

How Education Level, Gender, and Social Network Correlate With Migrant Workers' Starting Income in China's Urban Cities^{*}

Dandan Chen

The George Washington University,
Washington, D.C., USA

Jingning Guan

University of Wisconsin-Madison,
Madison, USA

The research question for this study is how education level, gender, and social network affect migrant workers' starting income for their first job in China's urban cities. Our objective is to reveal the interplay of education level, gender, and social network in determining migrant workers' income, which are to the core of the current academic debate. The explanatory factor analysis was used to identify relevant variables, and a multivariate regression analysis was conducted to reveal the relationships among education level, gender, social network, and migrant workers' starting income in China's urban cities.

Keywords: migrant workers, education, gender, social network, income, urban development

Introduction

The year of 2010 marked the first time in China's history for the urban population to surpass the rural one (Bloomberg, 2012), witnessing the internal migrant population, made an unprecedented jump to involve more than 200 million people. By the end of 2013, this internal migrant population was reported to have grown up to 245 million, which equals at least one migrant in every six residents of this nation (The National Health and Family Planning Commission, 2014).

The term "migrant workers" has been frequently used without an unequivocal definition. However, it is widely adopted to generally refer to, in China's internal migration, the people who grew up in rural areas and have moved into urban cities to work. This population has caught much attention in the current migration studies relating to China due to the significant scale, diversity, and complexity of this group.

This research paper aims to provide some empirical evidence that would support existing policy suggestions urging migrant workers' working conditions to be enhanced. These policy suggestions have appealed for making such a difference by effectively improving migrant workers' education background and their soft skills like networking skills, regardless of their gender. The research question for this study is how education level, gender, and social network affect migrant workers' starting income in China's urban cities. Our objective is to reveal the interplay of education level, gender, and social network, which are to the core of the current academic debate.

^{*} The ideas and opinions expressed in the article are those of the authors and do not necessarily represent the views of UNESCO and do not commit the organization.

Dandan Chen, M.A., Graduate School of Education and Human Development, The George Washington University.

Jingning Guan, M.A., Department of Educational Leadership and Policy Analysis, University of Wisconsin-Madison.

Literature Review

According to Roberts (2001), factors like the education level, gender, age, marital status, company ownership, and place of origin could have an impact on migrant workers' income. Nevertheless, this study was based on the data on Shanghai's migrant population in 1993, which was a long time ago; besides, it only reflected the situation in one single city.

Similarly, a study conducted by Lu and Song (2006) found that the education, age, work duration, experience, hukou (household registration status), and company ownership are significant determinants of migrants' hourly wages, based on a survey conducted in 2005 in China's Tianjin City. Interestingly, this study singled out the gender, asserting the gender does not significantly affect migrant workers' hourly wages. In the meantime, contrarily, a study conducted by Wang (2005) on the gender-wage differential for migrant workers claimed females tend to get lower income than their male counterparts, so a gender disparity does exist. Unfortunately, the data sets for both studies are inaccessible to the public, which makes it very challenging to verify their findings or to help reach a consensus.

Recently, Wang, Lao, and Zhou (2014) asserted that the social network (i.e., the network of social connections each individual person possesses) is a significant determinant of migrant workers' income, using the survey data collected by the government in 2008. However, their study focused on some very specific aspects of the social network effect itself in lieu of the correlations between the social network and other determinants, and consequently how the social network gets to play with other social factors in determining migrant workers' income has yet to be unveiled.

Overall, through literature review, the education level and the social network have been found frequently confirmed as two critical determinants of migrant workers' income, while how the social network interacts with some other factors has yet to be clearly demonstrated. Also, the effect of the gender has been argued to be an important factor. This study thus focuses on the correlations among the education level, social network, gender, and migrant workers' income to bridge the literature gap and to support the current policymaking.

Methodology

Framework

This research was built upon the theoretical framework of Costello and Osborne (2005), which clarifies the myths in processing appropriate the exploratory factor analysis appropriately to select the most relevant variables for further analysis. Costello and Osborne's framework, simply put, asserted that, in the exploratory factor analysis, the maximum likelihood factor analysis is the best choice for data that are relatively normally distributed, whereas the principle axis factor analysis should be used instead when the assumption of normal distribution is severely violated. The best choice for researchers in retaining variables, according to Costello and Osborne, is the scree test, which retains variables above the break point where the scree curve flattens out. In addition, according to the two authors, oblique rotation should render a more accurate solution than orthogonal rotation in social science studies as some correlation among social factors would be considered in the former, which should be the case when analyzing social issues.

Analysis

Selection of data set. There are limited national data and restricted access to census data on migrant workers. Additionally, when investigating the factors that influence migrant workers' income, a survey should

be designed to collect data (Lu & Song, 2006; Han & Yuan, 2009). Taking into account of the findings from the literature review, the only open-access survey data found to be most relevant to this study is the data from the China Urban Labor Survey, collected in 2001 (China Survey Data Network [CSDN], 2008), by local offices of China's National Statistical Bureau. This data set included detailed information from questionnaire items on the individual and household demographics based on a sample of 8,109 individuals who aged above 16 and already joined the workforce in five large urban cities: Shanghai, Shenyang, Wuhan, Xi'an, and Fuzhou.

Selection of observations. Considering that the year of 1996 is the turning point of China's labor market and that the original data set includes a variable that asks whether the respondents started their first job before 1996, we decided to use this variable to preserve only the information provided by the respondents who started their first job after 1996. This study also aims to target those migrant workers who were "employed" rather than "unemployed" the time when the survey was conducted, and who were "hired by others" rather than "self-employed" for the first job, as suggested in the data set. Thus, only the observations that were coded as "employed by others" were preserved. In this way, a total of 634 from the original 2,998 observations were selected to be the observations for the further analysis in this study.

Selection of regression model. The regression analysis is conducted using fixed effects model. First, the following regression model is assumed:

$$Y = \beta X_i + \mu_i \quad (1)$$

Y_i is the monthly income migrant workers (i) earn for their first job at the beginning stage. X_i is the vector of independent variables we targeted, namely, the education level, social network, and gender. μ_i represents the stochastic error.

Second, the regression model in the following equation is considered:

$$Y = \beta X_i + n_i + e_i \quad (2)$$

The selected data for this study vary among different migrant workers, including the variance among the variables to be controlled in this study. n_i and e_i , replacing the error μ_i in (1), capture the variance and the error respectively.

There are two ways to properly deal with n_i above so as not to bias the coefficient for the variable X_i : fixed effects model and random effects model. The former is better for a data set that is not normally distributed, which is the case here, and therefore, the fixed effects model was adopted for this study.

Using the measure of mean for each of its variables except n_i , we got the regression model (3) as follows:

$$\bar{Y} = \beta \bar{X}_i + n_i + \bar{e}_i \quad (3)$$

By deducting the model (3) from the model (2), n_i gets deleted. Consequently, the potential bias from the between-entity variance can be resolved through this subtraction. The model we get through this process was a fixed effects model this study was looking for (as cited in Kim & Park, 2012).

To answer the research question for this study, the fixed effects regression model was specified as below:

$$\text{Income} = \beta_0 + \beta_1 \text{Education Level} + \beta_2 \text{Gender} + \beta_3 \text{Social Network} + i. \text{Variance Within Entities}$$

The Income was the dependent variable, and Education Level, Gender, and Social Network were the independent variables. "i. Variance Within Entities" represents the effects of some other important factors that were held constant using the fixed effects.

The null hypotheses considered for this study are: 1. The coefficient of Education Level equals zero

($\beta_1 = 0$); 2. The coefficient of Gender equals zero ($\beta_2 = 0$); and 3. The coefficient of Social Network equals zero ($\beta_3 = 0$) when holding all other variables as covered in “variance within entities” constant.

Selection of variables. Now that the fixed effects regression equation was clarified, the corresponding variables in the survey data set need to be identified for data processing. Given the available variables in the data set, only the migrant workers' beginning monthly income for their first job is found to the interest of this study, and thus this indicator would be used as a proxy indicator for migrant workers' starting income. The variables displaying migrant workers' education level and their gender, as already existing in the data set, were designated as the variables to be used for the analysis of Education Level and Gender. Please note, the variable for gender was found necessary to be used as a dummy variable, and so the original variable was recoded, using “0” for “male” and “1” for “female.”

Results and Discussion

In the original data set,¹ there are multiple items relating to the social network and various other factors which were expected to be studied. To integrate the most relevant variables in the data set into the fixed effects regression design, the Shapiro-Francia test was used in the first place to examine the assumption of normal distribution, only to find that these subset data were not normally distributed, as shown in Table 1.

Table 1

Results From Shapiro-Francia Test

Variable	Observations	W	V	Z	p-value
Year of birth (a1_a)	634	0.80318	82.048	10.706	0.0000
Education level (b1)	634	0.94564	22.659	7.581	0.0000
Years of stay (a14)	631	0.72922	112.394	11.469	0.0000

In this condition, following the theoretical framework from Costello and Osborne (2005), the principal axis factor analysis, scree test, and oblique rotation method were used to identify the relevant variables. Given the results, as shown in Table 2, only the variables that met these conditions were retained:

1. The value in uniqueness was smaller than 0.50;
2. The value in loading for factors was larger than 0.50;
3. The eigenvalue was above 1;
4. The variable lied above the break point of the scree plot in the scree test.

Table 2

Results From Principal Axis Factor Analysis

Variable		Factor 1	Factor 2	Uniqueness
	a10*	0.8703	<i>Note: Eigenvalue < 1</i>	0.1046
	a11*	0.8982		0.0756
	a15*	0.8564		0.1308
Hukou	a16*	0.8983		0.0911
	a37		0.6927	0.5107
	a38		0.7019	0.4910
	a40			0.9414

(Table 2 to be continued)

¹ The link to download the original data set (downloadable after registration): http://www.chinasurveycenter.org/csdn_en/DownloadChannel_new/detail.aspx?ClassID=4&DataID=9

Social network	a44_1			0.3031
	a44_2*	0.9001		0.1556
	a44_3*		0.8342	0.1957
	a44_4	0.3568		0.6618
	a45_1			0.2664
	a45_2*	0.8709		0.1878
	a45_3*		0.8976	0.0704
	a45_4			0.4067
Family	c102_11			0.2218
	c103_11*		0.8940	0.0855
	c104_11	0.8415		0.1471
	c105_11*	0.6712		0.1977
	c102_12			0.3189
	c103_12*		0.8729	0.1387
	c104_12*	0.8763		0.1428
	c105_12*	0.7202		0.2407

Notes. Blank represents factor loading smaller than 0.3; the symbol * indicates this variable is identified to be relevant and is to be used for the analysis.

The variables finally selected for the fixed effects regression analysis were listed below in Table 3.

Table 3

Selected Variables for the Fixed Effect Regression Analysis

Variables used in this study		Variable names in CSDN's survey data
Dependent variable	Monthly income at the beginning for the first job	f44_a_1: "What is your monthly income (including wage, monthly bonus, and subsidies)?" (only look at the first job at the beginning stage)
Independent variables to be targeted	Gender	a2: "Interviewers, please write down the sex of the respondent": 1. Male 2. Female
	Social network	a44_2: "How many relatives did you know when you just came to this city?" a45_2: "How many relatives do you know now in this city?" a44_3: "How many hometown fellows did you know when you just came to this city?" a45_3: "How many hometown fellows do you know now in this city?"
	Education level	b1: "What is your education level?" (code) _____
Independent variables to be held constant	Year of birth	a1_a: "When were you born?" Year _____ Month _____ (only look at the number for year)
	Marital status	a3: "What is your marriage status?" 1. Single 2. Married 3. Divorced 4. Widowed
	Hukou status	a10: "Did you have city hukou or rural hukou before you were 16 years old?" 1. City hukou 2. Rural hukou a11: "When did you mainly live before you were 16 years old?" 1. This city and this community 2. This city and other community 3. Other city 4. Other town or county 5. Other village 6. Others a15: "Which kind of hukou do you have now?" 1. City hukou 2. Rural hukou a16: "Where is your hukou now?" 1. This city this community 2. This city and other community 3. Other city 4. Other town or county 5. Other village 6. Others
	Years of staying in this city	a14: "How long have you been living in this city?" _____ year
	Family background	c103_11: "What is the type of the work unit for your father's primary job?" (code) _____

(Table 3 to be continued)

Independent variables to be held constant	Family background	c103_12: "What is the type of the work unit for your mother's primary job?" (code) _____ c104_11: "What is the industry of your father's primary job?" c104_12: "What is the industry of your mother's primary job?" c105_11: "What is your father's occupation?" 1. Peasants 2. Workers 3. Self-employed 4. Owner of private business 5. Administrators 6. Professionals 7. Administrators and professionals 8. Others c105_12: "What is your mother's occupation?" 1. Peasants 2. Workers 3. Self-employed 4. Owner of private business 5. Administrators 6. Professionals 7. Administrators and professionals 8. Others
Independent variables used to select observations	Whether you started the first job after 1996	a19: "Did you start your first job before 1996?" 1. Yes 2. No
	Whether you are employed now	a36: "What is your working status now?" (code) _____
	Whether you were employed by others for your first job	f4_1: "Are you self-employed or hired by others?" 1. Self-employed 2. Hired by others

Four variables were identified to represent Social Network, and one variable was designated for Education Level and Gender respectively. Considering the results from the literature review, the variables to be held constant were relating to the "year of birth," "marital status," "hukou status," "years of staying in this city," and "family background."

A fixed effects regression analysis was then conducted to estimate the effect of the education level, gender, and social network on migrant workers' beginning monthly income from their first job, holding other factors constant. The results in Table 4 showed that $R^2 = 0.59$, and the adjusted $R^2 = 0.28$. The p -value was smaller than 0.01 for the variable for Gender ("a2") and was smaller than 0.05 for the variable for Education Level ("b1"). These p -values showed that Gender is significant enough at the 0.01 level and Education Level significant at the 0.05 level to reject the null hypotheses 1 and 2. They suggested that female migrant workers tend to earn lower beginning monthly income than their male counterparts, whereas a higher education level would offset this effect. The p -values for the four variables representing Social Network ("a44_2," "a44_3," "a45_2," and "a45_3") were larger than 0.05, which failed to reject the null hypothesis 3, suggesting Social Network should not be a significant factor in this analysis.

Table 4

Results From the Two Regression Analyses

Starting monthly income (f44_a_1)	Coef.	Robust S.E.	t	$p > t $	95% conf. interval	
Gender (a2)	-188.4974**	54.4759	-3.46	0.001	-296.062	-80.933
Social network 1 (a44_2)	0.02453	11.6282	0.00	0.998	-22.936	22.985
Social network 2 (a45_2)	-3.1484	9.5713	-0.33	0.743	-22.047	15.751
Social network 3 (a44_3)	-0.7622	3.5287	-0.22	0.829	-7.730	6.205
Social network 4 (a45_3)	0.6964	3.4616	0.20	0.841	-6.139	7.532
Education level (b1)	21.8178*	10.7905	2.02	0.045	0.512	43.124
Intercept	1,077.642**	380.3137	2.83	0.005	326.6994	1,828.585

Notes. * $p < 0.05$; ** $p < 0.01$.

In the Stata output of this analysis, the F statistic and the overall significance value were missing, perhaps this was because too many variables were computed in this analysis. Nevertheless, an F -test for the joint significance of the identified variables for Education Level, Gender, and Social Network demonstrated the three

factors made a significant effect on migrant workers' beginning monthly income ($\text{Pro} > F = 0.0427 < 0.05$), when holding other factors constant.

Conclusion

This paper studied how education level, gender, and social network correlate with migrant workers' starting income, based on the survey data for China's five urban cities: Shanghai, Shenyang, Wuhan, Xi'an, and Fuzhou. A fixed effects regression model was built to guide the exploration:

$$\text{Income} = \beta_0 + \beta_1 \text{Education Level} + \beta_2 \text{Gender} + \beta_3 \text{Social Network} + i. \text{Variance Within Entities}$$

Here are the null hypotheses for this study:

1. The coefficient of Education Level equals zero ($\beta_1 = 0$);
2. The coefficient of Gender equals zero ($\beta_2 = 0$);
3. The coefficient of Social Network equals zero ($\beta_3 = 0$) when holding all other variables as covered in "variance within entities" constant.

In the fixed effect regression analysis, the variable for Gender was found significant enough at the 0.01 level and Education Level significant at the 0.05 level to reject the null hypotheses 1 and 2; the variables representing Social Network were larger than 0.05, which failed to reject the null hypothesis 3. These findings suggested that female migrant workers tend to earn lower beginning monthly income than their male counterparts, whereas a higher education level would offset this effect; the network of social connections one migrant worker possesses do not have a significant impact on this person's beginning monthly income.

The results of this study suggested that there is indeed a gender disparity among migrant workers in urban cities regarding their starting monthly income, with female migrant workers earning less than the male migrant workers. However, this income gap is not that obvious for male and female migrant workers who possess a higher level of education. The results also suggested that the social network should not be a significant determinant of migrant workers' starting monthly income.

References

- Bloomberg. (2012, January 17). China's urban population exceeds countryside for first time. *Bloomberg News*. Retrieved from <http://www.bloomberg.com/news/articles/2012-01-17/china-urban-population-exceeds-rural>
- China Survey Data Network (CSDN). (2008). *China Urban Labor Survey, 2001*. Retrieved from http://www.chinasurveycenter.org/CSDN_EN/DownloadChannel_new/detail.aspx?ClassID=4&DataID=9
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9.
- Han, L., & Yuan, X. (2009). The income increase of the rural migrant worker and its determinants. *Population Journal, 1*, 37-43.
- Kim, K., & Park, D. (2012). Impacts of urban economic factors on private tutoring industry. *Asia Pacific Education Review, 13*, 273-280.
- Lu, Z., & Song, S. (2006). Statistical analysis on micro-determinants that affects wages of rural-urban migrants. *Modern Finance and Economics, 10*(26), 77-81.
- Roberts, K. D. (2001). The determinants of occupational choice of labor migrants to Shanghai. *China Economic Review, 12*(1), 15-39.
- The National Health and Family Planning Commission. (2014). *Development report on China's migrant population*. The National Health and Family Planning Commission.

- Wang, C., Lao, H., & Zhou, X. (2014). The impact mechanism of social networks on Chinese rural-urban migrant workers' behavior and wages. *Economic and Labour Relations Review*, 25(2), 353-371.
- Wang, M. (2005). Zhongguo chengshi laodongli shichang shangde xingbie gongzi chayi (Gender wage differentials in China's urban labor market). *Economic Research*, 12, 35-44.
- World Bank. (2009). *China—From poor areas to poor people: China's evolving poverty reduction agenda—An assessment of poverty and inequality*. Washington, D.C.: World Bank.