Source-Country Income Inequality and Immigrant Self- Selection: A Difference-in-Difference Approach

Bollenbacher, Zak; Gihoon, Hong

To cite this article: Bollenbacher, Zak, and Hong Gihoon. "Source-Country Income Inequality and Immigrant Self- Selection: A Difference-in-Difference Approach." *Empirical Economics Letters*, vol. 13, no. 9, Sept. 2014, pp. 959–65.

Source-Country Income Inequality and Immigrant Self-Selection: A Difference-in-Difference Approach

Zak Bollenbacher* **and Gihoon Hong****

Indiana University South Bend, South Bend, IN 46634-7111, USA

Abstract. This paper revisits the income maximization hypothesis on immigrant self-selection. The traditional Roy model predicts that negative selection would arise when income distribution in the source country becomes more unequal relative to the destination country. However, the previous literature provides mixed evidence. Using data from the World Bank and U.S. Census between 2000 and 2010, we estimate a new specification that controls for country-specific fixed effects and unobserved global trends in immigration. The estimation results show that there exists no statistically significant relationship between immigrant skill composition and source-country income inequality, indicating that cross-sectional analysis suffers from omitted variable biases.

Keywords: Immigrant Self-selection, Income Inequality, Difference in Difference Estimation

JEL Classification Number: J11, J61, D63

1. Introduction

Perhaps, immigration is one of the most important ongoing political issues in the United States. As of 2014, a bipartisan immigration bill which involves a comprehensive immigration reform is being discussed at the Congress. Although political debates over immigration are typically centered on the type of immigrants who come to the U.S., the existing literature on immigration seems to offer diverse predictions on the pattern of immigrant self-selection. For example, Borjas (1987) argues that, holding the costs of migration constant, immigrants are negatively selected as the source country income distribution becomes more unequal relative to the destination country. On the other hand, Chiswick (1999) asserts that immigrants are positively selected because they are expected to be more able, ambitious, and entrepreneurial, given the significant costs of migration.

In this paper, we study the importance of source-country income inequality as a determinant of the self-selection of immigrants into the United States. Using data on income inequality and immigrant skill composition of 31 countries in 2010, we first test the Roy model prediction and confirm that source-country income inequality is positively associated with negative self-selection among immigrants, consistent with Borjas (1987). However, we argue that the results from cross-sectional analyses may be confounded

* Email: zbollenb@iusb.edu; ** Corresponding Author. Email: honggi@iusb.edu

when country-specific effects are not controlled. For example, it is conceivable that countries with high-income inequality are systematically different in other attributes from countries with low income inequality. To the extent that this unobserved country effects are correlated with the dependent variable, the OLS estimates are vulnerable to omitted variable bias. We address the issue by employing the difference-in-difference estimation method. Using data on the two census years of 2000 and 2010and removing unobserved country specific factors by first differencing, we show that the results are qualitatively different in that the OLS estimates are no longer significant, indicating that cross-sectional analysis suffers from omitted variable biases.

This paper proceeds as follows. Section 2 presents the economic theory underlying the Roy model and develops testable hypotheses. Section 3 discusses data and constructs the key variables. The results are presented in Section 4. We conclude in Section 5.

2. Economic Theory and Hypothesis Development

A few economic models drive the study of migration between countries. Borjas (1987) goes into great detail using the Roy model-the underpinning of this paper. The Roy model attempts to predict the skills of immigrants into a destination country based on the difference in returns-to-skill in the source and destination countries. In this sense, groups of workers are attempting to maximize their earning potential based on their skill level. For simplicity of analysis, two assumptions are made. First, although many other factors actually influence earning potential, only skill level is used in determining future earnings. Second, the skills of an individual will perfectly translate in both countries.

Within the Roy model framework, the greater income inequality in the destination country is associated with higher returns-to-skill, resulting in positive selection of immigrants. This implies that immigrants are positively selected from the upper tail of income distribution in the source country. On the other hand, negative selection arises when the source country income distribution becomes more unequal relative to the destination country, implying that workers with fewer skills will have a higher earning potential in the destination country than in the source country. Using the predictions of the Roy model, Borjas (1987) shows that there is negative correlation between source-country income inequality and the earnings of immigrants in the destination country. This leads to the following hypothesis:

> *Hypothesis 1: Higher income inequality of a source-country leads to negative selectivity of migrants from the country.*

Therefore, the most straightforward approach to test the hypothesis would be to estimate the following regression equation:

$$college_i = \alpha + \beta(gini_i) + \varepsilon_i,\ \varepsilon_i \sim i.i.d\ Normal(0, \sigma^2) \tag{1}$$

where $college_i$ is the share of college graduate workers among the immigrants from country i, which is a proxy for skill level; $gini_i$ is the Gini coefficient for country i; ε_i represents a random shock that is not explained by other variables in the model; α, β, and σ are model parameters to be estimated.A statistically significant negative value of β would support Hypothesis 1.

However, one concern arises in estimating Equation (1) using cross-sectional analysis due to non-random assignment of sample countries. For example, it can be reasonably assumed that availability of or access to higher education is fundamentally different across countries depending on cultural factors or the state of economic development. To the degree that the unobserved country effects shaping the mean skill level of the labor force, as measured by educational attainment, are also correlated with income dispersion within the country, the Ordinary Least Squares (OLS) estimates would be biased. In order to account for the sample selection issue, we propose to employ a difference-in-difference estimation approach. By creating an average outcome for the treatment and the control group both before and after treatment, the double difference estimator controls for unobserved country effects as well as any trends in immigration common to all the countries in the sample.

> *Hypothesis 1a: After accounting for unobserved country-specific effects, an increase in income inequality within a source country would lead to more negative selectivity of immigrants out of the country.*

Then the estimation equation in (1) is modified as follows:

$$\begin{gathered} college_i^{2010} - college_i^{2000} = \alpha + \beta(gini_i^{2010} - gini_i^{2000}) + \varepsilon_i, \\ \varepsilon_i \sim i.i.d\ Normal(0, \sigma^2) \end{gathered} \tag{2}$$

Following with the predictions of the Roy model, we expect the coefficient of the difference in Gini Coefficients to be negative. This implies that as income inequality increases, the fraction of college graduates from the source-country will decrease.

3. Data

As a measure of income inequality of the source countries, the Gini indices are constructed from the World Development Indicators published by the World Bank.[1] Unfortunately, we find that many countries do not have Gini index reported often or predictably. In addition,

[1] The data are publicly available at http://data.worldbank.org/data-catalog/world-development-indicators.

the data points are sporadic for many countries. In an effort to get as many countries as possible in the sample, we use three-year averages at 1999-2001 and 2009-2011 respectively. Any countries with no data point within the three-year time span are deleted. As a result, 31 countries are included in the final sample. Table 1 presents the 31 countries along with the Gini coefficients for the two census years. Although the sample size is relatively small, we find that there is much variation in the sample in terms of income inequality, in that 10 out of the 31 countries experienced an increase in the Gini index.

Table 1: List of Countries Included in the Sample

Country	Gini 2000	Gini 2010	Country	Gini 2000	Gini 2010
Mexico	51.87	47.16	Hungary	27.34	31.20
Belize	54.87	56.95	Poland	32.95	33.52
Costa Rica	48.37	50.73	Romania	30.41	27.22
El Salvado	52.92	48.33	Lithuania	32.13	37.60
Panama	57.30	51.98	Moldavia	38.97	33.53
Dominican	51.22	48.03	Armenia	36.12	31.30
Argentina	51.43	45.31	China	39.23	42.06
Bolivia	59.74	56.30	Cambodia	42.05	36.03
Brazil	59.96	54.69	Indonesia	28.99	36.86
Chile	55.26	52.06	Malaysia	43.70	46.21
Colombia	58.48	56.29	Philippine	46.09	42.98
Ecuador	58.36	49.35	Thailand	42.97	39.70
Paraguay	56.55	51.73	Bangladesh	33.46	32.12
Peru	53.82	48.60	Pakistan	33.02	30.00
Uruguay	45.28	45.80	Uganda	43.07	44.30
Venezuela	47.50	44.80			

As the dependent variable in the regression equations, we consider the fraction of college graduate among U.S. immigrants. In constructing the dependent variable for estimation, we extract the 1 percent ACS samples from the 2000 and 2010 census.[2] In order to compute the share of college graduated immigrants for each country, we first create a dummy variable which indicates whether an individual has a college degree.[3] Then,

[2]The data are publicly available at https://usa.ipums.org/usa/

[3]The U.S. Census datareport respondents' educational attainment in EDUC. We consider any individual with the value of EDUC greater than 100 ashaving a bachelor degree.

computing the mean of the dummy variable for each country produces the proportion of college graduates out of the entire labor force from a given country, which is used as the dependent variable in the regressions.

We merge the Gini coefficients for both the 2000 and 2010 samples with the sample compiled from the U.S. Census. In order to only count people who could reasonably have the opportunity to gain an education, all data points with ages greater than 65 and less than 16 are deleted from the sample. As a sensitivity check, we also consider a sample with a different threshold age (22 instead of 16) in the regression analysis. Table 2 presents descriptive statistics associated with the key variables used in the regressions. It shows the wide range of values assigned to each country for both the mean education level of immigrants and the Gini index.

Table 2: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Country	31	37981.61	12329.49	20000	60055
College-graduate share in 2000	31	0.3462	0.1671	0.0385	0.7143
College-graduate share in 2010	31	0.3387	0.1368	0.0570	0.5790
Gini index in 2010	31	43.6357	8.6820	27.22	56.95
Gini index in 2000	31	45.5933	10.1648	27.3367	59.955

4. Results

Table 3 presents the main results from estimating Equations (1) and (2). First and secondly, a standard regression model is estimated for each of the 2000 and 2010 years, respectively. The first two columns shows that consistent with Hypothesis 1, source-country income inequality and the fraction of college graduates are negatively correlated at -0.007 and -0.006, respectively. These values indicate that for each 1% increase in the Gini index, the share of college graduates among immigrants from the source country increases by 0.7%in 2000 and 0.6% in 2010, respectively. Further, the estimated coefficients are statistically significant with t-statistics of 2.84 and 2.65.

However, one concern arises when making a causal interpretation on the estimates in Columns (1) and (2) in Table 3. Due to the cross-sectional nature of the econometric analysis, the OLS estimation approach fails to account for unobserved country-specific effects. For example, it is conceivable that countries have fundamentally different dispositions to education or access to education. To the extent that the unobserved factors affecting the dependent variable are also correlated with source-country income inequality, the OLS estimates would be subject to omitted variable bias. To account for fundamental differences in unobserved country characteristics, we use the difference in difference estimation method as shown in equation (2).The results in Column (3) show that the change in income inequality in the source country is no longer statistically significant, with a t-statistic of 0.26.As a result, no statistical inferences can be made about the importance of income inequality on selection from source to destination countries.

Table 3: Affect of GINI of Selection (OLS Model)

	(1)	(2)	(3)	(4)
Variables	**Standard 2000**	**Standard 2010**	**Difference**	**Only Age >22**
Gini	-.0077**	-.0069**	-.0010	.0011
R-Squared	.2179	.1951	.0023	.0025
Adj R-Squared	.1909	.1673	-.0321	-.0318
Observations	31	31	31	31

Note: 1% ACS IPUMS Data. ** denotes statistical significance at the 1% level.

The estimation results indicate that when countries are systematically different along some unobserved attributes, failing to account for omitted variables leads to qualitatively biased results. In the context of immigrant self-selection, potentially important determinants of immigrant skill composition would also include distance between the two countries and the size of network at the destination as well as cultural or historical link between the source and destination countries. Although it is beyond the scope of this paper to provide a complete analysis of immigrant self-selection, we believe that the results presented here emphasize the importance of a correctly specified model in an economic analysis.

Finally, as a robustness check, we use a different threshold for the age value to beincluded in the sample by deleting all immigrants less than 22 years old rather than 16. The thought process for this isthat between 16 and 22, immigrants still have an option to pursue college education, and therefore classifying immigrant skill based on the highest degree completed may not be valid. The results shown in Column (4) indicate the estimated coefficient of source-country income inequality is still statistically insignificant.

5. Conclusion

In this paper, we have studied the impacts of source-country income distribution on self-selection of immigrants in the U.S. Accounting for unobserved country characteristics and global trends in immigration, we find that there exists no statistically significant relationship between the pattern of immigrant self-selection and source-country income inequality. The results suggest that a careful econometric design is needed in immigration research when sample selection is not completely random, in which case cross-sectional analysis suffers from omitted variable biases.

The estimation results also suggest that other factors that are non-accessible for the use of this paper may better explain the decision to immigrate. For example, cultural forces or the cost of moving as a function of distance between the two countries as well as the size of immigrant network at the destination may prove to be important determinants of immigrant self-selection (McKenzie and Rapoport, 2010).

References

Akee, R, 2010, Who Leaves? Deciphering Immigrant Self-Selection from a Developing Country, Economic Development & Cultural Change, 58, 323-344.

Borjas, G. J., 1987, Self-Selection and the Earnings of Immigrants, American Economic Review 77, 531-553.

Borjas, G. J., 1990, Self-Selection and the Earnings of Immigrants: Reply, American Economic Review80, 305-308.

Chiquiar, D. and Hanson, G.H., 2005, International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and The United States, The Journal of Political Economy 2, 239-281.

Chiswick, B. R., 1999, Are Immigrants Favorably Self-Selected? American Economic Review 89, 181-185.

Clark, X., Hatton, T. and Williamson, J., 2007, Explaining US immigration,1971–1998,The Review of Economics and Statistics 89, 359–373.

Liebig, T. and Sousa-Poza, A., 2004, Migration, Self-Selection and Income Inequality: An International Analysis, *Kyklos* 57, 125-146.

Mayda, A. M., 2010, International migration: a panel data analysis of the determinants of bilateral flows, Journal of Population Economics 23, 1249–1274.

McKenzie, D. and Rapoport, H., 2010, Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks, The Review of Economics and Statistics, 92, 811-821.