The effects of English proficiency on labour market performance in China

By Bing Li

(8062339)

Major paper presented to the

Department of Economics of the University of Ottawa

in partial fulfillment of the requirements of the M.A. Degree

Supervisor: Professor Gilles Grenier

ECO6999

Ottawa, Ontario

April 2018

Content

Abstract

With economic development and globalization, mastering English has become a crucial skill in the labour market in China and proficiency in English is expected to increase earnings. This paper uses data from the Chinese General Social Survey (CGSS) of 2013 to estimate the determinants of English fluency and its effects on earnings in China's labour market. The results show that gender, geography and education are the most crucial factors that are related to English speaking fluency. Males are less likely to be fluent in English than females and living in the coastal area is strongly correlated with English speaking. When the education level increases, the probability of being fluent in English also increases. Most importantly, using an OLS regression, I estimate that speaking fluent English can increase wages by about 9%. Gender, geographic variables and education are also important determinants of earnings. An IV regression is also estimated with parental education as instrument for English fluency. The returns to English fluency are higher than in the OLS regressions, but they are extremely large and imprecise, so that the results should be considered as inconclusive at this point.

1. Introduction

Language skills are regarded as an important human capital component in the labour market (Grenier, 2015). In China, the first language that most people speak is Mandarin, but with economic development and globalization, more and more young Chinese people learn English. Mastering English has become a crucial skill in the labour market and in general proficiency in English is expected to increase earnings in China.

In this paper, I examine the determinants of English speaking proficiency and its effects on earnings in China's labour market. In order to do so, I use data from the Chinese General Social Survey (CGSS) of 2013. This survey contains data on English proficiency and on the potential factors that may affect English speaking fluency and earnings. Besides English proficiency, I include other variables such as gender, geographic variables, age, education and work-related characteristics. The main finding is that speaking fluent English is positively correlated with wages. In addition, gender, geographic variables and education are important factors that contribute to the determination of English speaking fluency and earnings in China.

This paper is organized as follows. Section 2 provides a literature review on the impact of language skills on labour market performance in different countries, including China. Section 3 shows the details of the data and variables and presents some descriptive statistics. In section 4, I provide the econometric model and its

specifications. Section 5 presents and discusses the empirical results. Section 6 is the conclusion of this paper.

2. Literature Review

Numerous scholars have analyzed the role of language skills on the labour market performance and some have done it specifically for China. Therefore, in this section, I will summarize some of the key studies that have examined the relationship between language skills and labour market performance in various (mainly developed) countries, and then I will consider the effects of language skills on earnings in China.

2.1 The impact of language skills on labour market performance

Much of the early research on that topic was done in Canada. For instance, Carliner (1981) studied wage differentials among different language groups based on the 1971 Canadian Census. The results showed that the native French-speaking men, even those who have learned English, earned less than monolingual English-speaking men in Quebec. The analysis also suggested that the wages tended to be higher in Toronto, Montreal, and Quebec than in the rest of Canada. Subsequently, Shapiro and Stelcner (1997) studied the linguistic earnings differentials in Quebec by gender using the 1991 Canadian Census data and then compared the results with 1971 and 1981 Census data. They pointed out that the earnings gap had narrowed between Anglophones and Francophones based on the estimates for 1980, and that Allophones are the most disadvantaged group. Continuing in the analysis of the situation of

Quebec, Albouy (2008) analyzed the wage gap between Francophone and Anglophone men in Quebec using the Canadian Censuses between 1971 and 2001. The author concluded that the wage gap between Francophones and Anglophones narrowed mainly because the relative wages of Quebec Anglophones decreased. Apart from the wage decline, another explanation for the fall of the wage gap is the increasing level of education level for Francophones.

Following these Canadian studies, Christofides and Swidinsky (2010) analyzed the returns to speaking a second official language at work in Quebec and the rest of Canada (ROC) with 2001 Canadian Census data. They found that in the rest of Canada, the bilingual men who frequently use French at work earn the most. At the same time, the findings for women in the ROC are similar and there even exists a more significant relationship between language of work and earnings. However, the findings for individuals in Quebec are very different. Bilingual individuals who frequently use English at work earn the most among both men and women.

This topic was also studied in the context of other countries. In the United States, an early study by Grenier (1984) investigated the effects of language characteristics on the wages of Hispanic males, using data from the 1976 Survey of Income and Education (SIE). He pointed out that Hispanic males earned 25.8% less than whites. Specifically, more than one third of this difference in the log wage can be attributed to the language handicap.

Chiswick and Miller (1995) examined the relationship between earnings and language skills. They used Census data from Australia between 1981 and 1986. In the 1981 census, individuals fluent in English earn 5.3% more than those from non-English-speaking countries. In the 1986 census, the earnings are 8.3% higher than those from non-English-speaking countries based on OLS regression, which suggests that the returns to English language skills increased over time.

Hayfron (2001) studied the labour market returns to Norwegian language proficiency for Third World male immigrants in Norway. The main finding is that immigrants who participated in a language training program are more likely to acquire proficiency in speaking and reading than those who did not. Younger immigrants and those who stay in Norway for a longer time also acquire proficiency in the Norwegian language more easily. However, the results suggest that Norwegian language proficiency has no significant effect on earnings. One of the possible reasons that the author proposed is that although immigrants need Norwegian language proficiency to be employed in the Norwegian labour market, once they have jobs, their wages are not necessarily determined by that proficiency. Moreover, immigrants with perfect English literacy easily acquire literacy in the Norwegian language. The reason is that both English and Norwegian languages have similar grammatical structure and vocabulary.

In a study on the United Kingdom, Shields and Price (2002) explored two crucial aspects of assimilation experience, the determinants of English fluency and the effects of English fluency on occupational success, using data from the Fourth National Survey of Ethnic Minorities undertaken in 1994. They showed that English language speaking fluency is the second most important determinant of occupational success, the first one being the possession of a degree or equivalent highest qualification. Using instrumental variables for language fluency (language of interview, whether married to a UK born spouse and number of dependent children in the household), the results suggest that occupational wage increases by approximately 16.5% for those with fluent English speaking compared to individuals who do not have such proficiency. However, those results were based on a sample of men in the position of paid employment, meaning that if individuals who are in the sample are more able or motivated that those who are not in it, the study may underestimate the effect of language fluency.

Subsequently to that study, Dustmann and Fabbri (2003) studied the effect of language on earnings based on two UK surveys: The Family and Working Lives Survey (FWLS) and The Fourth National Survey on Ethnic Minorities (FNSEM). They found that language proficiency has a positive effect on employment probability and on earnings and that lack of English fluency may lead to earning losses. However, the data that they used included only immigrants who belong to the ethnic minority communities, that is, only 49% of the total immigrants in the UK. Therefore, they

concluded that more comprehensive surveys are necessary to study the effect of language on earnings for immigrants in the UK.

Bleakley and Chin (2004) studied the effect of language skills on earnings by using individual-level data from the 1990 U.S. Census. They asked whether younger children learn languages more easily than older ones and used age at arrival as an instrumental variable for language proficiency. They found that most of the effect of language skills on wages appears to be mediated by the effect on years of schooling. This means that the role of language proficiency as an input to the production of human capital is more important than the direct effect of language on the marginal product of labour. Subsequently to that research, Wang and Wang (2011) estimated the language effects on the earnings distribution among adults who immigrated to the United States as children and drew the conclusion that there exists a significant heterogeneity in language premiums across the earnings distribution and that the patterns changed over time.

Maxwell (2010) examined the requirement of low-skilled jobs to find the effect of the joint use of English and other skills on wages. The author used the Bay Area Longitudinal Surveys (BALS) which is collected from a single labour market with a high proportion of immigrant enclaves in three counties of the San Francisco Bay Area from June 1998 to October 2002. The results suggest that English skills are required even in low-skilled jobs which are opened to workers with limited English.

For instance, over two-thirds of low-skilled jobs require workers with limited English proficiency to have the ability to read work-related schedules, written instructions, and reading product labels and manuals.

There are also scholars who explored the relationship between English proficiency and earnings in developing countries. Azam, Chin and Prakash (2010) used data set from the 2005 India Human Development Survey (IHDS2005) to investigate the effects of English speaking on wages in India. The study suggests that, based on OLS results, the hourly wages of men who are fluent in English are 34% higher than those of men who do not speak English. For women, being fluent in English increases the hourly wages by 22%. Casale and Posel (2011) investigated the relationship between English proficiency and earnings in South Africa. The data set that they used is from the National Income Dynamics Survey (NIDS) of 2008. By using home language proficiency among Africans as instrumental variable, they found that men in the South African labour market who are English proficient earn 39% more than those who are not. Furthermore, the analysis shows that South African men with post-secondary education earn approximately 90 percent more if they are also English proficient.

In a recent investigation, Yao and van Ours (2015) conducted a study to measure the effect of Dutch language skills in the Netherlands using the Longitudinal Internet Studies for the Social sciences (LISS) panel. They used two instrumental variables:

the language spoken during childhood and age at arrival in the Netherlands. The 2SLS results indicated that the wages of female immigrants without language problems are 48% higher than those with language problems after controlling for personal characteristics. In addition, language problems have no effects on employment and working time for all immigrants. However, for male immigrants, Dutch language skills seem to be less important than for female immigrants.

To sum up, the above literature indicates that language skills have significantly positive impacts on earnings, especially the English language. English language proficiency affects the labour market more than other languages. English skills are required even in low-skilled jobs which are opened to workers with limited English. In addition, education is another important factor that is positively associated with earnings.

2.2 The effects of language skills in China

There are some studies that analyzed the economic return to speaking Mandarin. For instance, Gao and Smyth (2010) examined the returns to speaking standard Mandarin among internal migrants in China's urban labour market. They collected data from the 2005 China Urban Labour Survey. They found that speaking standard Mandarin could generate significant economic returns based on a 2SLS regression, where age at arrival is used as instrumental variable. Specifically, the coefficient on fluency in standard Mandarin for females (51%) is statistically significant and larger than that of

males (34%). A plausible explanation for this phenomenon is that women are more likely to be in occupations that need contact with local people than men.

Guo and Sun (2014) analyzed the economic returns to English proficiency for college graduates in Mainland China using OLS regressions. The data set is the 2010 Chinese College Student Survey, provided by the China Data Center of Tsinghua University. The results indicate that English proficiency, which is based on the College English Test-Band 4 (CET-4) scores of the participants, has positive effects on college graduates' starting salary, on their probability of changing permanent residence status from rural to urban, and on their future earnings potential. The results also suggest that higher English proficiency did not seem to have been achieved at the expense of the other human capital investments that positively affect college graduates' earnings.

Liu (2014) examined the relationship between English proficiency and incomes in China's urban labour market based on the data from the China's General Social Survey (CGSS2006). The article used an extended Mincer earnings equation to estimate the rate of return of foreign language proficiency and found that proficiency in foreign language ability has a significant positive effect on individual earnings based on OLS model.

A recent article by Wang, Smyth and Cheng (2017) studied the economic returns to English proficiency in China by using panel data from the China Labor-Force

Dynamics Survey (CLDS). The authors regressed the log of the hourly wage on years of schooling, the key independent variable, English language proficiency, and a full set of control variables. According to the results of their study, the economic returns to proficiency in English are substantial (16%) based on the 2SLS estimates. Consequently, those people who are truly proficient in English can earn a large premium for speaking English well. The authors also pointed out that the economic returns to English proficiency are higher in urban areas than rural areas in China.

In conclusion, the above studies demonstrated that language skills are positively associated with earnings in different countries. Most of these studies used OLS regressions to reach this conclusion. In addition, in order to better address the problem of causality and to correct for unobserved heterogeneity and measurement errors, some studies used 2SLS by adding instrumental variables. The studies also showed that there are many variables such as gender and education, which affect language variables and wages. In this paper, I will control for those variables and provide a detailed analysis of the effects of English language proficiency on earnings in China's labour market using a recent dataset.

3. Data and Descriptive Statistics

3.1 Dataset

In this paper, I use data from the Chinese General Social Survey (CGSS) of 2013, which is a component of one of the earliest national comprehensive academic surveys in China. The CGSS system collects data from multiple elements of the society, including community, family and individuals, which provide essential information on demographic, economics, personal characteristics and many other popular features in China. CGSS 2013 is the fourth annual survey of the second stage of the CGSS (2010-2019), and it is the tenth year of CGSS.

The survey collected data from 100 counties (districts) across the country, plus five major cities, which are Beijing, Shanghai, Tianjin, Guangzhou and Shenzhen, as primary sampling units. The data contains a total of 11,438 individual records and 650 variables. Among all those variables, there are four that concern language skills: Mandarin listening, Mandarin speaking, English listening and English speaking. Since the purpose of this paper is to examine the determinants of English proficiency and the impact of English skills on China's labour market performance, I will focus on the variables of English proficiency.

3.2 Sample Restrictions

In order to eliminate some extraneous factors and to effectively measure the labour market performance of individuals effectively in this sample, I impose some restrictions on the dataset. Firstly, since this paper focuses on the individuals who belong to the labour force in China, I set a constraint on age by excluding younger

individuals who are under 16 years old; I also exclude males who are above 60 and females who are above 55, since the retirement age in China is 60 for males and 55 for females. Then I drop the individuals whose annual wages are equal to zero China Yuan (CNY). Finally, I eliminate the lowest one percent and highest one percent of individual wages since the very small and the very large observations may be considered as outliers. After combining all those restrictions and dropping some observations with missing values on variables such as age and education, the final sample contains 5,772 usable observations.

3.3 Variables

In this section, I define the variables that I need in my paper.

3.3.1 Dependent Variables

In order to measure how the labour market performance in China is affected by English language proficiency, I use two dependent variables: the proficiency in English speaking and the natural logarithm of annual wages of individuals in 2012. The first dependent variable is defined as a binary variable.

3.3.2 Independent Variables

According to the previous literature, many factors can influence language skills and annual wages, such as gender, age, education level, place of residence, language skills and work status. These independent variables can be divided by two categories:

human capital and working status. In this part, I provide a description of the main explanatory variables.

Human capital variables:

Language proficiency

The key independent variable in the wage regression is the language proficiency which is also the dependent variable in the first regression. The questionnaire of the CGSS includes questions on Mandarin and English. In this paper, I focus on English speaking proficiency. In the survey, self-assessed proficiency of English speaking is measured in five levels: 1 = cannot speak English; 2 = can speak a few English words; 3 = moderately fluent in English; 4 = fluent in English; 5 = very fluent in English. I divide the English-speaking proficiency into two groups, defining level 3,4 and 5 as fluent, and levels 1and 2 as not fluent. I then set the English-speaking proficiency as a dummy variable: 1 = fluent, and 0 = not fluent.

Education

Education is another essential human capital characteristic for labour market performance. In this paper, I specify the education level into five categories: Lower Education, Middle, High, College and University. The first group is the lowest education level, which contains people who were never educated or who only finished primary education. The second group, called “middle”, includes people who finished secondary school education. The third group is called “high” and includes people who

finished either their high school or graduated from a technical school. The fourth dummy variable is called “college” and includes people who hold a college diploma. The final group is “university”, which means people who hold a bachelor’s or a higher-level degree. The lowest education level is taken as a benchmark.

Region

The dataset contains 29 geographical variables, which I divide into coastal and central regions. The way of dividing regions is based on the economic development level of the Chinese provinces. The most developed regions in China are the coastal regions, which contain (from North to South) Hebei, Beijing, Tianjin, Shandong, Jiangsu, Shanghai, Zhejiang, Fujian and Guangdong provinces. The central regions include all the other provinces that are included in the survey. It is expected that people in the coastal areas earn more and are more fluent in English, as they also need to face more challenge and competition.

Gender, marital status and place of residence

I use dummy variables to define gender, marital status and place of residence. For gender, I use the value of one to represent males and the value of zero to represent females. For marital status, I set the variable which is called “marriage” equal to one if the individual is married (first time or more than one time) or separated but not divorced. The value of zero includes those who are currently single, never legally married, divorced or widowed. Regarding the place of residence, the dummy variable

called “rural” define the area where people live. The value of one includes people who are currently living in rural areas, while the value of zero is for individuals who are living in the urban areas.

Work-related variables:

Industry

One of the key factors that is linked to the use of English and to the labour market performance is the kind of economic sector where people work. I divide sectors into 11 categories based on the standard classification of China’s industries: Manufacturing, Construction, Transportation/post, Whole Sale/Retail/Cafeteria, Social Service, Farm/Forestry/Fishing, Mining, Electricity/Gas/Water, Education/Science/Broadcasting and Government/Social Organization. I use Social service as the reference group.

Occupation

Occupation is also an important component of individuals’ work. I divide occupation into four groups: Officials/Manager, Professional/Technician, Service/Sale Workers and Plant and Machine Operators/Assemblers in order to represent the different category of occupations of China’s people. I use Official/Manager as a benchmark.

3.4 Descriptive Statistics

Table 1 shows the mean values of the selected variables in my dataset, first for the total sample, and then separately for males and females, for those who are fluent and those who are not fluent in English, for rural and urban areas, and for coastal and central areas.

In this sample, 59.8% of the individuals who work are males and 40.2% are females. The mean log of annual wages is 9.868, where the value for males is 10.039 and the one for females is 9.615. Therefore, it can be seen that on average males earn more than females in China by about 42 percent (the difference between the two log wages).

Regarding the variable of English speaking proficiency, we can see that the large majority (86.6%) of Chinese are not fluent in English. There is a smaller proportion of males who are fluent than of females (difference of about 5 percentage points). The geographic locations are strongly related to English speaking proficiency. People who live in urban areas are 15.7 percentage point more fluent in English than those who live in rural areas, and people who live in the coastal regions are 16.2 percentage point more fluent in English than those who live in the central regions. For those who are not fluent in English speaking, the average age is 41.8 years old, while people who are fluent in English are much younger, with an average age of 33.2 years. The mean of log of annual wages for people with fluent English is 10.579, which is quite larger than that of people without fluency in English (9.758).

Apart from the language variable, the geographical variables are also important for this study. It is noticeable that urban residents earn more than rural residents; similarly, the average of log of earnings for people who live in coastal areas is much higher than that for people who live in central areas.

The table also shows the mean values of the independent variables including age, education level, marital status and work-related variables. For the age variable, the total average age for people who participate in this survey about 40.6, where the average age for males is 41.8 and is 38.9 for females.

Comparing the education levels, more males finished middle school and high school education than females. However, more females hold bachelor's and master's degrees than males. About 23.3% of the total participants were never educated at all or only finished a primary education. About 32.8 percent of participants finished middle school education and 21.9 percent of participants finished high school education. Moreover, there are about 11.3% of the total participants who hold a college diploma and around 10.7% who finished a bachelor's or master's degree. In addition, from the table we can see that 71.9% of people who can speak fluent English finished college or university education while only 14.3% of people who cannot speak fluent English hold a college or university diploma, which means that fluent people have much higher education levels than the non-fluent ones.

Regarding marital status, 84.0% of the total participants are married and 16.0% are single or divorced or widowed, with relatively few more married females than males (difference of about 2 percentage points). However, only 68.0% of people who speak fluent English are married, but this may be related to the fact that they are younger than those who are not fluent. For the work-related variables, there are relatively more male participants who work on the Plant and machine operators and assemblers' occupations. The proportion of the female participants who are employed in the service and sale workers is higher (8 percentage points) than that of the male participants.

In addition, for the industry variables, more than one fifth of people who speak fluent English work in Social Service. As for the occupation variables, a large proportion of 39.2% of people who speak fluent English are employed in the Professional/Technician occupation, compared to only 9.6% of the non-fluent people who work in the same occupation.

Table 1. Mean Values of the Variables for Different Sub-samples

Mean Values	**Total**	**Male**	**Female**	**Fluent**	**Not fluent**	**Rural**	**Urban**	**Coastal**	**Central**
Mean of Log of Wages	9.868	10.039	9.615	10.579	9.758	9.250	10.255	10.391	9.552
A. English Speaking									
Fluent	0.134	0.117	0.161	-	-	0.038	0.195	0.235	0.073
Not Fluent	0.866	0.883	0.839	-	-	0.962	0.805	0.765	0.927
B. Gender									
Male	0.598	-	-	0.519	0.610	0.591	0.602	0.593	0.600
Female	0.402	-	-	0.481	0.390	0.409	0.398	0.407	0.400

C. Age									
Age	40.6	41.8	38.9	33.2	41.8	42.8	39.2	39.6	41.3
D. Education Level									
Lower Education	0.233	0.201	0.281	0.010	0.268	0.441	0.103	0.118	0.303
Middle	0.328	0.346	0.301	0.076	0.367	0.414	0.274	0.273	0.361
High	0.219	0.242	0.185	0.193	0.223	0.114	0.284	0.261	0.193
College	0.113	0.109	0.120	0.251	0.092	0.023	0.170	0.170	0.079
University	0.107	0.102	0.114	0.468	0.051	0.008	0.169	0.178	0.064
E. Marital Status									
Married	0.840	0.831	0.853	0.680	0.864	0.891	0.807	0.801	0.863
Not Married	0.160	0.169	0.147	0.320	0.136	0.109	0.193	0.199	0.137
F. Place of Residence									
Rural Area	0.385	0.381	0.391	0.110	0.427	-	-	0.195	0.499
Urban Area	0.615	0.619	0.609	0.890	0.573	-	-	0.805	0.501
G. Regions									
Coastal Area	0.377	0.374	0.381	0.660	0.333	0.191	0.493	-	-
Central Area	0.623	0.626	0.619	0.340	0.667	0.809	0.507	-	-
H. Industries									
Manufacturing	0.134	0.141	0.124	0.197	0.124	0.085	0.165	0.222	0.081
Construction	0.049	0.074	0.012	0.031	0.052	0.041	0.055	0.046	0.051
Transportation/Post	0.052	0.076	0.017	0.048	0.053	0.023	0.070	0.068	0.042
Whole Sale/Retail/Cafeteria	0.120	0.096	0.156	0.124	0.119	0.047	0.166	0.164	0.094
Social Service	0.126	0.110	0.150	0.227	0.111	0.037	0.182	0.168	0.101
Farm/Forestry/Fishing	0.009	0.011	0.006	0.005	0.010	0.009	0.010	0.008	0.010
Mining	0.007	0.010	0.002	0.005	0.007	0.009	0.006	0.010	0.005
Electricity/Gas/Water	0.011	0.013	0.007	0.011	0.010	0.005	0.014	0.012	0.010
Education/Science/Broadcasting	0.052	0.040	0.069	0.162	0.035	0.018	0.073	0.058	0.048
Government/Social Organization	0.046	0.050	0.041	0.063	0.044	0.014	0.066	0.051	0.043
I. Occupations									
Officials/Manager	0.067	0.077	0.050	0.120	0.058	0.023	0.094	0.100	0.046
Professional/Technician	0.136	0.124	0.153	0.392	0.096	0.034	0.199	0.186	0.105
Service/Sale Workers	0.146	0.113	0.195	0.140	0.147	0.058	0.200	0.184	0.123
Plant and Machine Operators/Assemblers	0.062	0.088	0.022	0.036	0.066	0.046	0.071	0.080	0.051
Observations	5,772	3,450	2,322	776	4,996	2,220	3,552	2,176	3,596

4. Econometric Model

In this section, I provide an overview of the econometric models estimated in this paper. I use ordinary least square regressions (OLS) with robust standard errors to analyze the crucial determinants of English language proficiency and the effect of English skills on wages in China. Thus, I ran two groups of regression models with different subsets of independent variables. In order to support the OLS regression results for English speaking proficiency which is a binary variable, I also estimated probit regressions that gave similar results. In addition, for the wage equation, instrumental variable regressions are also reported later in this paper.

4.1 Model one: English speaking

Firstly, I examine the determinants of ***English speaking***, which is a binary variable and ε_i is the error term. Thus, I analyze the determinants of English language proficiency by controlling successively for human capital and work-related variables.

Specification 1

$$Pr\,(English\ speaking = 1 \mid X) = \beta_0 + \beta_1 Gender + \varepsilon_i$$

The first specification only includes the *Gender* variable, which is a binary explanatory variable that is equal to one if *individual i* is male and equal to zero for a female. By estimating this simple model, I can obtain the gross effect of gender on being fluent in English speaking.

Specification 2

$$Pr\,(Englishspeaking = 1 \mid X)$$

$$= \beta_0 + \beta_1 Gender + \beta_2 Place\ of\ residence + \beta_3 Region + \varepsilon_i$$

For the second specification, I add the geographic variables *Place of residence* and *Region,* which have crucial impacts on English speaking. Both of them are dummy variables. For the *Place of residence*, it equals one if *individual i* live in a rural area and it equals zero otherwise. For the *Region*, it equals one if the province of *individual i* belongs to the coastal areas and it equals zero otherwise.

Specification 3

$$Pr\,(Englishspeaking = 1 \mid X)$$

$$= \beta_0 + \beta_1 Gender + \beta_2 Place\ of\ residence$$

$$+ \beta_3 Region + \beta_4\, Age + \beta_5 Education_i + \beta_6 Marital\ status + \varepsilon_i$$

In order to detect the effect of human capitals on English speaking, I add more independent variables: *Age*, *Education* and *Marital status*. Those variables were explained in the previous section.

Specification 4

$$Pr\,(Englishspeaking = 1 \mid X)$$

$$= \beta_0 + \beta_1 Gender + \beta_2 Place\ of\ residence$$

$$+ \beta_3 Region + \beta_4\, Age + \beta_5 Education_i + \beta_6 Marital\ status$$

$$+ \beta_7 Industry_i + \beta_8 Occupation_i + \varepsilon_i$$

The *Industry* and *Occupation* variables are added to the model to account the work-related factors. *Industry* is divided into ten binary variables and *Occupation* is divided into four binary variables, as defined in the previous section.

Specification 5

$$Pr\,(Englishspeaking = 1 \mid X)$$
$$= \beta_0 + \beta_1 Gender + \beta_2 Place\ of\ residence$$
$$+ \beta_3 Region + \beta_4\ Age + \beta_5 Education_i + \beta_6 Marital\ status$$
$$+ \beta_7 Industry_i + \beta_8 Occupation_i + \beta_9 Parental\ education + \varepsilon_i$$

Finally, I add the *Parental education* which is represented by how many years of education the father and mother received, which will be used later as instrumental variables in the wage equation. For this specification, I control for the full range of potential factors that can be included in this study.

4.2 Model two: Income

The main purpose of the second model is to estimate earnings differences between those who are English speaking fluent and those who are not. The dependent variable is $lnwage_i$, which is the natural logarithm of annual wages of individuals and μ_i is the error term. The first specification investigates the gross effect of English speaking proficiency on wages, and the other specifications add variables similarly to the models for English speaking fluency.

Specification 1

$$lnwage_i = \alpha_0 + \alpha_1 English\ speaking + \mu_i$$

Specification 2

$$lnwage_i = \alpha_0 + \alpha_1 English\ speaking + \alpha_2 Gender + \mu_i$$

Specification 3

$$lnwage_i = \alpha_0 + \alpha_1 English\ speaking + \alpha_2 Gender + \alpha_3 Place\ of\ residence + \alpha_4 Region + \mu_i$$

Specification 4

$$lnwage_i = \alpha_0 + \alpha_1 English\ speaking + \alpha_2 Gender + \alpha_3 Place\ of\ residence + \alpha_4 Region + \alpha_5 Age + \alpha_6 Education_i + \alpha_7 Marital\ status + \mu_i$$

Specification 5

$$lnwage_i = \alpha_0 + \alpha_1 English\ speaking + \alpha_2 Gender + \alpha_3 Place\ of\ residence + \alpha_4 Region + \alpha_5 Age + \alpha_6 Education_i + \alpha_7 Marital\ status + \alpha_8 Industry_i + \alpha_9 Occupation_i + \mu_i$$

5. Empirical Results

In this section, I present the empirical results separately for the English speaking and earnings OLS estimates discussed in the previous section. In addition, since the OLS estimates of the coefficient on fluency of English speaking may be biased due to the potential endogeneity of language skills, I estimate the economic returns to people who are fluent in English speaking using instrumental variables (IV) regressions.

5.1 Regression analysis of the determinants on English speaking

Table 2 shows the OLS results for the five specifications of the determinants of English speaking model for the total sample. The first column shows the raw differences by gender without any other control variables (specification 1). The marginal effect sign of gender is negative, and the coefficient value is -0.050 at the 1% significance level, which means that the probability of males to be fluent in English is five percentage point less than that of females.

The geographic variables are key variables, in specification 2, living in a rural area also presents a negative marginal effect that shows that rural residents are 11.4 percentage point less likely to be fluent in English than urban residents. The reason is that people who live in urban areas get easier access English knowledge, or people fluent in English move to cities for work opportunities. Furthermore, people who live in the coastal areas are 12.4 percentage points more likely to be fluent in English than in those who live in the central areas. This is because the coastal area is more developed than central area in China and then people who live in the coastal area can get more English knowledge.

After adding the human capital factors in specification 3, gender still keeps a negative marginal effect, but it decreases slightly by 2 percentage points. However, the coefficient of rural area changes its sign and shows a positive marginal effect (1.6 percentage points). In addition, the magnitude of the coastal area effect is reduced by

around 5 percentage points compared with specification 2. This is because part of the estimated effect in specification 2 captures differences in the human capital factors of age, education and marital status. Age and marital status show negative effects on English speaking. As people get older by 1 year, the probability of being fluent in English decreases by 0.5 percentage points. Similarly, people who are married have a 4.6 percentage point lower probability to speak fluent English than those who are not married. Comparing the education levels, the regression results show that people who hold college or university degrees are much more likely to speak fluent English than those who do not hold such degrees (23.1 and 49.8 percentage points respectively). This is because most of the English knowledge must be obtained from higher level of education, such as college and university.

In specification 4, after adding the work-related variables, all the signs of the previously discussed variables are consistent with the previous results and the marginal effects of those variables change slightly or remain steady. Compared to the reference industry of Social Service, the Manufacturing, Whole Sale/Retail/Cafeteria and Education/Science/ Broadcasting industries have positive marginal effects, while the other industries have negative marginal effects but most are not significant. As for the occupation variables, people who work in the Professional/Technician occupation have a 4.6 percentage point higher probability to speak fluent English than in the reference category.

Specification 5 controls for all potential factors and adds the parental education. The sample size is slightly smaller than the previous one because of missing values for parental education. The coefficient of the other variables changes little and the effects of one year increases in parental education are 0.3 and 0.5 percentage points respectively for the father and the mother. The coefficients are statistically significant and the positive signs suggest that it is relatively easier to acquire fluent English if the parents have received more education.

Table 2. Regression results of the determinants on English speaking by different control variables-OLS results

	Gender	**Gender, Geographic**	**Gender, Geographic, Human Capitals**	**Gender, Geographic, Human Capitals, Work**	**Gender, Geographic, Human Capitals, Work, Parents Education**
Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Gender (Ref. Female)					
male	-0.050***	-0.047***	-0.028***	-0.024**	-0.021*
	(0.010)	(0.009)	(0.008)	(0.009)	(0.009)
Geographic					
Place of Residence (Ref. Urban Area)					
Rural Area		-0.114***	0.016*	0.012	0.018*
		(0.008)	(0.007)	(0.008)	(0.008)
Region (Ref. Central Area)					
Coastal Area		0.124***	0.072***	0.070***	0.067***
		(0.010)	(0.009)	(0.009)	(0.009)
Human Capitals					
Age			-0.005***	-0.005***	-0.004***
			(0.000)	(0.000)	(0.000)
Education (Ref. Lower Education)					
Middle			0.001	0.003	-0.008
			(0.005)	(0.006)	(0.006)

High			0.082***	0.080***	0.059***
			(0.010)	(0.010)	(0.011)
College			0.231***	0.218***	0.185***
			(0.018)	(0.019)	(0.020)
University			0.498***	0.474***	0.436***
			(0.021)	(0.024)	(0.025)
Marital Status (Ref. Not Married)					
Married			-0.046***	-0.047***	-0.042**
			(0.013)	(0.013)	(0.013)
Industry (Ref. Social Service)					
Manufacturing				0.024	0.019
				(0.014)	(0.014)
Construction				-0.017	-0.015
				(0.015)	(0.015)
Transportation/Post				-0.022	-0.030
				(0.020)	(0.020)
Whole				0.009	0.009
Sale/Retail/Cafeteria				(0.015)	(0.016)
Farm/Forestry/Fishing				-0.067	-0.065
				(0.043)	(0.044)
Mining				-0.001	-0.0003
				(0.038)	(0.040)
Electricity/Gas/Water				-0.010	-0.004
				(0.039)	(0.038)
Education/Science/				0.036	0.035
Broadcasting				(0.030)	(0.030)
Government/Social				-0.079***	-0.078***
Organization				(0.023)	(0.023)
Occupation (Ref. Officials/Manager)					
Professional/Technician				0.046*	0.044*
				(0.020)	(0.020)
Service/Sale Workers				-0.025	-0.024
				(0.014)	(0.015)
Plant and Machine				-0.017	-0.015
Operators/Assemblers				(0.016)	(0.016)
Parents Education					
Father Education					0.003**
					(0.001)
Mother Education					0.005***
					(0.001)
Constant	0.167***	0.159***	0.247***	0.248***	0.177***

	(0.008)	(0.010)	(0.020)	(0.021)	(0.024)
R^2	0.005	0.078	0.304	0.312	0.318
Observations	5,516	5,516	5,516	5,516	5,377

Note: 1) Robust standard errors in parentheses.
2) * p<0.05; ** p<0.01; *** p<0.001

I further study the difference in the probability of speaking fluent English between males and females using specification 5. The results are shown in Table 3.

Firstly, regarding the geographic variables, compared with respondents in coastal areas, males are more likely to speak fluent English than those in central areas and the difference is large than that of females. In terms of the human capitals factors, the gaps are quite large between males and females. With respect to the education levels, compared to the reference category, males have a lower probability of speaking fluent English than females for all education levels. For example, for the education level of high school, females have an 11.1 percentage point higher probability of speaking fluent English than those with lower education, while males have only a 2.4 percentage point higher probability to be in such a situation. For people who have a college degree, females have a 21.8 percentage point higher probability of speaking fluent English than those who only have lower education, while for males it is only 15.7 percentage points.

The work-related factors show that there are some gaps between males and female for the probability of speaking fluent English in some special industries but there are less significant differences with respect to occupations. For the

Education/Science/Broadcasting industry, compared to the reference category, males have a 3.8 percentage (not significant) negative effect on English speaking, while females have a 9.4 percentage point positive effect on English speaking.

Table 3. Regression results of the determinants on English speaking by gender-OLS results

Variables	Male	Female
Geographic		
Place of Residence (Ref. Urban Area)		
Rural Area	0.013	0.029
	(0.009)	(0.015)
Region (Ref. Central Area)		
Coastal Area	0.080***	0.049**
	(0.012)	(0.015)
Human Capitals		
Age	-0.003***	-0.005***
	(0.001)	(0.001)
Education (Ref. Lower Education)		
Middle	-0.022**	0.004
	(0.007)	(0.010)
High	0.024*	0.111***
	(0.012)	(0.022)
College	0.157***	0.218***
	(0.025)	(0.033)
University	0.414***	0.463***
	(0.032)	(0.040)
Marital Status (Ref. Not Married)		
Married	-0.030	-0.061**
	(0.016)	(0.023)
Industry (Ref. Social service)		
Manufacturing	0.022	0.012
	(0.018)	(0.023)
Construction	-0.015	-0.043
	(0.016)	(0.051)
Transportation/Post	-0.025	-0.044
	(0.022)	(0.055)
Whole Sale/Retail/Cafeteria	0.008	0.004
	(0.020)	0.024
Farm/Forestry/Fishing	-0.088	-0.004
	(0.049)	(0.088)

Mining	0.011	-0.066
	(0.042)	(0.111)
Electricity/Gas/Water	0.021	-0.068
	(0.046)	(0.065)
Education/Science/Broadcasting	-0.038	0.094*
	(0.042)	(0.043)
Government/Social Organization	-0.077**	-0.081
	(0.027)	(0.043)
Occupation (Ref. Officials/Manager)		
Professional/Technician	0.050	0.037
	(0.026)	(0.031)
Service/Sale Workers	-0.024	-0.022
	(0.020)	(0.022)
Plant and Machine	-0.024	0.034
Operators/Assemblers	(0.017)	(0.047)
Parents Education		
Father Education	0.003**	0.003
	(0.001)	(0.002)
Mother Education	0.005**	0.006*
	(0.002)	(0.002)
Constant	0.131***	0.244***
	(0.028)	(0.045)
R^2	0.310	0.334
Observations	3,248	2,129

Note: 1) Robust standard errors in parentheses.
2) * p<0.05; ** p<0.01; *** p<0.001

5.2 OLS regression analysis of the effect of English speaking on earnings

Table 4 shows the OLS results for the five specifications of earning differences for the total sample. Similar to the previous section, the first column represents a simple regression that shows a raw earnings differences by controlling for English speaking without any other independent variables (specification 1); then I add more control variables into the next four specifications.

A simple regression that controls only for English speaking shows that those fluent in English speaking earn a huge 72.5% more than those not fluent in English. Furthermore, by adding the gender factor into specification 2, the coefficient of English speaking increases slightly by another 4 percentage points. Gender brings a positive effect on earnings which means that males earn 38.1% more than females in China.

However, after adding the geographic variables into the regression in specification 3, the magnitude of the English speaking coefficient reduces by 33.8 percentage points, while the magnitude of gender coefficient only decreases by 0.4 percentage points. This means that English speaking has lower effects on earnings when taking geographic factors into account. From the results, the earnings of rural residents are 63.6% less than those of the urban residents, and people who are from the coastal areas have 46.6% higher earnings than those from central areas.

I add human capital variables into specification 4. From the regression results, the effect of English speaking on earnings decreases dramatically; people who can speak fluent English can earn 9.5% more than those who cannot. The age has a small negative effect on earnings while education levels and marital status present significant positive effects on earnings.

Specification 5 shows all relevant variables that affect earnings. From the results, being fluent in English increases earnings by 8.9% in China. With respect to the education levels, people who have college and university degrees earn 59.1% and 79.0% more than people who only received lower education. In addition, married people earn 17.5% more than people who are not married. The mining industry brings the highest wage for people, which is 36.1% higher than the reference group (Social Service). Construction and Transportation/Post industries also provide higher earnings for people, at 29.7% and 23.7% respectively. With respect to the occupations, all of the occupations have positive effects on earnings compared to the reference group, especially for the Professional/Technician occupation (20.9% more than the reference group).

Table 4. Regression results on earnings by different control variables-OLS results

	English Speaking	**English Speaking, Gender**	**English Speaking, Gender, Geographic**	**English Speaking, Gender, Geographic, Human Capitals**	**English Speaking, Gender, Geographic, Human Capitals, Work**
Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
English Speaking (Ref. Not Fluent)					
Fluent	0.725***	0.764***	0.426***	0.095**	0.089**
	(0.030)	(0.029)	(0.028)	(0.030)	(0.030)
Gender (Ref. Female)					
Male		0.381***	0.377***	0.383***	0.360***
		(0.023)	(0.020)	(0.019)	(0.020)
Geographic					
Place of Residence (Ref. Urban Area)					
Rural Area			-0.636***	-0.422***	-0.352***
			(0.023)	(0.024)	(0.024)

Region (Ref. Central Area)			
Coastal Area	0.466***	0.424***	0.384***
	(0.021)	(0.020)	(0.020)
Human Capitals			
Age		-0.009***	-0.008***
		(0.001)	(0.001)
Education (Ref. Lower Education)			
Middle		0.188***	0.159***
		(0.029)	(0.028)
High		0.400***	0.351***
		(0.033)	(0.033)
College		0.620***	0.591***
		(0.037)	(0.039)
University		0.820***	0.790***
		(0.040)	(0.044)
Marital Status (Ref. Not Married)			
Married		0.191***	0.175***
		(0.028)	(0.027)
Industry (Ref. Social Service)			
Manufacturing			0.176***
			(0.028)
Construction			0.297***
			(0.045)
Transportation/Post			0.237***
			(0.042)
Whole			0.145***
Sale/Retail/Cafeteria			(0.034)
Farm/Forestry/Fishing			-0.020
			(0.096)
Mining			0.361***
			(0.088)
Electricity/Gas/Water			0.126
			(0.077)
Education/Science/Broa			-0.017
dcasting			(0.041)
Government/Social			-0.010
Organization			(0.035)
Occupation (Ref. Officials/Manager)			
Professional/Technician			0.209***

					(0.031)
Service/Sale Workers					0.156***
					(0.030)
Plant and Machine					0.149***
Operators/Assemblers					(0.037)
Constant	9.856***	9.622***	9.725***	9.611***	9.445***
	(0.013)	(0.019)	(0.021)	(0.055)	(0.055)
R^2	0.078	0.121	0.338	0.406	0.426
Observations	5,516	5,516	5,516	5,516	5,516

Note: 1) Robust standard errors in parentheses.
2) * p<0.05; ** p<0.01; *** p<0.001

Table 5 shows the earnings regressions separately for males and females. Males who are fluent in English earn 8.5% more than males who are not, while it is about the same for females who earn 8.9% more. Looking at the geographic factors, for example, males earn 41.0% more if they are in the coastal area while females earn 35.4% more, and there exist negative effect on earnings for both male and female rural residents.

Regarding the human capitals variables, it is noticeable that the effects of age and marital status factors for males are larger than those for females. For instance, married males earn 29.8% more than not married males while marital status has almost no effect for females. Considering the education level, we can find that differences between males and females is widened with the increases in the education level. Specifically, males earn 12.4% more than the reference group if they finished middle school and it is about the same for females who earn 18.7% more; however, for the university group, males only earn 67.8% more than reference group while females earn 94.0% more than reference group. The potential reason may be that

high-educated people are more likely to do some professional and precise work, providing an advantage to females.

Regarding to the work-related variables, there are some significant differences between males and females. For example, for the industry variable, the Education/Science/ Broadcasting and Government/Social Organization industries have a positive effect on male wages while these industries have negative effect on female wages. For the occupation variables, only the Service/Sale Workers have higher wage returns for females than for males.

Table 5. Regression results on earnings by gender-OLS results

Variables	**Male**	**Female**
English Speaking (Ref. Not Fluent)		
Fluent	0.085*	0.089*
	(0.041)	(0.044)
Geographic		
Place of Residence (Ref. Urban Area)		
Rural Area	-0.338***	-0.360***
	(0.031)	(0.039)
Region (Ref. Central Area)		
Coastal Area	0.410***	0.354***
	(0.026)	(0.032)
Human Capitals		
Age	-0.012***	0.0003
	(0.001)	(0.002)
Education (Ref. Lower Education)		
Middle	0.124***	0.187***
	(0.037)	(0.044)
High	0.286***	0.442***
	(0.043)	(0.052)

College	0.490***	0.717***
	(0.053)	(0.058)
University	0.678***	0.940***
	(0.058)	(0.066)
Marital Status (Ref. Not Married)		
Married	0.298***	-0.004
	(0.034)	(0.044)
Industry (Ref. Social service)		
Manufacturing	0.183***	0.181***
	(0.036)	(0.044)
Construction	0.266***	0.530***
	(0.047)	(0.137)
Transportation/Post	0.236***	0.214*
	(0.047)	(0.101)
Whole Sale/Retail/Cafeteria	0.153***	0.133**
	(0.046)	(0.050)
Farm/Forestry/Fishing	0.051	-0.169
	(0.115)	(0.170)
Mining	0.327**	0.421*
	(0.100)	(0.178)
Electricity/Gas/Water	0.139	0.101
	(0.088)	(0.162)
Education/Science/Broadcasting	0.015	-0.055
	(0.058)	(0.057)
Government/Social Organization	0.042	-0.082
	(0.045)	(0.057)
Occupation (Ref. Officials/Manager)		
Professional/Technician	0.211***	0.197***
	(0.041)	(0.045)
Service/Sale Workers	0.149***	0.172***
	(0.042)	(0.045)
Plant and Machine	0.152***	0.079
Operators/Assemblers	(0.041)	(0.084)
Constant	9.920***	9.258***
	(0.071)	(0.091)
R^2	0.393	0.439
Observations	3,332	2,184

Note: 1) Robust standard errors in parentheses.
2) * p<0.05; ** p<0.01; *** p<0.001

5.3 Instrumental variable regressions of the effect of English speaking on earnings

The OLS results only show correlations between English knowledge and earnings. To address the issue of potential causality between language and earnings, I now estimate instrumental variable equations. Table 6 shows the IV regression results on earnings with the same control variables as those of Table 4 for the total sample. Various instrumental variables have been utilized in the literature on language skills. I use the education of the respondents' father and mother as instruments. Parental education has been used in the human capital literature, but not in the previously cited studies on language skills. However, it was the best candidate that was available in my dataset. The rationale is that Chinese parents who have high education are likely to provide good education for their children, including learning new languages, which presumes that the years of parental education are correlated with their children's English fluency, but not directly with their children's earnings. In specification 5 of Table 3, the years of education of father and of mother are highly significant in the determination of English fluency. This satisfies the first condition for good instrumental variables, but the condition that parental education does not affect earnings cannot be verified directly.

From table 6, we can see that for the five regression results, the IV estimates of the English speaking are statistically significant and larger than OLS results, although there is a decreasing trend in the coefficient of English speaking when adding the

different control variables. Other studies also found that the IV coefficients of the effects of language skills on earnings are larger than the OLS ones. However, the coefficients in Table 6 are extremely large and imprecise, implying that English fluency would increase earnings by more than 500%. This is too large to be believed. A possible reason is that the instruments are not good since education of the parents may also affect their children's potential earnings since it provides them with skills other than language. For that reason, the IV analysis should be considered as inconclusive at this point and I do not trust much the results.

Table 6. Regression results on earnings by different control variables-IV results

	English Speaking	**English Speaking, Gender**	**English Speaking, Gender, Geographic**	**English Speaking, Gender, Geographic, Human Capitals**	**English Speaking, Gender, Geographic, Human Capitals, Work**
Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
English Speaking (Ref. Not Fluent)					
Fluent	2.534***	2.625***	1.721***	1.910***	1.810***
	(0.112)	(0.112)	(0.112)	(0.451)	(0.443)
Gender (Ref. Female)					
Male		0.473***	0.439***	0.433***	0.400***
		(0.030)	(0.024)	(0.027)	(0.027)
Geographic					
Place of Residence (Ref. Urban Area)					
Rural Area			-0.481***	-0.446***	-0.368***
			(0.028)	(0.030)	(0.029)
Region (Ref. Central Area)					
Coastal Area			0.306***	0.295***	0.265***
			(0.028)	(0.042)	(0.040)
Human Capitals					
Age				-0.001	0.001
				(0.002)	(0.002)

Education (Ref. Lower Education)					
Middle				0.196***	0.165***
				(0.034)	(0.033)
High				0.259***	0.221***
				(0.054)	(0.053)
College				0.208	0.224*
				(0.114)	(0.108)
University				-0.084	-0.024
				(0.231)	(0.218)
Marital Status (Ref. Not Married)					
Married				0.271***	0.251***
				(0.040)	(0.039)
Industry (Ref. Social Service)					
Manufacturing					0.133***
					(0.038)
Construction					0.326***
					(0.055)
Transportation/Post					0.280***
					(0.058)
Whole					0.126**
Sale/Retail/Cafeteria					(0.043)
Farm/Forestry/Fishing					0.094
					(0.122)
Mining					0.321*
					(0.138)
Electricity/Gas/Water					0.127
					(0.110)
Education/Science/Broa					-0.070
dcasting					(0.061)
Government/Social					0.110
Organization					(0.066)
Occupation (Ref. Officials/Manager)					
Professional/Technician					0.122**
					(0.046)
Service/Sale Workers					0.201***
					(0.040)
Plant and Machine					0.182***
Operators/Assemblers					(0.052)
Constant	9.614***	9.316***	9.519***	9.147***	9.004***
	(0.021)	(0.030)	(0.030)	(0.130)	(0.129)

R^2	0.140	0.143	0.166	0.310	0.315
Observations	5,377	5,377	5,377	5,377	5,377

Note: 1) Robust standard errors in parentheses.
2) * p<0.05; ** p<0.01; *** p<0.001

6. Conclusion

In this paper, I have examined the determinants of English speaking fluency and its effects on earnings in China's labour market by using data from the Chinese General Social Survey (CGSS) of 2013. The two outcomes are ***Englishspeaking*** and ***the natural logarithm of annual wages***. The paper provides different specifications to estimate the determinants and effects of English speaking. The main findings can be summarized as follows.

Firstly, the results show that speaking fluent English can increase wages in an OLS regression by 8.9% (specification 5 of Table 4). As for the determinants of English speaking, gender and marital status have negative relationships on the probability of being fluent in English. However, these variables have positive effects on earnings. There may exists gender discrimination in China's labour market since males earn 36% more than of females.

Secondly, the geographic variables are crucial to the two outcomes. People who live in the coastal area are more likely to be fluent in English and have higher earnings than those who live in other areas. This can be explained by the fact that in China the coastal area is more developed than the central area.

Furthermore, the positive marginal effect on the probability of being fluent in English and the economic returns also increase with education. This argument applies more to females than to males since the economic returns to education for females are larger than those for males.

Finally, to address the issue of causality, I have attempted to estimate the effect of English fluency on earnings with instrumental variables like some previous studies have done. I used the education of the respondents' father and mother as instruments. However, the IV coefficients are extremely large and imprecise, so the IV analysis is not inconclusive at this point. Better instruments should be used with other data sets in further research.

References

Albouy, D. (2008). The wage gap between Francophones and Anglophones: A Canadian perspective, 1970–2000. *Canadian Journal of Economics*, 41(4), 1211-1238.

Azam, M., & Chin, A., & Prakash, N. (2010). The Returns to English-Language Skills in India. *IDEAS Working Paper Series from RePEc*.

Bleakley, H., & Chin, A. (2004). Language skills and earnings: Evidence from childhood immigrants. *Review of Economics and Statistics*, 86(2), 481-496.

Carliner, G. (1981). Wage Differences by Language Group and the Market for Language Skills in Canada. *The Journal of Human Resources*, 16(3), 384–399.

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

Chiswick, B. R., & Miller, P. W. (1995). The Endogeneity between Language and Earnings: International Analyses. *Journal of Labor Economics*, 13(2), 246-288.

Christofides, L. N., & Swidinsky, R. (2010). The economic returns to the knowledge and use of a second official language: English in Quebec and French in the Rest-of-Canada. *Canadian Public Policy*, 36(2), 137-158.

Dustmann, C., & Fabbri, F. (2003). Language proficiency and labour market performance of immigrants in the UK. *The Economic Journal*, 113(489), 695-717.

Gao, W., & Smyth, R. (2010). Economic returns to speaking 'standard Mandarin' among migrants in China's urban labour market. *Economics of Education Review*, 30(2), 342-352.

Grenier, G. (1984). The Effects of Language Characteristics on the Wages of Hispanic-American Males. *The Journal of Human Resources*, 19(1), 35–52.

Grenier, G. (2015). The value of language skills. *IZA World of Labor*. 205.

Guo, Q., & Sun, W. (2014). Economic returns to English proficiency for college graduates in mainland China. *China Economic Review*, 30, 290-300.

Hayfron, J. (2001). Language training, language proficiency and earnings of immigrants in Norway. *Applied Economics*, 33(15), 1971-1979.

Liu, Q. (2014). Foreign language skills and income - Evidence from China's urban labor market. *NANKAI ECONOMIC STUDIES*, (03):137-153.

Maxwell, N. L. (2010) English Language and Low-Skilled Jobs: The Structure of Employment. *Industrial Relations: A Journal of Economy and Society*, 49(3), 457-465.

Shapiro, D. M., & Stelcner, M. (1997). Language and Earnings in Quebec: Trends over Twenty Years, 1970-1990. *Canadian Public Policy*, 23(2), 115–140.

Shields, M. A., & Price, S. W. (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.

Wang, C., & Wang, L. (2011). Language Skills and the Earnings Distribution Among Child Immigrants. *Industrial Relations*, 50(2), 297-322.

Wang, H., & Smyth, R., & Cheng, Z. (2017). The economic returns to proficiency in English in China. *China Economic Review*, 43, 91-104.

Yao, Y., & van Ours, J. C. (2015). Language skills and labor market performance of immigrants in the Netherlands. *Labour Economics*, 34, 76-85.