Urbanization and Returns to Human Capital Investment

Sung Soo Lim

***Abstract*—Many developing economies are undergoing urbanization processes through rural-urban migration. One of stylized facts in migration is that education is positively correlated with migration. This study investigates and tests the underlying cause of the positive education effects. In particular, using Heckman's two step procedure, this study finds differential returns to human capital investment across education groups in urban areas of Indonesia. The results show that more educated people have a greater incentive to migrate to urban areas.**

***Index Terms*—Urbanization, migration, returns to education, human capital.**

I. Introduction

Many developing economies are undergoing urbanization processes through rural-urban migration. Empirical studies of migration provide a number of stylized facts, one of which is that education will tend to increase migration. So why is migration more likely for more educated people? In the neoclassical framework, migration occurs if a potential migrant's expected earnings in the city net of her expected earnings in the village and net of the direct costs of migration are strictly positive. In fact, the returns to higher education are generally low in the rural areas of most developing countries. Thus, it may be necessary for rural residents to move to a Metropolitan area to reap the monetary rewards of higher education. An alternative, yet similar view is that cities are places for accumulating human capital [1], [2]. According to this view, benefits of cities accrue over time and the urban wage premium is a wage growth effect rather than a wage level effect. The positive effect of education on migration may then be explained by greater income growth rates of more educated groups in urban areas. Empirical studies testing these explanations are rare, although the identification of true causes has significant importance for policy makers [3], [4].

To investigate and test the validity of these explanations, one may compare returns to education between urban and rural areas. Higher returns to education or higher income growth rates in urban areas would support the explanations. For example, between 2000 and 2007, the growth of real income of male college graduates in cities was much higher than that found in rural areas of Indonesia (Table I). In particular, the mean income growth of the former is approximately 40 percentage points higher thanlatter for males and 15 percentage points higher for female workers with college education in urban areas. However, there are two concerns with this test. First, high urban wages foundin many countries may induce low-skill workers to migrate to low-productivity city jobs and unemployment, who might otherwise be more productive in rural sectors [3], [5]. If this is true, then estimations of returns to education may be misleading when one simply compares urban and rural wages. Another concern with this test is that it is based on an assumption that migrants correctly expect their earnings post migration. However, perceived returns and actual returns to education may differ. For instance, Jensen [6] finds that there is discrepancy between the measured returns to schooling in the Dominican Republic and the returns perceived by students. Likewise, returns to education perceived by potential migrants may be different from measured returns of the sample.

To circumvent these issues, this study compares, using Heckman's [7], [8] two step procedure, migrants' post migration wages with what the migrants would have earned had they not moved, and tests whether migrants and stayers make rational migration decisions consistent with the neoclassical explanations of education effects. Furthermore, this study investigates the growth rate of migrants' post migration wages and tests if it is consistent with migration growth rates of each educational group. The results of this study provide strong evidence for higher returns to human capital investment in urban areas, supporting the explanation that migrants are motivated by higher income gains in urban areas.

II. Survey Data and Characteristics of the Sample Population

Indonesia, the fourth most populous nation in the world, has enjoyed rapid economic growth over the past three decades. Furthermore, since 1980Indonesia has rapidly urbanized, exceeding many other developing countries in urbanization rates. This paper uses three waves (1993, 2000, and 2007) of panel data from the Indonesian Family Life Survey (IFLS), a high-quality panel survey of individuals, households, and communities. The survey contains information from more than 10,000 households representing about 83% of the Indonesian population, those who live in the 13 most populous of the nation's 26 provinces. The IFLS collected a broad array of demographic, socioeconomic, and health information on individuals, households, and communities while capturing the cultural and socioeconomic diversity of Indonesia. The survey data also contain detailed migration information. In particular, it contains information on each subsequent location of residence for all moves that crossed a village boundary and lasted for longer than 6 months. One unique aspect of this data set is that it contains information on attitudes towards risk derived from a series of questions asked in the 2007 wave of the IFLS. To measure risk attitudes, respondents were asked to choose between a

Manuscript received November 13, 2014; revised January 24, 2015.

Sung Soo Lim is with the School of Business and Administration, American University in Dubai, Media City, UAE (e-mail: slim@aud.edu).

sure amount of money and lotteries of various types. Based on their answers to these questions, the respondents were classified into four groups according to their preferences with respect to risky lotteries. This direct measure of risk attitudes allows one to control for the effects of risk attitudes when analyzing the effects of uncertainty on migration decisions. In addition to the IFLS data, this study utilizes information on travel distance by a car to major cities in Indonesia obtained via Google Maps. The distance information was collected at the kecamatan level, a subdivision of a city.

TABLE I: MEAN REAL INCOME AND INCOME GROWTH OF MALES AND FEMALES (2000 AND 2007)

			Real Income	% change 2000-2007	Real Income	% change 2000-2007	Real Income	% change 2000-2007
Year	Region	Gender	college graduates		high school		less than high school	
2007	Urban	Male	8501	69%	3954	40%	3094	27%
2000			5019		2831		2428	
2007	Rural		5622	30%	3551	47%	2331	51%
2000			4327		2408		1539	
2007	Urban	Female	5676	62%	3282	72%	2183	56%
2000			3514		1898		1339	
2007	Rural		4062	46%	3201	162%	1609	120%
2000			2775		1219		729	

Source: IFLS 3 & 4, Unit: thousand Rupiah

Due to data limitations on less populated islands, regression analyses are conducted only for the sample population who live in Sumatra or Java, which includes approximately 70% of the households in the IFLS. The working sample consists of 6696 individuals in the 18 to 65 age bracket in 2000, for whom complete internal migration and employment histories for the 2000 and 2007 as well as information on other variables of interest were available. Migrants in this study are defined to be individuals between the ages of 18 and 65 who crossed a village boundary to move to urban areas for the first time in their lives between the years 2000 and 2007 and who stayed in the destination for longer than 6 months.

III. METHODOLOGY

Heckman's self-selection model recognizes the existence of measured and unmeasured heterogeneous skills among agents. Denoting by y_m^i the log wages of migrants and by y_n^i those of nonmigrants, the two earnings equations can be written as

$$y_m^i = \gamma_m X_m^{i'} + \varepsilon_m^i \quad (1)$$

$$y_n^i = \gamma_n X_n^{i'} + \varepsilon_n^i, \quad (2)$$

where γ_m and γ_n are vectors of exogenous variables. When standard Heckman [7], [8] sample selection bias correction formulae are used, the expected values of the wage data for migrant (i) and nonmigrant (j) are

$$E\left(y_m^i|\mathrm{G_i} > 0\right) = \gamma_m X_m^{i'} + \theta_m \mu_m^i, \quad (3)$$

$$E\left(y_n^j|\mathrm{G_j} \le 0\right) = \gamma_n X_n^{j'} + \theta_n \mu_n^j, \quad (4)$$

where $\mathrm{G_i}$ is a latent variable that represents gains to migration, and

$$\theta_m = \frac{\sigma_{\eta m}}{\sigma_\eta}, \ \theta_n = \frac{\sigma_{\eta n}}{\sigma_\eta} \quad (5)$$

Unbiased estimates of γ_m and γ_n can be obtained by OLS only when $\theta_m = 0$ and $\theta_n = 0$, respectively. The selectivity control variables μ_m^i and μ_n^j are denoted as follows.

$$\mu_m^i = \frac{f(\gamma' \mathrm{Z_i})}{F(\gamma' \mathrm{Z_i})}, \quad \mu_m^i = \frac{-f(\gamma' \mathrm{Z_i})}{1 - F(\gamma' \mathrm{Z_i})}, \quad (6)$$

where $F(\gamma' \mathrm{Z_i})$ is the cumulative density function such that $F(\gamma' \mathrm{Z_i}) = Pr(M_i > 0)$ and $f(\gamma' Z_i)$ is the density of the normal distribution.Predicting the wage gains to migration for each education group, however, requires the counterfactual earnings of migrants of each group. That is, one needs to measure what migrants would have earned had they not moved.For migrant *i*, the wage gain equals the expected wage in the migration regime less that in the nonmigration regime, conditional on the fact that the agent is a migrant. Following Davanzo and Hosek [9] and Tunali [10], I obtain the counterfactual wage by means of the consistent estimates of γ_n, θ_n, and μ_n:

$$E\left(y_n^i|\mathrm{G_i} > 0\right) = \gamma_n X_n^{i'} + \theta_n \mu_n^i \quad (7)$$

Thus, a migrant's expected wage gain from migration equals

$$E\left(y_m^i|\mathrm{G_i} > 0\right) - E\left(y_n^i|\mathrm{G_i} > 0\right), \quad (8)$$

which compares the migrant's predicted wage in the migration regime with her predicted counterfactual wage in the nonmigration regime. A nonmigrant's expected wage gain from staying can also be estimated by

$$E\left(y_n^i|\mathrm{G_i} \le 0\right) - E\left(y_m^i|\mathrm{G_i} \le 0\right), \quad (9)$$

which compares the nonmigrant's predicted wage in the nonmigration regime with her predicted counterfactual wage in the migration regime. Comparing the expected gain for each educational group, one can innocuously test if migrants are motivated by expected income gains.

TABLE II: COMPARISON OF EARNINGS BETWEEN MIGRATNTS AND STAYERS IN 2007

	Model 1	Model 2	Model 3	Model 4	Mode 5	Model 6
Dependent Var	Income 2007 Migrants	Heckman 2007	Heckman 2000	Income 2007 Stayers	Heckman 2007	Heckman 2000
Personal						
Educational						
middle school	0.135	0.168	0.387	0.278***	0.297***	0.321***
	(0.796)	(1.046)	(1.560)	(5.405)	(8.139)	(8.241)
high school	0.709***	0.770***	0.640***	0.662***	0.706***	0.676***
	(4.808)	(5.010)	(2.691)	(11.558)	(20.185)	(17.361)
university	1.372***	1.438***	1.130***	1.098***	1.203***	1.097***
	(8.270)	(8.295)	(4.222)	(13.460)	(25.005)	(19.636)
Demographic						
age	0.042	0.033	0.060	0.072***	0.067***	0.113***
	(1.158)	(0.903)	(1.015)	(6.066)	(7.834)	(11.140)
age squared	0.000	0.000	0.000	-0.001***	-0.001***	-0.001***
	(-0.677)	(-0.538)	(-0.504)	(-6.496)	(-8.314)	(-10.394)
female	-0.297***	-0.293***	0.225	-0.612***	-0.505***	-0.470***
	(-2.670)	(-2.642)	(1.360)	(-15.484)	(-19.682)	(-16.076)
Vocational						
self employed	0.206	0.204	0.092	0.366***	0.356***	0.268***
	(0.988)	(0.997)	(0.350)	(7.005)	(8.750)	(6.831)
government	0.398*	0.404*	0.363	1.154***	0.982***	0.705***
	(1.687)	(1.736)	(1.071)	(14.387)	(17.978)	(11.266)
private	0.189	0.217	0.215	0.543***	0.554***	0.445***
	(1.037)	(1.201)	(0.859)	(9.284)	(13.348)	(10.486)
Behavioral						
risk attitude 2	-0.03	-0.032	-0.219	0.005	0.053*	0.017
	(-0.263)	(-0.285)	(-1.219)	(0.106)	(1.772)	(0.504)
risk attitude 3	0.210*	0.209*	-0.028	0.080*	0.109***	0.075**
	(1.716)	(1.744)	(-0.162)	(1.879)	(3.599)	(2.198)
risk attitude 4	-0.015	0.003	0.036	0.209**	0.325***	0.193**
	(-0.049)	(0.010)	(0.108)	(2.214)	(5.043)	(2.418)
Village						
distance (ln)	-0.119***	-0.131***	-0.187***	-0.106***	-0.152***	-0.112***
	(-3.174)	(-3.238)	(-2.705)	(-5.157)	(-13.435)	(-9.005)
Island						
Java island	0.105	0.099	0.301*	-0.300***	-0.269***	-0.186***
	(0.915)	(0.866)	(1.698)	(-7.261)	(-8.650)	(-5.474)
constant	14.878***	14.725***	11.184***	14.554***	14.871***	12.826***
	(27.734)	(27.769)	(12.776)	(69.407)	(99.590)	(67.531)
R-2	.32			.31		
Inverse mills ratio		.162	1.033***		-.683	-.462
Standard Error		.158	.298		.066	.102

* $p<0.05$, ** $p<0.01$, *** $p<0.001$, All standard errors were adjusted for within cluster dependence.

IV. RESULTS

Table II shows that results of earnings regressions estimated by the Heckman two step procedure. The regression results of the selection equations are presented in the Appendix II. As exclusion restrictions, the variable "married" is included for probit regressions for both migrants and stayers. In addition, the variable "land" is included in the selection model of migrants. Intuitively, urban income is less likely to be affected by the possession of land in rural areas while it is more likely to affect the migration propensities. The regression results are found to be consistent with this conjecture.

The results show that the coefficients on the education dummies have all expected signs and magnitudes, and the results are consistent across all models. Assuming the coefficients on the education dummies properly estimate differential returns to education for each education group in urban and rural areas, the results appear to be consistent with the hypothesis that college graduates have greater incentives to migrate to urban areas in Indonesia between the years 2000 and 2007. In particular, migrants with college education in 2007 have roughly 24% greater returns to education than those in rural areas (Model 2 and Model 4). The gain in returns is comparatively greater than migrants with high school education, and the opposite is true for migrants with middle school education. The results are consistent when education dummies are replaced with education years although the results are not shown in this study. On the other hand, as expected, the income gains are larger when the selection bias is not controlled (Model 1 and 3), implying the existence of a selection bias.

One concern with this test is that it is based on one period of time (2000 -2007) and may not justify the assumption that migrants have correct expectations on their earnings *post* migration. Ideally, to justify the assumption, one would need data showing historical variations in returns to education as well as changes in migration propensities.This study instead utilizes three waves of survey data and tests if there exist greater growth rates in returns to higher education for migrants, given that the increase in migration propensities between 1993-2000 and 2000-2007 is greater for more educated people. Table 3 shows supportive results.

A comparison of migration propensities between two

periods (1993-2000 and 2000-2007) indicates that the growth in migration propensities is greater for more educated groups. It shows that one year of education is associated with a marginal increase of migration propensities by 0.2% between 2000 and 2007, which is twice the increase in propensities between 1993 and 2000. On the other hand, the growth in returns to education appears to be lower between 2000 and 2007, compared with the growth rate between 1993and 2000. The results are consistent at all education levels, regardless of migrants or stayers. The decrease, however, appears to be much greater for rural stayers than migrants to urban areas. In other words, there is *relatively* a greater incentive to migrate to urban areas for more educated people. Results are consistent whether the estimation models use categorical variables or continuous variables for education levels. Again, the results support the hypothesis that there is a greater incentive for migration for more educated groups.

V. Conclusion

Overall, the results of this study support the conjecture that there are greater economic gains from migration for more educated people. In particular, comparing migrants' post wages with their counterfactual wages in rural areas, this study finds significantly larger income gains for more educated groups. The results of this study also support the view that cities are places for accumulating human capital [1], [2]. If we adopt this perspective, then the coefficients of education variables would be returns on human capital investment for each educational group. For instance, each additional year of education is associated with an accumulation of a 5 times greater earnings premium for migrants, compared with rural stayers, during the period of 2000 and 2007. More intuitively, for instance, educated workers may particularly benefit from skill spillovers in urban areas, where human capital accumulation is accelerated by a greater share of educated workers (e.g. [2], [10]). On the other hand, this study find no evidence economically justifying the rural-urban migration of less educated people. In fact, the returns appear to be greater in rural areas. This implies that policy makers need to pay special attention to the differential returns to human capital investment if they are concerned about a rapid urbanization.

Finally, while this study presents supportive results for the neoclassical explanation, there are some caveats to be noted. Ideally one needs to account for confounding factors such as ambition and motivation as well as risk attitudes. Assuming these factors count, the returns to education of migrants may have been somewhat overestimated. In fact, the reality of migration decisions may not be as straightforward as it is explained by the human capital theory of migration.

TABLE III: Changes in Migration Propensities and Returns to Education

Education level	**Trend**	**Migrants to big cities**		**Rural stayers**	
		1993-2000	2000-2007	1993-2000	2000-2007
education years	Growth in returns to education	0.082	0.050	0.102	0.011
	migration propensities	0.001	0.002		
middle school education vs. elementary education	Growth in returns to education	0.461	-0.219	0.285	-0.024
	migration propensities	0.005	0.007		
high school education vs. elementary education	Growth in returns to education	0.583	0.130	0.613	0.030
	migration propensities	0.011	0.028		
college education vs. elementary education	Growth in returns to education	0.841	0.308	0.994	0.106
	migration propensities	0.012	0.036		

Appendix I: Definition of Explanatory Variables and Descriptive Statistics

Abbreviation	**Definition**		
Dependent Var		Mean	St.Dev
city00-07	Migrated to big cities between 2000 and 2007	0.035	0.184
city93-00	Migrated to big cities between 1993 and 2000	0.027	0.185
Personal			
Elementary	1 if education is higher than elementary school (reference)	0.402	0.490
middle school	1 if education is above elementary school but less than high school	0.199	0.399
high school	1 if education is above middle school but below college	0.275	0.446
University	1 if education is higher than college	0.125	0.330
married	1 if married	0.425	0.494
age	age in years	37.612	11.009
female	1 if female	0.492	0.500
self employed	1 if self employed	0.356	0.479
government	1 if government worker	0.068	0.251
private	1 if private worker	0.332	0.471
unpaid	1 if unpaid family worker (reference)	0.125	0.331
casual[1]	1 if casual worker	0.119	0.324
risk attitude 1	most risk averse (reference)	0.378	0.485
risk attitude 2	risk averse	0.402	0.490
risk attitude 3	risk loving	0.145	0.352
risk attitude 4	most risk loving	0.074	0.262
Family			
family income (ln)	logarithm of total family yearly income	15.446	1.194
land possession	1 if land holders	0.316	0.328

Village			
earnings gap	mean income of the closest metro city[2] minus that of kecamatan of origin	2.276	4.778
standard deviation gap	Income standard deviation of the closest metro city minus that of kecamatan	0.848	0.695
distance (km)	travel distance by a car to the center of the closest metro city	83.454	82.001
Island			
Java island	1 if Java island	0.725	0.447

Source IFLS.* All variables were measured prior to any migration except risk aversion and time preference variables, which were available only in 2007 IFLS data. 1) Casual workers are defined to be people who supply services on an irregular or flexible basis. 2) Cities with population of one million or larger.

APPENDIX II: HECKMAN SELECTION MODEL IN EARNING ESTIMATIONS

Dependent Var	Migration=Yes(*N*=389)	
Personal	coefficient	(*t*-value)
Educational		
middle school	0.183*	(1.785)
high school	0.457***	(5.094)
university	0.570***	(4.871)
General		
married	-1.014***	(-12.274)
age	-0.012***	(-2.604)
female	-0.493***	(-6.688)
Vocational		
self	0.912**	(2.513)
gov	0.849**	(2.247)
private	1.153***	(3.179)
Casual	0.996***	(2.688)
Behavioral		
risk attitude 2	0.015	(0.174)
risk attitude 3	0.018	(0.213)
risk attitude 4	0.105	(0.582)
Family		
Yearly income(ln)	-0.160***	(-3.946)
land	-0.366***	(-4.289)
Village		
earngap	-0.097***	(-2.653)
std gap of earnings	0.045	(0.848)
distance(ln)	-0.135***	(-4.044)
Island		
Java	-0.134	(-1.476)
constant	1.311*	(1.705)

REFERENCES

[1] R. R. Lucas, "Life earnings and rural-urban migration," *Journal of Political Economy*, vol. 112, no. 1, pp. 29, 2004

[2] E. I. Glaeser and D. C. Mare, "Cities and skills," *Journal of Labor Economics*, vol. 19, no. 2, April, pp. 316-342. 2001

[3] B. Mc Cormick and J. Wahba, "Why do the young and educated in LDCs concentrate in large cities? Evidence from migration data," *Economica*, vol. 72, no. 285, pp. 39-67, 2005.

[4] Wozniak, "Why are college graduates more responsive to distant labor market opportunities?" University of Notre Dame, 2006.

[5] R. E. Lucas, "The effects of proximity and transportation on developing country population migrations," *Journal of Economic Geography*, vol. 1, no. 3, pp. 323, 2001.

[6] R. Jensen, "The (perceived) returns to education and the demand for schooling," *The Quarterly Journal of Economics*, vol. 125, no. 2, pp. 515, 2010.

[7] J. J. Heckman, "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," NBER, 1976.

[8] J. J. Heckman, "Sample selection bias as a specification error," *Econometrica: Journal of the Econometric Society*, pp. 153-161, 1979

[9] J. D. Vanzo and J. R. Hosek., *Does Migration Increase Wage Rates?: An Analysis of Alternative Techniques for Measuring Wage Gains to Migration*, Santa Monica, CA.: Rand, 1981.

[10] E. I. Glaeser, "Learning in cities," *Journal of Urban Economics*, vol. 46, pp. 254-257, 1999.

Sung Soo Lim was born in South Korea. He received his B.A. degree from University of Wisconsin at Madison and his M.Sc. from University of Illinois at Urbana-Champaign. He also holds his Ph.D. degree in applied economics, which was awarded from University of Minnesota at Twin Cities in 2012. Dr. Lim is currently an assistant professor of economics at American University in Dubai. His main research area is in development economics.