)]-E[y(t1)]	0.533	0.000	0.533	0.005	0.457

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included. (99 repetitions)

Table 1.6: Bounds of ln(hourly wage) and average treatment effect by different age groups

Year	16-45	ETS	MTR+MTS		MTR+MTS+MIV	
2000			LB	UB	LB	UB
	E[y(t3)]-E[y(t1)]	0.287	0.000	0.324	0.000	0.307
	E[y(t4)]-E[y(t1)]	0.394	0.000	0.394	0.000	0.370
	45-64	ETS	MTR+MTS		MTR+MTS+MIV	
			LB	UB	LB	UB
	E[y(t3)]-E[y(t1)]	0.374	0.000	0.425	0.000	0.365
	E[y(t4)]-E[y(t1)]	0.548	0.000	0.548	0.007	0.439
2010	16-45	ETS	MTR+MTS		MTR+MTS+MIV	
			LB	UB	LB	UB
	E[y(t3)]-E[y(t1)]	0.288	0.000	0.342	0.000	0.309
	E[y(t4)]-E[y(t1)]	0.443	0.000	0.443	0.000	0.374
	45-64	ETS	MTR+MTS		MTR+MTS+MIV	
			LB	UB	LB	UB
	E[y(t3)]-E[y(t1)]	0.385	0.000	0.467	0.000	0.459
	E[y(t4)]-E[y(t1)]	0.664	0.000	0.664	0.000	0.605

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included. (99 repetitions)

Table 1.7: Bounds of ln(hourly wage) and average treatment effect by gender

Year		ETS	MTR+MTS		MTR+MTS+MIV	
2000	Male		LB	UB	LB	UB
	E[y(t3)]-E[y(t1)]	0.330	0.000	0.360	0.000	0.327
	E[y(t4)]-E[y(t1)]	0.429	0.000	0.429	0.000	0.398
	Female	ETS	MTR+MTS		MTR+MTS+MIV	
			LB	UB	LB	UB
	E[y(t3)]-E[y(t1)]	0.251	0.000	0.316	0.000	0.308
	E[y(t4)]-E[y(t1)]	0.417	0.000	0.417	0.000	0.397
2010	Male	ETS	MTR+MTS		MTR+MTS+MIV	
			LB	UB	LB	UB
	E[y(t3)]-E[y(t1)]	0.321	0.000	0.373	0.000	0.341
	E[y(t4)]-E[y(t1)]	0.491	0.000	0.491	0.000	0.426
	Femlae	ETS	MTR+MTS		MTR+MTS+MIV	
			LB	UB	LB	UB
	E[y(t3)]-E[y(t1)]	0.302	0.000	0.387	0.000	0.348
	E[y(t4)]-E[y(t1)]	0.526	0.000	0.526	0.000	0.482

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included. (99 repetitions)

Table 1.8: Bounds of ln(hourly wage) and average treatment effect by Metropolitan

Year	Metropolitan	ETS	MTR+MTS		MTR+MTS+MIV	
2000			LB	UB	LB	UB
	E[y(t3)]-E[y(t1)]	0.336	0.000	0.372	0.000	0.342
	E[y(t4)]-E[y(t1)]	0.441	0.000	0.441	0.000	0.419
	Non-Metropolitan	ETS	MTR+MTS		MTR+MTS+MIV	
			LB	UB	LB	UB
	E[y(t3)]-E[y(t1)]	0.240	0.000	0.282	0.000	0.275
	E[y(t4)]-E[y(t1)]	0.369	0.000	0.369	0.000	0.360
Year	Metropolitan	ETS	MTR+MTS		MTR+MTS+MIV	
2010			LB	UB	LB	UB
	E[y(t3)]-E[y(t1)]	0.323	0.000	0.387	0.000	0.360
	E[y(t4)]-E[y(t1)]	0.514	0.000	0.514	0.000	0.472
	Non-Metropolitan	ETS	MTR+MTS		MTR+MTS+MIV	
			LB	UB	LB	UB
	E[y(t3)]-E[y(t1)]	0.305	0.000	0.349	0.000	0.317
	E[y(t4)]-E[y(t1)]	0.440	0.000	0.440	0.000	0.372

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included. (99 repetitions)

Table 1.9: Empirical quantiles of ln(hourly wage) and distribution of t

	Year				2000					
T/τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Mean
T1	1.352	1.590	1.730	1.826	1.949	2.060	2.199	2.389	2.712	1.998
T2	1.466	1.685	1.833	1.966	2.078	2.216	2.360	2.537	2.842	2.127
T3	1.590	1.822	1.995	2.144	2.283	2.423	2.570	2.753	3.040	2.303
T4	1.609	1.875	2.060	2.238	2.401	2.557	2.730	2.943	3.257	2.415

	Year				2010					
T/τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Mean
T1	1.590	1.836	1.984	2.060	2.177	2.283	2.423	2.575	2.842	2.214
T2	1.734	1.942	2.071	2.199	2.303	2.436	2.582	2.753	3.001	2.346
T3	1.813	2.060	2.242	2.378	2.506	2.660	2.813	2.976	3.264	2.534
T4	1.854	2.165	2.363	2.544	2.692	2.871	3.048	3.264	3.578	2.712

Note: This table shows the conditional quantile for each treatment group t at $\tau's$ percentile in the year 2000 and 2010.

Table 1.10: Bounds on 0.1 quantile of ln(hourly wage) and quantile treatment effect in 2000

Assumption	Conditional Quantile			EE		95% Conf in.	
				LB	UB	LB	UB
Q[y(t1)]	1.352			1.000	2.140	1.000	2.154
Q[y(t2)]	1.466			1.000	1.926	1.000	1.926
Q[y(t3)]	1.590			1.000	2.177	1.000	2.177
Q[y(t4)]	1.609			1.000	2.060	1.000	2.066
Q[y(t4)]-Q[y(t3)]	0.020			-1.177	1.060	-1.177	1.066
Q[y(t3)]-Q[y(t2)]	0.123			-0.926	1.177	-0.926	1.177
Q[y(t2)]-Q[y(t1)]	0.115			-1.140	0.926	-1.154	0.926
Q[y(t3)]-Q[y(t1)]	0.238			-1.140	1.177	-1.154	1.177
Q[y(t4)]-Q[y(t1)]	0.258			-1.140	1.060	-1.154	1.066
Assumption	**MTRQMTS**		**95% Conf in.**	**MTRQMTSMIV**		**95% Conf in.**	
	LB	UB	LB	UB	LB	UB	LB	UB

Assumption	MTRQMTS LB	MTRQMTS UB	95% Conf in. LB	95% Conf in. UB	MTRQMTSMIV LB	MTRQMTSMIV UB	95% Conf in. LB	95% Conf in. UB
Q[y(t1)]	1.351	1.526	1.322	1.528	1.351	1.526	1.283	1.528
Q[y(t2)]	1.447	1.538	1.427	1.538	1.450	1.549	1.427	1.549
Q[y(t3)]	1.504	1.590	1.500	1.593	1.521	1.590	1.500	1.590
Q[y(t4)]	1.510	1.629	1.506	1.629	1.526	1.625	1.506	1.631
Q[y(t4)]-Q[y(t3)]	0.000	0.125	0.000	0.129	0.000	0.105	0 .000	0.131
Q[y(t3)]-Q[y(t2)]	0.000	0.143	0.000	0.166	0.000	0.139	0.000	0.163
Q[y(t2)]-Q[y(t1)]	0.000	0.188	0.000	0.227	0.000	0.198	0.000	0.266
Q[y(t3)]-Q[y(t1)]	0.000	0.239	0.000	0.271	0.000	0.239	0.000	0.306
Q[y(t4)]-Q[y(t1)]	0.000	0.278	0.000	0.307	0.000	0.274	0.000	0.348

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.11: Bounds on 0.2 quantile of ln(hourly wage) and quantile treatment effect in 2000

Assumption	Conditional Quantile	EE		95% Conf in.	
		LB	UB	LB	UB
Q[y(t1)]	1.590	1.000	5.000	1.000	5.000
Q[y(t2)]	1.685	1.000	2.405	1.000	2.423
Q[y(t3)]	1.822	1.000	2.871	1.000	2.889
Q[y(t4)]	1.875	1.000	2.549	1.000	2.570
Q[y(t4)]-Q[y(t3)]	0.053	-1.871	1.549	-1.889	1.570
Q[y(t3)]-Q[y(t2)]	0.137	-1.405	1.871	-1.423	1.889
Q[y(t2)]-Q[y(t1)]	0.095	-4.000	1.405	-4.000	1.423
Q[y(t3)]-Q[y(t1)]	0.233	-4.000	1.871	-4.000	1.889
Q[y(t4)]-Q[y(t1)]	0.285	-4.000	1.549	-4.000	1.570

Assumption	MTRQMTS		95% Conf in.		MTRQMTSMIV		95% Conf in.	
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	1.590	1.758	1.590	1.772	1.590	1.758	1.590	1.772
Q[y(t2)]	1.671	1.772	1.659	1.772	1.678	1.772	1.662	1.772
Q[y(t3)]	1.749	1.835	1.743	1.852	1.749	1.835	1.707	1.852
Q[y(t4)]	1.764	1.877	1.753	1.880	1.772	1.865	1.749	1.877
Q[y(t4)]-Q[y(t3)]	0.000	0.129	0.000	0.137	0.000	0.116	0.000	0.170
Q[y(t3)]-Q[y(t2)]	0.000	0.165	0.000	0.193	0.000	0.157	0.000	0.190
Q[y(t2)]-Q[y(t1)]	0.000	0.182	0.000	0.182	0.000	0.182	0.000	0.182
Q[y(t3)]-Q[y(t1)]	0.000	0.246	0.000	0.262	0.000	0.245	0.000	0.262
Q[y(t4)]-Q[y(t1)]	0.000	0.288	0.000	0.290	0.014	0.275	0.000	0.288

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.12: Bounds on 0.3 quantile of ln(hourly wage) and quantile treatment effect in 2000

Assumption	Conditional Quantile	EE		95% Conf in.	
		LB	UB	LB	UB
Q[y(t1)]	1.730	1.000	5.000	1.000	5.000
Q[y(t2)]	1.833	1.000	5.000	1.000	5.000
Q[y(t3)]	1.995	1.000	5.000	1.000	5.000
Q[y(t4)]	2.060	1.000	3.227	1.000	3.245
Q[y(t4)]-Q[y(t3)]	0.065	-4.000	2.227	-4.000	2.245
Q[y(t3)]-Q[y(t2)]	0.163	-4.000	4.000	-4.000	4.000
Q[y(t2)]-Q[y(t1)]	0.102	-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t1)]	0.265	-4.000	4.000	-4.000	4.000
Q[y(t4)]-Q[y(t1)]	0.330	-4.000	2.227	-4.000	2.245

Assumption	MTRQMTS		95% Conf in.		MTRQMTSMIV		95% Conf in.	
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	1.734	1.910	1.723	1.926	1.735	1.910	1.723	1.926
Q[y(t2)]	1.814	1.926	1.812	1.936	1.814	1.926	1.813	1.936
Q[y(t3)]	1.897	2.011	1.888	2.021	1.897	2.011	1.888	2.021
Q[y(t4)]	1.917	2.060	1.903	2.066	1.922	2.060	1.903	2.060
Q[y(t4)]-Q[y(t3)]	0.000	0.163	0.000	0.178	0.000	0.163	0.000	0.172
Q[y(t3)]-Q[y(t2)]	0.000	0.197	0.000	0.209	0.000	0.197	0.000	0.208
Q[y(t2)]-Q[y(t1)]	0.000	0.192	0.000	0.212	0.000	0.191	0.000	0.213
Q[y(t3)]-Q[y(t1)]	0.000	0.277	0.000	0.298	0.000	0.276	0.000	0.298
Q[y(t4)]-Q[y(t1)]	0.000	0.326	0.000	0.343	0.010	0.324	0.000	0.337

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.13: Bounds on 0.4 quantile of ln(hourly wage) and quantile treatment effect in 2000

Assumption	Conditional Quantile				EE		95% Conf in.	
					LB	UB	LB	UB
Q[y(t1)]	1.826				1.000	5.000	1.000	5.000
Q[y(t2)]	1.966				1.000	5.000	1.000	5.000
Q[y(t3)]	2.144				1.000	5.000	1.000	5.000
Q[y(t4)]	2.238				1.000	5.000	1.000	5.000
Q[y(t4)]-Q[y(t3)]	0.095				-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t2)]	0.177				-4.000	4.000	-4.000	4.000
Q[y(t2)]-Q[y(t1)]	0.140				-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t1)]	0.317				-4.000	4.000	-4.000	4.000
Q[y(t4)]-Q[y(t1)]	0.412				-4.000	4.000	-4.000	4.000
Assumption	MTRQMTS		95% Conf in.		MTRQMTSMIV		95% Conf in.	
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	1.833	2.060	1.822	2.060	1.833	2.060	1.822	2.060
Q[y(t2)]	1.944	2.065	1.931	2.085	1.944	2.065	1.944	2.079
Q[y(t3)]	2.036	2.177	2.026	2.177	2.036	2.177	2.036	2.177
Q[y(t4)]	2.060	2.238	2.056	2.247	2.060	2.228	2.060	2.231
Q[y(t4)]-Q[y(t3)]	0.000	0.202	0.000	0.221	0.000	0.192	0.000	0.195
Q[y(t3)]-Q[y(t2)]	0.000	0.234	0.000	0.246	0.000	0.233	0.000	0.233
Q[y(t2)]-Q[y(t1)]	0.000	0.232	0.000	0.262	0.000	0.232	0.000	0.257
Q[y(t3)]-Q[y(t1)]	0.000	0.345	0.000	0.355	0.000	0.345	0.000	0.355
Q[y(t4)]-Q[y(t1)]	0.000	0.405	0.000	0.425	0.000	0.395	0.000	0.409

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.14: Bounds on 0.5 quantile of ln(hourly wage) and quantile treatment effect in 2000

Assumption	Conditional Quantile				EE		95% Conf in.	
					LB	UB	LB	UB
Q[y(t1)]		1.949			1.000	5.000	1.000	5.000
Q[y(t2)]		2.078			1.000	5.000	1.000	5.000
Q[y(t3)]		2.283			1.000	5.000	1.000	5.000
Q[y(t4)]		2.401			1.000	5.000	1.000	5.000
Q[y(t4)]-Q[y(t3)]		0.118			-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t2)]		0.205			-4.000	4.000	-4.000	4.000
Q[y(t2)]-Q[y(t1)]		0.130			-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t1)]		0.334			-4.000	4.000	-4.000	4.000
Q[y(t4)]-Q[y(t1)]		0.452			-4.000	4.000	-4.000	4.000
Assumption	**MTRQMTS**		**95% Conf in.**		**MTRQMTSMIV**		**95% Conf in.**	
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	1.954	2.193	1.933	2.206	1.961	2.193	1.933	2.206
Q[y(t2)]	2.060	2.228	2.060	2.231	2.060	2.228	2.060	2.231
Q[y(t3)]	2.177	2.305	2.171	2.324	2.177	2.305	2.171	2.324
Q[y(t4)]	2.196	2.396	2.186	2.404	2.197	2.378	2.186	2.388
Q[y(t4)]-Q[y(t3)]	0.000	0.219	0.000	0.234	0.000	0.201	0.000	0.217
Q[y(t3)]-Q[y(t2)]	0.000	0.245	0.000	0.264	0.000	0.245	0.000	0.264
Q[y(t2)]-Q[y(t1)]	0.000	0.273	0.000	0.299	0.000	0.267	0.000	0.298
Q[y(t3)]-Q[y(t1)]	0.000	0.350	0.000	0.391	0.000	0.344	0.000	0.391
Q[y(t4)]-Q[y(t1)]	0.000	0.442	0.000	0.472	0.004	0.417	0.000	0.455

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.15: Bounds on 0.6 quantile of ln(hourly wage) and quantile treatment effect in 2000)

Assumption	Conditional Quantile	EE		95% Conf in.	
		LB	UB	LB	UB
Q[y(t1)]	2.060	1.000	5.000	1.000	5.000
Q[y(t2)]	2.216	1.000	5.000	1.000	5.000
Q[y(t3)]	2.423	1.000	5.000	1.000	5.000
Q[y(t4)]	2.557	1.000	5.000	1.000	5.000
Q[y(t4)]-Q[y(t3)]	0.134	-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t2)]	0.207	-4.000	4.000	-4.000	4.000
Q[y(t2)]-Q[y(t1)]	0.156	-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t1)]	0.363	-4.000	4.000	-4.000	4.000
Q[y(t4)]-Q[y(t1)]	0.497	-4.000	4.000	-4.000	4.000

Assumption	MTRQMTS		95% Conf in.		MTRQMTSMIV		95% Conf in.	
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	2.060	2.347	1.933	2.360	2.060	2.347	2.045	2.360
Q[y(t2)]	2.182	2.378	2.060	2.378	2.182	2.378	2.060	2.378
Q[y(t3)]	2.317	2.465	2.171	2.467	2.317	2.465	2.303	2.467
Q[y(t4)]	2.347	2.556	2.186	2.568	2.347	2.545	2.334	2.545
Q[y(t4)]-Q[y(t3)]	0.000	0.239	0.000	0.398	0.000	0.228	0.000	0.243
Q[y(t3)]-Q[y(t2)]	0.000	0.283	0.000	0.407	0.000	0.283	0.000	0.407
Q[y(t2)]-Q[y(t1)]	0.000	0.318	0.000	0.446	0.000	0.318	0.000	0.333
Q[y(t3)]-Q[y(t1)]	0.000	0.405	0.000	0.534	0.000	0.405	0.000	0.422
Q[y(t4)]-Q[y(t1)]	0.000	0.497	0.000	0.636	0.000	0.486	0.000	0.500

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.16: Bounds on 0.7 quantile of ln(hourly wage) and quantile treatment effect in 2000

Assumption	Conditional Quantile			EE		95% Conf in.	
				LB	UB	LB	UB
Q[y(t1)]	2.199			1.000	5.000	1.000	5.000
Q[y(t2)]	2.360			1.000	5.000	1.000	5.000
Q[y(t3)]	2.570			1.000	5.000	1.000	5.000
Q[y(t4)]	2.730			1.000	5.000	1.638	5.000
Q[y(t4)]-Q[y(t3)]	0.159			-4.000	4.000	-3.362	4.000
Q[y(t3)]-Q[y(t2)]	0.211			-4.000	4.000	-4.000	4.000
Q[y(t2)]-Q[y(t1)]	0.160			-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t1)]	0.371			-4.000	4.000	-4.000	4.000
Q[y(t4)]-Q[y(t1)]	0.530			-4.000	4.000	-3.362	4.000

Assumption	MTRQMTS		95% Conf in.		MTRQMTSMIV		95% Conf in.	
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	2.213	2.515	2.189	2.526	2.213	2.515	2.189	2.526
Q[y(t2)]	2.333	2.537	2.332	2.545	2.333	2.537	2.332	2.545
Q[y(t3)]	2.473	2.619	2.465	2.631	2.473	2.619	2.470	2.631
Q[y(t4)]	2.515	2.725	2.506	2.739	2.519	2.708	2.506	2.721
Q[y(t4)]-Q[y(t3)]	0.000	0.252	0.000	0.273	0.000	0.235	0.000	0.251
Q[y(t3)]-Q[y(t2)]	0.000	0.286	0.000	0.300	0.000	0.286	0.000	0.299
Q[y(t2)]-Q[y(t1)]	0.000	0.324	0.000	0.356	0.000	0.324	0.000	0.356
Q[y(t3)]-Q[y(t1)]	0.000	0.406	0.000	0.442	0.000	0.406	0.000	0.442
Q[y(t4)]-Q[y(t1)]	0.000	0.512	0.000	0.549	0.000	0.495	0.000	0.532

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.17: Bounds on 0.8 quantile of ln(hourly wage) and quantile treatment effect in 2000

Assumption	Conditional Quantile				EE		95% Conf in.	
					LB	UB	LB	UB
Q[y(t1)]	2.389				1.000	5.000	1.000	5.000
Q[y(t2)]	2.537				1.788	5.000	1.772	5.000
Q[y(t3)]	2.753				1.744	5.000	1.723	5.000
Q[y(t4)]	2.943				2.244	5.000	2.231	5.000
Q[y(t4)]-Q[y(t3)]	0.190				-2.756	3.256	-2.769	3.277
Q[y(t3)]-Q[y(t2)]	0.216				-3.256	3.212	-3.277	3.228
Q[y(t2)]-Q[y(t1)]	0.148				-3.212	4.000	-3.228	4.000
Q[y(t3)]-Q[y(t1)]	0.364				-3.256	4.000	-3.277	4.000
Q[y(t4)]-Q[y(t1)]	0.554				-2.756	4.000	-2.769	4.000
Assumption	MTRQMTS		95% Conf in.		MTRQMTSMIV		95% Conf in.	
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	2.391	2.730	2.378	2.740	2.401	2.730	2.378	2.740
Q[y(t2)]	2.523	2.753	2.506	2.753	2.526	2.753	2.534	2.753
Q[y(t3)]	2.670	2.813	2.659	2.842	2.670	2.813	2.685	2.842
Q[y(t4)]	2.734	2.943	2.721	2.953	2.727	2.925	2.721	2.938
Q[y(t4)]-Q[y(t3)]	0.000	0.273	0.000	0.293	0.000	0.255	0.000	0.253
Q[y(t3)]-Q[y(t2)]	0.000	0.291	0.000	0.336	0.000	0.287	0.000	0.308
Q[y(t2)]-Q[y(t1)]	0.000	0.361	0.000	0.375	0.000	0.352	0.000	0.375
Q[y(t3)]-Q[y(t1)]	0.000	0.422	0.000	0.464	0.000	0.412	0.000	0.464
Q[y(t4)]-Q[y(t1)]	0.000	0.551	0.000	0.575	0.000	0.523	0.000	0.560

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.18: Bounds on 0.9 quantile of ln(hourly wage) and quantile treatment effect in 2000

Assumption	Conditional Quantile				EE		95% Conf in.	
					LB	UB	LB	UB
Q[y(t1)]	2.712				1.772	5.000	1.000	5.000
Q[y(t2)]	2.842				2.277	5.000	1.772	5.000
Q[y(t3)]	3.040				2.382	5.000	1.723	5.000
Q[y(t4)]	3.257				2.738	5.000	2.231	5.000
Q[y(t4)]-Q[y(t3)]	0.217				-2.262	2.618	-2.769	3.277
Q[y(t3)]-Q[y(t2)]	0.198				-2.618	2.723	-3.277	3.228
Q[y(t2)]-Q[y(t1)]	0.130				-2.723	3.228	-3.228	4.000
Q[y(t3)]-Q[y(t1)]	0.328				-2.618	3.228	-3.277	4.000
Q[y(t4)]-Q[y(t1)]	0.545				-2.262	3.228	-2.769	4.000
Assumption	MTRQMTS		95% Conf in.		MTRQMTSMIV		95% Conf in.	
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	2.715	3.065	2.708	3.071	2.727	3.065	2.727	3.071
Q[y(t2)]	2.842	3.071	2.816	3.075	2.842	3.071	2.848	3.075
Q[y(t3)]	2.975	3.097	2.951	3.127	2.975	3.097	2.976	3.127
Q[y(t4)]	3.048	3.257	3.048	3.273	3.056	3.257	3.046	3.264
Q[y(t4)]-Q[y(t3)]	0.000	0.282	0.000	0.323	0.000	0.282	0.000	0.288
Q[y(t3)]-Q[y(t2)]	0.000	0.255	0.000	0.310	0.000	0.255	0.000	0.279
Q[y(t2)]-Q[y(t1)]	0.000	0.357	0.000	0.367	0.000	0.344	0.000	0.348
Q[y(t3)]-Q[y(t1)]	0.000	0.382	0.000	0.418	0.000	0.370	0.000	0.400
Q[y(t4)]-Q[y(t1)]	0.000	0.543	0.000	0.565	0.000	0.530	0.000	0.537

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.19: Bounds on 0.1 quantile of ln(hourly wage) and quantile treatment effect in 2010

Assumption	Conditional Quantile	EE		95% Conf in.	
		LB	UB	LB	UB
Q[y(t1)]	1.590	1.000	2.506	1.000	2.506
Q[y(t2)]	1.734	1.000	2.143	1.000	2.155
Q[y(t3)]	1.813	1.000	2.378	1.000	2.388
Q[y(t4)]	1.854	1.000	2.362	1.000	2.378
Q[y(t4)]-Q[y(t3)]	0.041	-1.378	1.362	-1.388	1.378
Q[y(t3)]-Q[y(t2)]	0.079	-1.143	1.378	-1.155	1.388
Q[y(t2)]-Q[y(t1)]	0.145	-1.506	1.143	-1.506	1.155
Q[y(t3)]-Q[y(t1)]	0.223	-1.506	1.378	-1.506	1.388
Q[y(t4)]-Q[y(t1)]	0.264	-1.506	1.362	-1.506	1.378

Assumption	MTRQMTS		95% Conf in.		MTRQMTSMIV		95% Conf in.	
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	1.590	1.772	1.590	1.772	1.618	1.772	1.630	1.772
Q[y(t2)]	1.723	1.772	1.696	1.772	1.703	1.772	1.685	1.772
Q[y(t3)]	1.772	1.836	1.772	1.837	1.772	1.772	1.726	1.772
Q[y(t4)]	1.772	1.852	1.772	1.877	1.772	1.772	1.772	1.772
Q[y(t4)]-Q[y(t3)]	0.000	0.080	0.000	0.105	0.000	0.000	0.000	0.046
Q[y(t3)]-Q[y(t2)]	0.000	0.114	0.000	0.140	0.000	0.069	0.000	0.087
Q[y(t2)]-Q[y(t1)]	0.000	0.182	0.000	0.182	0.000	0.154	0.000	0.142
Q[y(t3)]-Q[y(t1)]	0.000	0.247	0.000	0.247	0.000	0.154	0.000	0.142
Q[y(t4)]-Q[y(t1)]	0.000	0.262	0.000	0.288	0.000	0.154	0.000	0.142

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.20: Bounds on 0.2 quantile of ln(hourly wage) and quantile treatment effect in 2010

Assumption	Conditional Quantile	EE		95% Conf in.	
		LB	UB	LB	UB
Q[y(t1)]	1.836	1.000	5.000	1.000	5.000
Q[y(t2)]	1.942	1.000	2.583	1.000	2.590
Q[y(t3)]	2.060	1.000	2.996	1.000	3.015
Q[y(t4)]	2.165	1.000	2.871	1.000	2.881
Q[y(t4)]-Q[y(t3)]	0.105	-1.996	1.871	-2.015	1.881
Q[y(t3)]-Q[y(t2)]	0.118	-1.583	1.996	-1.590	2.015
Q[y(t2)]-Q[y(t1)]	0.105	-4.000	1.583	-4.000	1.590
Q[y(t3)]-Q[y(t1)]	0.223	-4.000	1.996	-4.000	2.015
Q[y(t4)]-Q[y(t1)]	0.329	-4.000	1.871	-4.000	1.881

Assumption	MTRQMTS		95% Conf in.		MTRQMTSMIV		95% Conf in.	
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	1.833	1.995	1.833	1.995	1.852	1.995	1.852	1.995
Q[y(t2)]	1.926	2.017	1.926	2.023	1.926	2.016	1.926	2.023
Q[y(t3)]	1.995	2.090	1.995	2.097	1.995	2.060	1.963	2.066
Q[y(t4)]	1.995	2.165	1.995	2.174	1.995	2.108	1.995	2.120
Q[y(t4)]-Q[y(t3)]	0.000	0.170	0.000	0.179	0.000	0.113	0.000	0.157
Q[y(t3)]-Q[y(t2)]	0.000	0.164	0.000	0.171	0.000	0.134	0.000	0.140
Q[y(t2)]-Q[y(t1)]	0.000	0.184	0.000	0.191	0.000	0.164	0.000	0.171
Q[y(t3)]-Q[y(t1)]	0.000	0.258	0.000	0.265	0.000	0.208	0.000	0.214
Q[y(t4)]-Q[y(t1)]	0.000	0.332	0.000	0.342	0.000	0.256	0.000	0.268

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.21: Bounds on 0.3 quantile of ln(hourly wage) and quantile treatment effect in 2010

Assumption	Conditional Quantile	EE		95% Conf in.	
		LB	UB	LB	UB
Q[y(t1)]	1.984	1.000	5.000	1.000	5.000
Q[y(t2)]	2.071	1.000	5.000	1.000	5.000
Q[y(t3)]	2.242	1.000	5.000	1.000	5.000
Q[y(t4)]	2.363	1.000	3.576	1.000	3.594
Q[y(t4)]-Q[y(t3)]	0.121	-4.000	2.576	-4.000	2.594
Q[y(t3)]-Q[y(t2)]	0.170	-4.000	4.000	-4.000	4.000
Q[y(t2)]-Q[y(t1)]	0.087	-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t1)]	0.258	-4.000	4.000	-4.000	4.000
Q[y(t4)]-Q[y(t1)]	0.379	-4.000	2.576	-4.000	2.594

Assumption	MTRQMTS		95% Conf in.		MTRQMTSMIV		95% Conf in.	
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	1.984	2.174	1.973	2.177	1.986	2.172	1.983	2.177
Q[y(t2)]	2.060	2.182	2.060	2.188	2.060	2.182	2.060	2.188
Q[y(t3)]	2.143	2.283	2.133	2.283	2.143	2.283	2.133	2.283
Q[y(t4)]	2.177	2.366	2.165	2.378	2.177	2.306	2.165	2.322
Q[y(t4)]-Q[y(t3)]	0.000	0.223	0.000	0.245	0.000	0.163	0.000	0.189
Q[y(t3)]-Q[y(t2)]	0.000	0.223	0.000	0.223	0.000	0.223	0.000	0.223
Q[y(t2)]-Q[y(t1)]	0.000	0.198	0.000	0.216	0.000	0.196	0.000	0.206
Q[y(t3)]-Q[y(t1)]	0.000	0.299	0.000	0.310	0.000	0.297	0.000	0.300
Q[y(t4)]-Q[y(t1)]	0.000	0.382	0.000	0.405	0.000	0.320	0.000	0.339

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.22: Bounds on 0.4 quantile of ln(hourly wage) and quantile treatment effect in 2010

Assumption	Conditional Quantile	EE		95% Conf in.	
		LB	UB	LB	UB
Q[y(t1)]	2.060	1.000	5.000	1.000	5.000
Q[y(t2)]	2.199	1.000	5.000	1.000	5.000
Q[y(t3)]	2.378	1.000	5.000	1.000	5.000
Q[y(t4)]	2.544	1.000	5.000	1.000	5.000
Q[y(t4)]-Q[y(t3)]	0.166	-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t2)]	0.179	-4.000	4.000	-4.000	4.000
Q[y(t2)]-Q[y(t1)]	0.140	-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t1)]	0.318	-4.000	4.000	-4.000	4.000
Q[y(t4)]-Q[y(t1)]	0.484	-4.000	4.000	-4.000	4.000

	MTRQMTS		95% Conf in.		MTRQMTSMIV		95% Conf in.	
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	2.060	2.303	2.060	2.303	2.061	2.303	2.060	2.303
Q[y(t2)]	2.177	2.322	2.177	2.325	2.177	2.322	2.177	2.325
Q[y(t3)]	2.283	2.423	2.283	2.430	2.283	2.423	2.283	2.430
Q[y(t4)]	2.303	2.545	2.283	2.545	2.303	2.506	2.283	2.506
Q[y(t4)]-Q[y(t3)]	0.000	0.262	0.000	0.262	0.000	0.223	0.000	0.223
Q[y(t3)]-Q[y(t2)]	0.000	0.245	0.000	0.253	0.000	0.245	0.000	0.253
Q[y(t2)]-Q[y(t1)]	0.000	0.262	0.000	0.265	0.000	0.261	0.000	0.265
Q[y(t3)]-Q[y(t1)]	0.000	0.363	0.000	0.371	0.000	0.362	0.000	0.370
Q[y(t4)]-Q[y(t1)]	0.000	0.486	0.000	0.486	0.000	0.445	0.000	0.446

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.23: Bounds on 0.5 quantile of ln(hourly wage) and quantile treatment effect in 2010

Assumption	Conditional Quantile	EE		95% Conf in.	
		LB	UB	LB	UB
Q[y(t1)]	2.177	1.000	5.000	1.000	5.000
Q[y(t2)]	2.303	1.000	5.000	1.000	5.000
Q[y(t3)]	2.506	1.000	5.000	1.000	5.000
Q[y(t4)]	2.692	1.000	5.000	1.000	5.000
Q[y(t4)]-Q[y(t3)]	0.186	-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t2)]	0.203	-4.000	4.000	-4.000	4.000
Q[y(t2)]-Q[y(t1)]	0.125	-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t1)]	0.329	-4.000	4.000	-4.000	4.000
Q[y(t4)]-Q[y(t1)]	0.514	-4.000	4.000	-4.000	4.000

Assumption	MTRQMTS		95% Conf in.		MTRQMTSMIV		95% Conf in.	
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	2.177	2.465	2.177	2.465	2.177	2.465	2.188	2.465
Q[y(t2)]	2.283	2.470	2.283	2.483	2.283	2.465	2.283	2.465
Q[y(t3)]	2.416	2.570	2.398	2.573	2.416	2.570	2.412	2.573
Q[y(t4)]	2.465	2.693	2.465	2.708	2.465	2.688	2.465	2.688
Q[y(t4)]-Q[y(t3)]	0.000	0.276	0.000	0.310	0.000	0.272	0.000	0.277
Q[y(t3)]-Q[y(t2)]	0.000	0.288	0.000	0.290	0.000	0.287	0.000	0.290
Q[y(t2)]-Q[y(t1)]	0.000	0.292	0.000	0.306	0.000	0.288	0.000	0.277
Q[y(t3)]-Q[y(t1)]	0.000	0.393	0.000	0.396	0.000	0.393	0.000	0.385
Q[y(t4)]-Q[y(t1)]	0.000	0.515	0.000	0.531	0.000	0.511	0.000	0.500

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.24: Bounds on 0.6 quantile of ln(hourly wage) and quantile treatment effect in 2010

Assumption	Conditional Quantile	EE		95% Conf in.	
		LB	UB	LB	UB
Q[y(t1)]	2.283	1.000	5.000	1.000	5.000
Q[y(t2)]	2.436	1.000	5.000	1.000	5.000
Q[y(t3)]	2.660	1.000	5.000	1.000	5.000
Q[y(t4)]	2.871	1.000	5.000	1.000	5.000
Q[y(t4)]-Q[y(t3)]	0.211	-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t2)]	0.224	-4.000	4.000	-4.000	4.000
Q[y(t2)]-Q[y(t1)]	0.153	-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t1)]	0.377	-4.000	4.000	-4.000	4.000
Q[y(t4)]-Q[y(t1)]	0.588	-4.000	4.000	-4.000	4.000

Assumption	MTRQMTS		95% Conf in.		MTRQMTSMIV		95% Conf in.	
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	2.283	2.619	2.283	2.619	2.283	2.619	2.283	2.619
Q[y(t2)]	2.423	2.631	2.423	2.639	2.423	2.631	2.423	2.639
Q[y(t3)]	2.545	2.718	2.526	2.724	2.545	2.718	2.526	2.718
Q[y(t4)]	2.619	2.871	2.610	2.877	2.619	2.871	2.613	2.871
Q[y(t4)]-Q[y(t3)]	0.000	0.325	0.000	0.352	0.000	0.325	0.000	0.345
Q[y(t3)]-Q[y(t2)]	0.000	0.296	0.000	0.302	0.000	0.295	0.000	0.295
Q[y(t2)]-Q[y(t1)]	0.000	0.348	0.000	0.357	0.000	0.348	0.000	0.356
Q[y(t3)]-Q[y(t1)]	0.000	0.435	0.000	0.441	0.000	0.435	0.000	0.435
Q[y(t4)]-Q[y(t1)]	0.000	0.588	0.000	0.595	0.000	0.588	0.000	0.588

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.25: Bounds on 0.7 quantile of ln(hourly wage) and quantile treatment effect in 2010

Assumption	Conditional Quantile			EE		95% Conf in.	
				LB	UB	LB	UB
Q[y(t1)]	2.423			1.000	5.000	1.000	5.000
Q[y(t2)]	2.582			1.000	5.000	1.000	5.000
Q[y(t3)]	2.813			1.000	5.000	1.000	5.000
Q[y(t4)]	3.048			1.859	5.000	1.852	5.000
Q[y(t4)]-Q[y(t3)]	0.235			-3.141	4.000	-3.148	4.000
Q[y(t3)]-Q[y(t2)]	0.232			-4.000	4.000	-4.000	4.000
Q[y(t2)]-Q[y(t1)]	0.159			-4.000	4.000	-4.000	4.000
Q[y(t3)]-Q[y(t1)]	0.391			-4.000	4.000	-4.000	4.000
Q[y(t4)]-Q[y(t1)]	0.626			-3.141	4.000	-3.148	4.000
Assumption	**MTRQMTS**		**95% Conf in.**		**MTRQMTSMIV**		**95% Conf in.**
	LB	UB	LB	UB	LB	UB	LB	UB

Assumption	MTRQMTS LB	MTRQMTS UB	95% Conf in. LB	95% Conf in. UB	MTRQMTSMIV LB	MTRQMTSMIV UB	95% Conf in. LB	95% Conf in. UB
Q[y(t1)]	2.423	2.783	2.423	2.788	2.430	2.783	2.461	2.788
Q[y(t2)]	2.546	2.802	2.545	2.813	2.556	2.802	2.570	2.813
Q[y(t3)]	2.688	2.898	2.688	2.900	2.688	2.898	2.696	2.900
Q[y(t4)]	2.785	3.048	2.770	3.056	2.785	3.048	2.768	3.056
Q[y(t4)]-Q[y(t3)]	0.000	0.360	0.000	0.368	0.000	0.360	0.000	0.360
Q[y(t3)]-Q[y(t2)]	0.000	0.352	0.000	0.355	0.000	0.342	0.000	0.330
Q[y(t2)]-Q[y(t1)]	0.000	0.379	0.000	0.391	0.000	0.371	0.000	0.352
Q[y(t3)]-Q[y(t1)]	0.000	0.475	0.000	0.478	0.000	0.468	0.000	0.439
Q[y(t4)]-Q[y(t1)]	0.003	0.626	0.000	0.633	0.000	0.618	0.000	0.595

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.26: Bounds on 0.8 quantile of ln(hourly wage) and quantile treatment effect in 2010

Assumption	Conditional Quantile			EE		95% Conf in.	
				LB	UB	LB	UB
Q[y(t1)]	2.575			1.000	5.000	1.000	5.000
Q[y(t2)]	2.753			2.060	5.000	2.060	5.000
Q[y(t3)]	2.976			2.060	5.000	2.060	5.000
Q[y(t4)]	3.264			2.545	5.000	2.526	5.000
Q[y(t4)]-Q[y(t3)]	0.288			-2.455	2.940	-2.474	2.940
Q[y(t3)]-Q[y(t2)]	0.223			-2.940	2.940	-2.940	2.940
Q[y(t2)]-Q[y(t1)]	0.178			-2.940	4.000	-2.940	4.000
Q[y(t3)]-Q[y(t1)]	0.401			-2.940	4.000	-2.940	4.000
Q[y(t4)]-Q[y(t1)]	0.689			-2.455	4.000	-2.474	4.000
Assumption	MTRQMTS		95% Conf in.		MTRQMTSMIV		95% Conf in.
	LB	UB	LB	UB	LB	UB	LB	UB
Q[y(t1)]	2.571	2.976	2.570	2.988	2.590	2.976	2.590	2.980
Q[y(t2)]	2.725	2.996	2.721	3.001	2.740	2.996	2.753	3.001
Q[y(t3)]	2.894	3.094	2.882	3.094	2.894	3.094	2.894	3.094
Q[y(t4)]	2.976	3.264	2.976	3.275	2.976	3.264	2.976	3.275
Q[y(t4)]-Q[y(t3)]	0.000	0.370	0.000	0.393	0.000	0.370	0.000	0.381
Q[y(t3)]-Q[y(t2)]	0.000	0.369	0.000	0.373	0.000	0.354	0.000	0.341
Q[y(t2)]-Q[y(t1)]	0.000	0.425	0.000	0.430	0.000	0.406	0.000	0.411
Q[y(t3)]-Q[y(t1)]	0.000	0.523	0.000	0.523	0.000	0.503	0.000	0.504
Q[y(t4)]-Q[y(t1)]	0.000	0.693	0.000	0.705	0.000	0.673	0.000	0.685

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Table 1.27: Bounds on 0.9 quantile of ln(hourly wage) and quantile treatment effect in 2010

Assumption	Conditional Quantile	EE LB	EE UB	95% Conf in. LB	95% Conf in. UB
Q[y(t1)]	2.842	1.920	5.000	1.905	5.000
Q[y(t2)]	3.001	2.506	5.000	2.501	5.000
Q[y(t3)]	3.264	2.659	5.000	2.654	5.000
Q[y(t4)]	3.578	3.048	5.000	3.025	5.000
Q[y(t4)]-Q[y(t3)]	0.314	-1.952	2.341	-1.975	2.346
Q[y(t3)]-Q[y(t2)]	0.263	-2.341	2.494	-2.346	2.499
Q[y(t2)]-Q[y(t1)]	0.158	-2.494	3.080	-2.499	3.095
Q[y(t3)]-Q[y(t1)]	0.421	-2.341	3.080	-2.346	3.095
Q[y(t4)]-Q[y(t1)]	0.735	-1.952	3.080	-1.975	3.095

Assumption	MTRQMTS LB	MTRQMTS UB	95% Conf in. LB	95% Conf in. UB	MTRQMTSMIV LB	MTRQMTSMIV UB	95% Conf in. LB	95% Conf in. UB
Q[y(t1)]	2.842	3.301	2.842	3.316	2.858	3.301	2.858	3.316
Q[y(t2)]	2.976	3.312	2.976	3.330	2.980	3.312	2.980	3.330
Q[y(t3)]	3.177	3.381	3.158	3.381	3.177	3.381	3.158	3.381
Q[y(t4)]	3.306	3.576	3.294	3.591	3.312	3.576	3.295	3.591
Q[y(t4)]-Q[y(t3)]	0.000	0.399	0.000	0.433	0.000	0.399	0.000	0.433
Q[y(t3)]-Q[y(t2)]	0.000	0.405	0.000	0.405	0.000	0.401	0.000	0.401
Q[y(t2)]-Q[y(t1)]	0.000	0.470	0.000	0.488	0.000	0.454	0.000	0.472
Q[y(t3)]-Q[y(t1)]	0.000	0.539	0.000	0.539	0.000	0.523	0.000	0.523
Q[y(t4)]-Q[y(t1)]	0.005	0.733	0.000	0.749	0.011	0.718	0.000	0.733

Note: MTRQMTS means imposing MTR and Quantile-MTS assumptions jointly. MTRQMTSMIV means imposing MTR, Quantile-MTS and Quantile-MIV assumptions jointly.

Chapter 2 Tables

Table 2.1: Summary statistics for English proficiency and monthly wage

	Listening		Speaking		Monthly wage		Sample Size
	Mean	SD	Mean	SD	Mean	SD	
2012	0.702	0.797	0.624	0.764	2857.937	2935.020	3,783
2013	0.716	0.796	0.612	0.763	3149.399	3275.924	3,881
2015	0.764	0.806	0.696	0.779	3689.370	4285.422	3,163
By Sex							
Males	0.638	0.769	0.558	0.734	3469.531	3786.112	6,588
Females	0.866	0.827	0.775	0.804	2778.865	2850.574	4,007
By Hukou							
Urban Hukou	0.927	0.822	0.835	0.809	3695.558	3873.566	6,113
Rural Hukou	0.463	0.685	0.389	0.629	2569.564	2748.355	4,714
By Education							
Less than high school	0.282	0.546	0.222	0.483	2290.565	2288.369	4,762
High school and above	1.073	0.794	0.969	0.791	3923.529	4033.107	6,065
By Ethnic							
Han	0.732	0.800	0.648	0.770	3238.805	3507.339	10,151
Others	0.627	0.779	0.538	0.738	2702.315	2890.493	676
By Age							
17-29	1.134	0.782	1.038	0.781	3088.470	2621.299	2,178
30–39	0.902	0.810	0.796	0.793	3610.061	3951.781	3,184
40–49	0.527	0.728	0.457	0.687	3186.375	3708.737	3,557
50-60	0.332	0.614	0.270	0.555	2698.540	2902.082	1,908
Full Sample	0.725	0.799	0.641	0.769	3205.308	3474.346	10,827

Note: Both Listening and Speaking have 3 levels: 0-no English Skill, 1-little English skill, 2-good English skill or above

Table 2.2: Estimated effects of Listening ability on ln(monthly wage) from OLS model

	(1)	(2)	(3)	(4)	(5)	(6)
Listening ability (t=2)	0.297***		0.113***		0.0901***	
	(0.0153)		(0.0159)		(0.0153)	
Listening ability (t=3)	0.564***		0.205***		0.130***	
	(0.0178)		(0.0204)		(0.0194)	
Listening ability		0.284***		0.103***		0.0675***
		(0.00854)		(0.0101)		(0.00959)
_cons	7.569***	7.288***	5.566***	5.467***	5.484***	5.386***
	(0.00896)	(0.0155)	(0.0950)	(0.0965)	(0.0909)	(0.0943)
Personal Characteristics	No	No	Yes	Yes	Yes	Yes
Location Characteristics	No	No	No	No	Yes	Yes
Observations	10827	10827	10827	10827	10827	10827
R^2	0.099	0.099	0.239	0.239	0.308	0.308

Note: Standard errors in parentheses. *, ** and *** indicate that the estimate is significant at the 0.1, 0.05 and 0.01 levels, respectively. Personality Characteristics include respondents' education, age, age squared, gender, ethnic, hukou type and health status. Location Characteristics is a dummy represents coastal provinces. The odd columns report the estimation results from the multilevel model, and the even columns report the estimation results from the one-level model. Listening ability (t=1) is the reference group, which is not reported in the odd columns.

Table 2.3: Estimated effects of speaking ability on ln(monthly wage) from OLS model

	(1)	(2)	(3)	(4)	(5)	(6)
Speaking ability (t=2)	0.303***		0.108***		0.084***	
	(0.015)		(0.016)		(0.016)	
Speaking ability (t=3)	0.591***		0.220***		0.139***	
	(0.019)		(0.021)		(0.020)	
Speaking ability		0.297***		0.110***		0.0713***
		(0.009)		(0.010)		(0.010)
_cons	7.586***	7.291***	5.571***	5.460***	5.488***	5.383***
	(0.009)	(0.016)	(0.095)	(0.096)	(0.091)	(0.094)
Personal Characteristics	No	No	Yes	Yes	Yes	Yes
Location Characteristics	No	No	No	No	Yes	Yes
Observations	10827	10827	10827	10827	10827	10827
R^2	0.010	0.010	0.240	0.240	0.308	0.308

Note: Standard errors in parentheses. *, ** and *** indicate that the estimate is significant at the 0.1, 0.05 and 0.01 levels, respectively. Personality Characteristics include respondents' education, age, age squared, gender, ethnic, hukou type and health status. Location Characteristics is a dummy represents coastal provinces.
The odd columns report the estimation results from the multilevel model, and the even columns report the estimation results from the one-level model. Speaking ability (t=1) is the reference group, which is not reported in the odd columns.

Table 2.4: IV (2SLS) estimation in the one-dimensional English ability model

	(Listening)	(Speaking)
Listening ability	0.673***	
	(0.171)	
Speaking ability		0.701***
		(0.178)
_cons	4.248***	4.238***
	(0.313)	(0.316)
Controls	Yes	Yes
First Stage		
number of universities	0.015***	0.016***
	(0.007)	(0.007)
number of children	-0.076***	-0.072***
	(0.011)	(0.012)
F-Stats	24.8	24.5
Hasen J Stats	0.169	0.380
Observations	10815	10815
R^2	0.025	0.027

Note: Standard errors in parentheses. *,** and *** indicate that the estimate is significant at the 0.1, 0.05 and 0.01 levels, respectively.

Table 2.5: Bounds of listening ability effect on ln(monthly wage) and ATEs under ETS , Worst-Case, MTR & MTS

Assumptions	ETS		95% conf.in.		Worst-Case		95% Conf. in.	
			LB	UB	LB	UB	LB	UB
$E[y(t1)]$	7.569		7.552	7.585	6.776	8.797	6.759	8.821
$E[y(t2)]$	7.865		7.841	7.889	6.532	9.391	6.517	9.408
$E[y(t3)]$	8.133		8.104	8.162	6.469	9.589	6.451	9.605
$ATE(t_3-t_2)$	0.268		0.214	0.321	-2.922	3.057	-2.958	3.088
$ATE(t_2-t_1)$	0.297		0.256	0.337	-2.265	2.615	-2.304	2.649
$ATE(t_3-t_1)$	0.564		0.519	0.610	-2.328	2.813	-2.370	2.846
Assumptions	MTR		95% Conf. in.		MTS		95% Confi.in.	
	LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t1)]$	6.776	7.777	6.759	7.791	7.569	8.797	6.759	8.823
$E[y(t2)]$	7.308	8.980	7.290	9.001	6.943	8.335	7.290	8.357
$E[y(t3)]$	7.777	9.589	7.764	9.608	6.469	8.133	7.764	8.164
$ATE(t_3-t_2)$	0.000	2.281	0.000	2.317	-1.866	1.190	-0.593	0.873
$ATE(t_2-t_1)$	0.000	2.204	0.000	2.242	-1.855	0.766	-1.533	1.598
$ATE(t_3-t_1)$	0.000	2.813	0.000	2.849	-2.328	0.564	-1.060	1.405
Assumptions	MTR+MTS		95% Conf. in.					
	LB	UB	LB	UB				
$E[y(t1)]$	7.569	7.777	7.553	7.791				
$E[y(t2)]$	7.719	7.924	7.703	7.945				
$E[y(t3)]$	7.777	8.133	7.765	8.163				
$ATE(t_3-t_2)$	0.000	0.414	0.000	0.460				
$ATE(t_2-t_1)$	0.000	0.355	0.000	0.393				
$ATE(t_3-t_1)$	0.000	0.564	0.000	0.610				

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included. (99 repetitions)

Table 2.6: Bounds of speaking ability effect on ln(monthly wage) and ATEs under ETS , Worst-Case, MTR & MTS

Assumptions	ETS	95% Conf.in.		Worst-Case		95% Conf. in.	
		LB	UB	LB	UB	LB	UB
$E[y(t1)]$	7.586	7.566	7.606	6.855	8.698	6.838	8.724
$E[y(t2)]$	7.889	7.863	7.915	6.530	9.408	6.510	9.427
$E[y(t3)]$	8.177	8.139	8.215	6.392	9.671	6.377	9.685
$ATE(t_3 - t_2)$	0.288	0.224	0.352	-3.016	3.142	-3.050	3.174
$ATE(t_2 - t_1)$	0.303	0.257	0.349	-2.168	2.553	-2.213	2.589
$ATE(t_3 - t_1)$	0.591	0.533	0.648	-2.305	2.816	-2.347	2.847

Assumptions	MTR		95% Conf. in.		MTS		95% Confi.in.	
	LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t1)]$	6.855	7.777	6.836	7.791	7.586	8.698	7.569	8.721
$E[y(t2)]$	7.385	9.080	7.367	9.101	6.870	8.270	6.848	8.297
$E[y(t3)]$	7.777	9.671	7.764	9.686	6.392	8.177	6.375	8.207
$ATE(t_3 - t_2)$	0.000	2.286	0.000	2.319	-1.877	1.307	-1.922	1.359
$ATE(t_2 - t_1)$	0.000	2.224	0.000	2.265	-1.828	0.684	-1.873	0.728
$ATE(t_3 - t_1)$	0.000	2.816	0.000	2.851	-2.305	0.591	-2.347	0.638

Assumptions	MTR+MTS		95% Conf. in.	
	LB	UB	LB	UB
$E[y(t1)]$	7.586	7.777	7.572	7.792
$E[y(t2)]$	7.726	7.941	7.711	7.960
$E[y(t3)]$	7.777	8.177	7.766	8.210
$ATE(t_3 - t_2)$	0.000	0.451	0.000	0.499
$ATE(t_2 - t_1)$	0.000	0.355	0.000	0.388
$ATE(t_3 - t_1)$	0.000	0.591	0.000	0.638

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included (99 repetitions). ATE stands for average treatment effect.

Table 2.7: Bounds of English ability effect on ln(monthly wage) and ATEs under joint assumptions MTR-MTS-MIV

	LB	UB	95/% conf.int.	
			LB	UB
MTR-MTS-MIV (Listening ability)				
$ATE(t_3 - t_2)$	0.000	0.345	0.000	0.404
$ATE(t_2 - t_1)$	0.000	0.330	0.000	0.383
$ATE(t_3 - t_1)$	0.056	0.498	0.011	0.558
N	10872			
MTR-MTS-MIV (Speaking ability)				
$ATE(t_3 - t_2)$	0.000	0.378	0.000	0.441
$ATE(t_2 - t_1)$	0.000	0.321	0.000	0.376
$ATE(t_3 - t_1)$	0.056	0.521	0.012	0.582
N	10872			

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included (99 repetitions). ATE stands for average treatment effect.

Table 2.8: Bounds of listening ability effect on ln(monthly wage) and ATEs by education

Assumptions		ETS	MTR		MTS		MTR+MTS		MTR+MTS+MIV	
Education			LB	UB	LB	UB	LB	UB	LB	UB
Less than highschool										
	$ATE(t_3 - t_2)$	0.017	0.000	2.468	-1.655	1.258	0.000	0.139	0.000	0.118
	$ATE(t_2 - t_1)$	0.160	0.000	2.326	-1.675	0.274	0.000	0.161	0.000	0.139
	$ATE(t_3 - t_1)$	0.177	0.000	2.769	-1.975	0.177	0.000	0.177	0.000	0.160
High school and above										
	$ATE(t_3 - t_2)$	0.221	0.000	2.134	-1.911	0.774	0.000	0.267	0.000	0.183
	$ATE(t_2 - t_1)$	0.161	0.000	2.108	-1.970	0.884	0.000	0.240	0.000	0.188
	$ATE(t_3 - t_1)$	0.383	0.000	2.848	-2.606	0.383	0.000	0.383	0.089	0.307

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included (99 repetitions). ATE stands for average treatment effect.
We don't report confidence interval in the table. Any positive lower bounds are statistically excluded from zero, otherwise equal to zero.

Table 2.9: Bounds of speaking ability effect on ln(monthly wage) and ATEs by education

Assumptions		ETS	MTR		MTS		MTR+MTS		MTR+MTS+MIV	
Education			LB	UB	LB	UB	LB	UB	LB	UB
Less than high school										
	$ATE(t_3 - t_2)$	-0.065	0.000	2.477	-1.667	1.263	0.000	0.078	0.000	0.048
	$ATE(t_2 - t_1))$	0.178	0.000	2.362	-1.636	0.249	0.002	0.176	0.000	0.156
	$ATE(t_3 - t_1)$	0.112	0.000	2.741	-1.903	0.112	0.000	0.112	0.000	0.085
High shool and above										
	$ATE(t_3 - t_2)$	0.253	0.000	2.137	-1.913	0.900	0.000	0.303	0.000	0.224
	$ATE(t_2 - t_1)$	0.153	0.000	2.116	-1.959	0.757	0.000	0.228	0.000	0.170
	$ATE(t_3 - t_1)$	0.405	0.000	2.874	-2.622	0.405	0.000	0.405	0.089	0.328

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included (99 repetitions). ATE stands for average treatment effect.
We don't report confidence interval in the table. Any positive lower bounds are statistically excluded from zero, otherwise equal to zero.

Table 2.10: Bounds of listening ability on ln(monthly wage) and ATEs by gender and hukou type

Assumptions		ETS	MTR+MTS		MIV+MTR+MTS	
			LB	UB	LB	UB
Gender						
	Female					
	$ATE(t_3-t_2)$	0.327	0.000	0.484	0.000	0.417
	$ATE(t_2-t_1)$	0.306	0.000	0.383	0.000	0.367
	$ATE(t_3-t_1)$	0.634	0.000	0.634	0.042	0.564
	Male					
	$ATE(t_3-t_2)$	0.298	0.000	0.405	0.015	0.257
	$ATE(t_2-t_1)$	0.280	0.000	0.377	0.000	0.295
	$ATE(t_3-t_1)$	0.578	0.000	0.578	0.108	0.446
Hukou						
	Rural					
	$ATE(t_3-t_2)$	0.115	0.000	0.250	0.000	0.249
	$ATE(t_2-t_1)$	0.209	0.000	0.222	0.000	0.213
	$ATE(t_3-t_1)$	0.324	0.000	0.324	0.000	0.320
	Urban					
	$ATE(t_3-t_2)$	0.262	0.000	0.375	0.000	0.280
	$ATE(t_2-t_1)$	0.297	0.000	0.377	0.000	0.303
	$ATE(t_3-t_1)$	0.560	0.000	0.560	0.105	0.445

Note: LB, lower bound; UB, upper bound. ATE stands for average treatment effect. We do not include confidence intervals in the table.The positive lower bounds mean that they are statistically excluded from zero.

Table 2.11: Bounds of speaking ability effect on ln(monthly wage) and ATEs by gender and hukou type

Assumptions		ETS	MTR+MTS		MIV+MTR+MTS	
			LB	UB	LB	UB
Gender						
	Female					
	$ATE(t_3 - t_2)$	0.332	0.000	0.532	0.000	0.457
	$ATE(t_2 - t_1)$	0.340	0.000	0.388	0.000	0.369
	$ATE(t_3 - t_1)$	0.672	0.000	0.672	0.042	0.595
	Male					
	$ATE(t_3 - t_2)$	0.291	0.000	0.430	0.000	0.292
	$ATE(t_2 - t_1)$	0.302	0.000	0.371	0.000	0.277
	$ATE(t_3 - t_1)$	0.594	0.000	0.594	0.108	0.472
Hukou						
	Rural					
	$ATE(t_3 - t_2)$	0.135	0.000	0.272	0.000	0.257
	$ATE(t_2 - t_1)$	0.198	0.000	0.208	0.000	0.202
	$ATE(t_3 - t_1)$	0.333	0.000	0.333	0.000	0.316
	Urban					
	$ATE(t_3 - t_2)$	0.266	0.000	0.398	0.000	0.297
	$ATE(t_2 - t_1)$	0.313	0.000	0.381	0.000	0.307
	$ATE(t_3 - t_1)$	0.578	0.000	0.578	0.105	0.460

Note: LB, lower bound; UB, upper bound. ATE stands for average treatment effect. We do not include confidence intervals in the table.The positive lower bounds mean that they are statistically excluded from zero.

Table 2.12: Bounds of listening ability effect on ln(monthly wage) and ATEs by age

Assumptions	ETS	MTR+MTS		MIV+MTR+MTS	
		LB	UB	LB	UB
Age					
17-29					
$ATE(t_3 - t_2)$	0.223	0.000	0.245	0.000	0.176
$ATE(t_2 - t_1)$	0.087	0.000	0.172	0.000	0.141
$ATE(t_3 - t_1)$	0.311	0.000	0.311	0.000	0.229
30-45					
$ATE(t_3 - t_2)$	0.323	0.000	0.481	0.000	0.367
$ATE(t_2 - t_1)$	0.324	0.000	0.391	0.000	0.353
$ATE(t_3 - t_1)$	0.648	0.000	0.648	0.073	0.536
46-60					
$ATE(t_3 - t_2)$	0.233	0.000	0.520	0.000	0.433
$ATE(t_2 - t_1)$	0.403	0.000	0.423	0.000	0.396
$ATE(t_3 - t_1)$	0.635	0.000	0.635	0.000	0.577

Note: LB, lower bound; UB, upper bound. ATE stands for average treatment effect. The positive lower bounds mean that they are statistically excluded from zero.

Table 2.13: Bounds of speaking ability effect on ln(monthly wage) and ATEs by age

Assumptions	ETS	MTR+MTS		MIV+MTR+MTS	
		LB	UB	LB	UB
Age					
$17-29$					
$ATE(t_3-t_2)$	0.208	0.000	0.241	0.000	0.151
$ATE(t_2-t_1)$	0.113	0.000	0.180	0.000	0.128
$ATE(t_3-t_1)$	0.321	0.000	0.321	0.000	0.226
$30-45$					
$ATE(t_3-t_2)$	0.374	0.000	0.542	0.000	0.431
$ATE(t_2-t_1)$	0.320	0.000	0.388	0.000	0.333
$ATE(t_3-t_1)$	0.694	0.000	0.694	0.073	0.581
$46-60$					
$ATE(t_3-t_2)$	0.245	0.000	0.577	0.000	0.490
$ATE(t_2-t_1)$	0.441	0.000	0.458	0.000	0.430
$ATE(t_3-t_1)$	0.686	0.000	0.686	0.000	0.629

Note: LB, lower bound; UB, upper bound. ATE stands for average treatment effect. The positive lower bounds mean that they are statistically excluded from zero.

Table 2.14: Bounds of listening ability effect on ln(monthly wage) and ATEs over time

Assumptions		ETS	MTR+MTS		MTR+MTS+MIV	
			LB	UB	LB	UB
Year						
2012						
	$ATE(t_3 - t_2)$	0.243	0.000	0.399	0.000	0.260
	$ATE(t_2 - t_1)$	0.304	0.000	0.356	0.006	0.291
	$ATE(t_3 - t_1)$	0.548	0.000	0.548	0.100	0.425
2013						
	$ATE(t_3 - t_2)$	0.240	0.000	0.381	0.000	0.262
	$ATE(t_2 - t_1)$	0.283	0.000	0.334	0.015	0.284
	$ATE(t_3 - t_1)$	0.523	0.000	0.523	0.107	0.414
2015						
	$ATE(t_3 - t_2)$	0.321	0.000	0.458	0.000	0.399
	$ATE(t_2 - t_1)$	0.290	0.000	0.365	0.000	0.332
	$ATE(t_3 - t_1)$	0.611	0.000	0.611	0.023	0.550

Note: LB, lower bound; UB, upper bound. ATE stands for average treatment effect. The positive lower bounds mean that they are statistically excluded from zero.

Table 2.15: Bounds of speaking ability effect on ln(monthly wage) and ATEs over time

Assumptions		ETS	MTR+MTS		MTR+MTS+MIV	
			LB	UB	LB	UB
Year						
2012						
	$ATE(t_3 - t_2)$	0.254	0.000	0.430	0.000	0.308
	$ATE(t_2 - t_1)$	0.319	0.000	0.363	0.034	0.263
	$ATE(t_3 - t_1)$	0.573	0.000	0.573	0.100	0.489
2013						
	$ATE(t_3 - t_2)$	0.290	0.000	0.451	0.000	0.324
	$ATE(t_2 - t_1)$	0.287	0.000	0.337	0.015	0.279
	$ATE(t_3 - t_1)$	0.577	0.000	0.577	0.107	0.450
2015						
	$ATE(t_3 - t_2)$	0.318	0.000	0.461	0.000	0.395
	$ATE(t_2 - t_1)$	0.286	0.000	0.349	0.000	0.312
	$ATE(t3 - t1)$	0.604	0.000	0.604	0.023	0.537

Note: LB, lower bound; UB, upper bound. ATE stands for average treatment effect. The positive lower bounds mean that they are statistically excluded from zero.

Chapter 3 Tables

Table 3.1: Summary statistics

	Mandarin Proficiency		Monthly wage		Sample Size
	Mean	SD	Mean	SD	
Full sample	2.991	0.978	4165.015	4996.800	7,115
By Sex					
Males	2.923	0.979	4480.186	4335.027	4,094
Females	3.084	0.970	3737.903	5747.073	3,021
By Hukou					
Urban Hukou	3.361	0.804	5262.296	6249.025	2,797
Rural Hukou	2.752	1.006	3454.248	3816.763	4,318
By Education					
Less than high school	2.626	0.997	3240.817	3945.389	3,806
Just a high school degree	3.183	0.851	3947.374	3571.990	1,533
College	3.611	0.615	6327.510	7009.188	1,772
Living Place					
Urban	3.292	0.841	4918.216	5583.436	3,623
Rural	2.680	1.012	3383.560	4163.640	3,492
By Age					
16-29	3.410	0.727	4022.833	4995.000	1,404
30–45	3.158	0.895	4795.422	5835.368	2,912
45–65	2.608	1.035	3580.479	3844.342	2,799

Table 3.2: Mandarin fluency and ln(Monthly wage)

Mandarin Proficiency	Monthly wage		
	All	Male	Female
Excellent	8.254	8.370	8.134
Good but wih accent	8.007	8.127	7.823
Not very well	7.744	7.855	7.566
Not at all	7.659	7.791	7.442

Table 3.3: Determinants of monthly wages for full sample: OLS Model

	(1)	(2)	(3)	(4)
	model1	model2	model3	model4
Mandarin proficiency	0.213***			0.064***
	(0.009)			(0.010)
Mandarin proficiency (t=2)		0.085***	-0.001	
		(0.033)	(0.032)	
Mandarin proficiency (t=3)		0.349***	0.108***	
		(0.027)	(0.029)	
Mandarin proficiency (t=4)		0.595***	0.171***	
		(0.028)	(0.032)	
Age			0.047***	0.047***
			(0.006)	(0.006)
Age squared			-0.001***	-0.001***
			(0.000)	(0.000)
Education			0.084***	0.084***
			(0.004)	(0.004)
Gender			-0.313***	-0.313***
			(0.016)	(0.016)
Urban			0.090***	0.090***
			(0.024)	(0.024)
Marriage			0.112***	0.112***
			(0.027)	(0.027)
Constant	7.385***	7.659***	7.348***	7.266***
	(0.027)	(0.024)	(0.134)	(0.136)
City Fixed	No	No	Yes	Yes
Sector Fixed	No	No	Yes	Yes
N	7115	7115	7008	7008
R^2	0.078	0.080	0.294	0.294

Note: Standard errors in parentheses. *,** and *** indicate that the estimate is significant at the 0.1, 0.05 and 0.01 levels, respectively.

Table 3.4: Determinants of monthly wages for full sample: IV Model

	(1) All	(2) Male	(3) Female
Mandarin proficiency	0.316***	0.272***	0.380***
	(0.056)	(0.069)	(0.092)
Constant	6.339***	5.903***	5.787***
	(0.248)	(0.311)	(0.399)
Controls	Yes	Yes	Yes
City Fixed	Yes	Yes	Yes
Observations	7008	4035	2973
R^2	0.249	0.254	0.247
First Stage			
F-statistic	255.790	159.977	104.196

Note: Standard errors in parentheses. *,** and *** indicate that the estimate is significant at the 0.1, 0.05 and 0.01 levels, respectively.
Control variables include age, age squared, education, gender, hukou type, location (urban or rural), marital status, sector.

Table 3.5: Bounds of Mandarin proficiency effect on ln(monthly wage) and ATEs

Assumptions	Worst-Case		95% conf.in.		MTR		95% Conf. Int.	
	LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t1)]$	4.440	9.719	4.417	9.734	4.440	8.021	4.418	8.032
$E[y(t2)]$	4.458	9.724	4.428	9.743	4.898	8.302	4.879	8.316
$E[y(t3)]$	5.617	9.196	5.572	9.220	6.515	8.578	6.491	8.587
$E[y(t4)]$	5.506	9.382	5.462	9.403	8.021	9.382	8.010	9.396
$ATE(t_4 - t_3)$	-3.690	3.765	-3.758	3.832	0.000	2.867	0.000	2.905
$ATE(t_3 - t_2)$	-4.107	4.738	-4.172	4.792	0.000	3.680	0.000	3.709
$ATE(t_2 - t_1)$	-5.261	5.284	-5.306	5.327	0.000	3.862	0.000	3.897
$ATE(t_3 - t_1)$	-4.102	4.756	-4.162	4.803	0.000	4.138	0.000	4.169
$ATE(t_4 - t_1)$	-4.212	4.942	-4.271	4.986	0.000	4.942	0.000	4.978
Assumptions	MTS		95% Conf. in.		MTR+MTS		95% Confi.in.	
	LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t1)]$	7.659	9.719	7.610	9.737	7.659	8.021	7.606	8.035
$E[y(t2)]$	7.294	9.453	7.246	9.478	7.734	8.031	7.694	8.043
$E[y(t3)]$	7.036	8.713	6.991	8.744	7.933	8.095	7.912	8.111
$E[y(t4)]$	5.506	8.254	5.452	8.283	8.021	8.254	8.004	8.285
$ATE(t_4 - t_3)$	-3.207	1.219	-3.292	1.292	0.000	0.321	0.000	0.372
$ATE(t_3 - t_2)$	-2.417	1.419	-2.488	1.498	0.000	0.361	0.000	0.417
$ATE(t_2 - t_1)$	-2.424	1.794	-2.491	1.868	0.000	0.372	0.000	0.436
$ATE(t_3 - t_1)$	-2.683	1.054	-2.746	1.134	0.000	0.436	0.000	0.504
$ATE(t_4 - t_1)$	-4.212	0.595	-4.285	0.673	0.000	0.595	0.000	0.678
Assumptions	MTR+MTS+MIV		95% Conf. in.					
	LB	UB	LB	UB				
$E[y(t1)]$	7.688	8.008	7.583	8.026				
$E[y(t2)]$	7.821	8.018	7.714	8.038				
$E[y(t3)]$	8.024	8.080	7.971	8.098				
$E[y(t4)]$	8.113	8.187	8.066	8.227				
$ATE(t_4 - t_3)$	0.033	0.163	0.000	0.256				
$ATE(t_3 - t_2)$	0.006	0.258	0.000	0.384				
$ATE(t_2 - t_1)$	0.000	0.330	0.000	0.455				
$ATE(t_3 - t_1)$	0.016	0.392	0.000	0.516				
$ATE(t_4 - t_1)$	0.105	0.499	0.040	0.644				

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included. (99 repetitions)

Table 3.6: Bounds of ln(monthly wage) and ATEs by gender

Assumptions	ETS	95% conf.in.		MTR+MTS		95% Conf. Int.		MTR+MTS+MIV		95% Conf. Int.	
Male											
		LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t1)]$	7.791	7.733	7.850	7.791	8.124	7.738	8.148	7.802	8.111	7.673	8.135
$E[y(t2)]$	7.855	7.805	7.905	7.847	8.132	7.797	8.161	7.999	8.118	7.836	8.141
$E[y(t3)]$	8.127	8.085	8.168	8.048	8.203	8.021	8.234	8.149	8.188	8.088	8.211
$E[y(t4)]$	8.370	8.332	8.409	8.124	8.370	8.101	8.415	8.223	8.312	8.175	8.359
$ATE(t_4 - t_3)$	0.244	0.163	0.324	0.000	0.323	-0.133	0.394	0.035	0.163	0.000	0.271
$ATE(t_3 - t_2)$	0.272	0.180	0.363	0.000	0.356	-0.140	0.438	0.032	0.189	0.000	0.375
$ATE(t_2 - t_1)$	0.064	-0.045	0.172	0.000	0.341	-0.351	0.423	0.000	0.315	0.000	0.468
$ATE(t_3 - t_1)$	0.336	0.236	0.436	0.000	0.412	-0.127	0.496	0.038	0.385	0.000	0.538
$ATE(t_4 - t_1)$	0.579	0.482	0.676	0	0.579	-0.047	0.677	0.112	0.510	0.040	0.686
	ETS	95% conf.in.		MTR+MTS		95% Conf. Int.		MTR+MTS+MIV		95% Conf. Int.	
Female											
		LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t1)]$	7.442	7.378	7.506	7.442	7.881	7.372	7.905	7.442	7.867	7.270	7.896
$E[y(t2)]$	7.566	7.495	7.636	7.552	7.894	7.488	7.920	7.657	7.880	7.520	7.910
$E[y(t3)]$	7.823	7.784	7.862	7.754	7.950	7.722	7.978	7.833	7.934	7.774	7.967
$E[y(t4)]$	8.134	8.094	8.173	7.881	8.134	7.855	8.176	7.972	8.057	7.909	8.105
$ATE(t_4 - t_3)$	0.311	0.232	0.389	0.000	0.380	-0.123	0.454	0.039	0.224	0.000	0.331
$ATE(t_3 - t_2)$	0.257	0.147	0.368	0.000	0.398	-0.199	0.489	0.000	0.276	0.000	0.447
$ATE(t_2 - t_1)$	0.124	-0.011	0.258	0.000	0.452	-0.417	0.549	0.000	0.439	0.000	0.640
$ATE(t_3 - t_1)$	0.381	0.278	0.484	0.000	0.508	-0.184	0.606	0.000	0.492	0.000	0.697
$ATE(t_4 - t_1)$	0.692	0.589	0.794	0.000	0.692	-0.051	0.804	0.105	0.616	0.013	0.835

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included. (99 repetitions)

Table 3.7: Bounds of ln(monthly wage) and ATEs by age group

Assumptions	ETS	95% conf.in.		MTR+MTS		95% Conf. Int.		MTR+MTS+MIV		95% Conf. Int.	
16-45											
		LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t1)]$	7.743	7.664	7.822	7.743	8.111	7.665	8.134	7.918	8.095	7.706	8.129
$E[y(t2)]$	7.837	7.760	7.914	7.831	8.117	7.764	8.141	7.998	8.102	7.823	8.126
$E[y(t3)]$	8.050	8.023	8.077	8.013	8.148	7.987	8.170	8.100	8.133	8.043	8.157
$E[y(t4)]$	8.269	8.237	8.301	8.111	8.269	8.092	8.306	8.203	8.204	8.162	8.240
$ATE(t_4 - t_3)$	0.219	0.160	0.278	0.000	0.255	0.000	0.319	0.070	0.104	0.005	0.197
$ATE(t_3 - t_2)$	0.213	0.109	0.318	0.000	0.317	0.000	0.407	0.000	0.135	0.000	0.334
$ATE(t_2 - t_1)$	0.094	-0.062	0.249	0.000	0.374	0.000	0.476	0.000	0.184	0.000	0.419
$ATE(t_3 - t_1)$	0.307	0.201	0.413	0.000	0.405	0 .000	0.506	0.000	0.215	0.000	0.451
$ATE(t_4 - t_1)$	0.526	0.415	0.636	0.000	0.526	0.000	0.641	0.108	0.286	0.033	0.534
	ETS	95% conf.in.		MTR+MTS		95% Conf. Int.		MTR+MTS+MIV		95% Conf. Int.	
45-64											
		LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t1)]$	7.619	7.566	7.672	7.619	7.881	7.569	7.905	7.687	7.879	7.567	7.906
$E[y(t2)]$	7.682	7.624	7.740	7.669	7.894	7.632	7.925	7.677	7.891	7.578	7.927
$E[y(t3)]$	7.940	7.894	7.986	7.825	7.995	7.797	8.025	7.861	7.988	7.798	8.023
$E[y(t4)]$	8.206	8.143	8.269	7.881	8.206	7.858	8.261	7.901	8.164	7.839	8.247
$ATE(t_4 - t_3)$	0.266	0.158	0.375	0.000	0.381	0.000	0.464	0.000	0.303	0.000	0.449
$ATE(t_3 - t_2)$	0.258	0.154	0.362	0.000	0.327	0.000	0.393	0.000	0.310	0.000	0.445
$ATE(t_2 - t_1)$	0.063	-0.048	0.174	0.000	0.275	0.000	0.356	0.000	0.204	0.000	0.361
$ATE(t_3 - t_1)$	0.321	0.222	0.419	0.000	0.376	0.000	0.455	0.000	0.300	0.000	0.456
$ATE(t_4 - t_1)$	0.587	0.472	0.703	0.000	0.587	0.000	0.691	0.022	0.477	0.000	0.680

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included. (99 repetitions)

Table 3.8: Bounds of ln(monthly wage) and ATEs by location

Assumptions	ETS	95% conf.in.		MTR+MTS		95% Conf. Int.		MTR+MTS+MIV		95% Conf. Int.	
Urban											
		LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t1)]$	7.734	7.633	7.835	7.734	8.199	7.631	8.219	7.749	8.191	7.565	8.213
$E[y(t2)]$	7.905	7.837	7.974	7.896	8.208	7.829	8.227	8.006	8.199	7.867	8.227
$E[y(t3)]$	8.135	8.098	8.173	8.094	8.240	8.063	8.263	8.138	8.233	8.070	8.255
$E[y(t4)]$	8.351	8.318	8.383	8.199	8.351	8.173	8.375	8.237	8.320	8.214	8.366
$ATE(t_4 - t_3)$	0.215	0.146	0.285	0.000	0.257	0.000	0.312	0.004	0.181	0.000	0.297
$ATE(t_3 - t_2)$	0.230	0.124	0.336	0.000	0.344	0.000	0.434	0.000	0.226	0.000	0.388
$ATE(t_2 - t_1)$	0.171	0.002	0.341	0.000	0.474	0.000	0.596	0.000	0.451	0.000	0.661
$ATE(t_3 - t_1)$	0.401	0.263	0.540	0.000	0.506	0.000	0.632	0.000	0.484	0.000	0.689
$ATE(t_4 - t_1)$	0.617	0.483	0.750	0.000	0.617	0.000	0.744	0.046	0.571	0.001	0.801
	ETS	95% conf.in.		MTR+MTS		95% Conf. Int.		MTR+MTS+MIV		95% Conf. Int.	
Rural											
		LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t1)]$	7.636	7.586	7.686	7.636	7.836	7.584	7.857	7.690	7.827	7.545	7.851
$E[y(t2)]$	7.658	7.607	7.708	7.654	7.840	7.607	7.865	7.658	7.833	7.580	7.858
$E[y(t3)]$	7.893	7.856	7.931	7.807	7.922	7.778	7.953	7.872	7.911	7.779	7.941
$E[y(t4)]$	8.029	7.980	8.078	7.836	8.029	7.811	8.082	7.925	7.987	7.862	8.046
$ATE(t_4 - t_3)$	0.136	0.050	0.222	-0.086	0.222	-0.142	0.304	0.014	0.114	0.000	0.266
$ATE(t_3 - t_2)$	0.235	0.147	0.323	-0.034	0.269	-0.086	0.345	0.040	0.253	0.000	0.361
$ATE(t_2 - t_1)$	0.022	-0.079	0.123	-0.183	0.204	-0.250	0.280	0.000	0.143	0.000	0.313
$ATE(t_3 - t_1)$	0.257	0.170	0.345	-0.029	0.287	-0.079	0.368	0.045	0.221	0.000	0.396
$ATE(t_4 - t_1)$	0.393	0.294	0.492	0.000	0.393	-0.046	0.498	0.098	0.297	0.011	0.500

Note: LB, lower bound; UB, upper bound. Bootstrap 5th and 95th percentiles of lower and upper bounds are included. (99 repetitions)

Table 3.9: Self-reported satisfaction on Job Match: Ordered Probit Model

	(1) Full	(2) Male	(3) Female
Job Match			
Mandarin proficiency	0.052***	0.049**	0.047*
	(0.017)	(0.023)	(0.028)
Gender	-0.040	0.0000	0.0000
	(0.0270)	(.)	(.)
Age	0.005***	0.004**	0.006**
	(0.002)	(0.002)	(0.003)
Education group(2)	0.197**	0.238*	0.112
	(0.079)	(0.137)	(0.105)
Education group(3)	0.312***	0.321**	0.292**
	(0.085)	(0.142)	(0.118)
Education group(4)	0.470***	0.460***	0.471***
	(0.088)	(0.145)	(0.122)
Marital Status	-0.009	0.023	-0.041
	(0.042)	(0.054)	(0.067)
Urban dummy	-0.035	-0.091*	0.043
	(0.042)	(0.055)	(0.065)
Sector Fixed	Yes	Yes	Yes
City Fixed	Yes	Yes	Yes
N	7007	4034	2973

Note: Standard errors in parentheses. *,** and *** indicate that the estimate is significant at the 0.1, 0.05 and 0.01 levels, respectively.
Education has four groups, group 1 is the lowest level represents people have no education experience, which also serves as the reference group, hence not shown. Group 2 means less than a high school degree. Group 3 means just high school or equivalent degrees. Group 4 means college degree and above Marital status means if respondents are married, and non-married group serves as the reference group. Urban dummy represents if respondents live in the urban areas, and rural areas serve as the reference group.

Table 3.10: Self-reported satisfaction on Promotion Opportunity: Ordered Probit Model

	(1)	(2)	(3)
	Full	Male	Female
Promotion			
Mandarin proficiency	0.007	-0.005	0.005
	(0.022)	(0.029)	(0.036)
Gender	0.010	0.0000	0.0000
	(0.033)	(.)	(.)
Age	0.003^{*}	0.006^{**}	0.001
	(0.002)	(0.002)	(0.003)
Education group(2)	-0.025	0.057	-0.130
	(0.110)	(0.176)	(0.151)
Education group(3)	0.125	0.229	0.006
	(0.115)	(0.182)	(0.162)
Education group(4)	0.319^{***}	0.493^{***}	0.106
	(0.117)	(0.184)	(0.165)
Marital Status	-0.124^{***}	-0.183^{***}	-0.075
	(0.048)	(0.064)	(0.074)
Urban dummy	-0.022	-0.047	0.017
	(0.052)	(0.068)	(0.083)
Sector Fixed	Yes	Yes	Yes
City Fixed	Yes	Yes	Yes
N	4802	2756	2046

Note: Standard errors in parentheses. *,** and *** indicate that the estimate is significant at the 0.1, 0.05 and 0.01 levels, respectively.
Education has four groups, group 1 is the lowest level represents people have no education experience, which also serves as the reference group, hence not shown. Group 2 means less than a high school degree. Group 3 means just high school or equivalent degrees. Group 4 means college degree and above Marital status means if respondents are married, and non-married group serves as the reference group. Urban dummy represents if respondents live in the urban areas, and rural areas serve as the reference group.