The Journal of Development Studies

ISSN: 0022-0388 (Print) 1743-9140 (Online) Journal homepage: http://www.tandfonline.com/loi/fjds20

Who Benefits from Higher Education in Low- and Middle-Income Countries?

M. Najeeb Shafiq, Robert K. Toutkoushian & Alexandria Valerio

To cite this article: M. Najeeb Shafiq, Robert K. Toutkoushian & Alexandria Valerio (2018): Who Benefits from Higher Education in Low- and Middle-Income Countries?, The Journal of Development Studies, DOI: 10.1080/00220388.2018.1528351

To link to this article: https://doi.org/10.1080/00220388.2018.1528351

The Journal of Development Studies, 2018
https://doi.org/10.1080/00220388.2018.1528351

Who Benefits from Higher Education in Low- and Middle-Income Countries?

M. NAJEEB SHAFIQ*, ROBERT K. TOUTKOUSHIAN** & ALEXANDRIA VALERIO†
*School of Education and Graduate School of Public and International Affairs, University of Pittsburgh, Pittsburgh, PA, USA, **Institute of Higher Education, University of Georgia, Athens, GA, USA, †The World Bank, Washington, DC, USA

(Original version submitted July 2017; final version accepted September 2018)

ABSTRACT *In this article, we investigate how higher education contributes to the employment and earnings of individuals in labour markets, and whether social origins play a role in the financial benefits from higher education. We focus on these questions in nine low- and middle-income countries: Armenia, Bolivia, Colombia, Georgia, Ghana, Kenya, Laos, Macedonia, and Vietnam. We use the recent Skills Towards Employability and Productivity (STEP) surveys of urban labour force participants to examine individuals' educational attainment, labour market participation, and earnings. Using logistic regressions, we find that individuals from disadvantaged origins are less likely to obtain a higher education degree. We find that in most of these countries, individuals who have earned a higher education degree are significantly more likely to be in the labour force and find employment, and enjoy sizable earnings premia. The findings are fairly robust with regard to the samples of individuals examined, and the methods used to measure earnings premia. Finally, we find little evidence that the earnings premia from higher education vary by social origins or the likelihood of an individual completing a degree. These results suggest that the benefits from higher education are comparable for individuals from disadvantaged and advantaged social origins.*

1. Introduction

Public support for higher education is a contentious matter in low- and middle-income countries (Carnoy, Froumin, Loyalka, & Tilak, 2014; Carnoy et al., 2013; Husén, 1987; Post et al., 2004). Starting in the 1970s, many countries began to expand the number of higher education institutions that offered bachelor's, master's, and doctor's degrees (Altbach, 2016). By the 1990s, several low- and middle-income countries had higher bachelor's level enrolment rates than European countries did only a few decades ago (Schoefer and Meyer, 2005). Although such higher education investments have contributed to economic growth and development, such as technology adoption (McMahon, 2009), the payoffs to individuals from higher education and the contribution of higher education to social mobility is unclear (Ferreyra, Ciro, Javier, Paz, & Urzúa, 2017; Majgaard & Mingat, 2012; Shavit, Arum, & Gamoran, 2010). Critics argue that public funds given to higher education systems primarily subsidise people from privileged social origins, that is, those who grew up with educated, wealthy, and engaged parents, because individuals from disadvantaged origins are less likely to access higher education (Shah & McKay, 2018; Sullivan, Parsons, Wiggins, Heath, & Green, 2014). Critics further argue that even if individuals from disadvantaged origins obtain higher education

Correspondence Address: M. Najeeb Shafiq, Economics and International Affairs, University of Pittsburgh, 230 S. Bouquet Street, 5907 Posvar Hall, Pittsburgh, PA 15260, USA. Email: mnshafiq@pitt.edu
Supplementary Materials are available for this article which can be accessed via the online version of this journal available at https://doi.org/10.1080/00220388.2018.1528351

credentials, they receive lower earnings premia because they attend lower-quality institutions and have weak social networks for obtaining higher-paying jobs (Gerber & Cheung, 2008).

In this article, we investigate the magnitude of labour market benefits to individuals in nine low- and middle-income countries with regard to higher education degree attainment. We first focus on the role of social origins (including socio-economic and other factors) in determining whether individuals in these countries earn a higher education degree. We then turn to whether degree attainment helps enable individuals to participate in the labour market and find work. Finally, we use a series of statistical methods to estimate the relationship between degree attainment and earnings, and whether these relationships vary by the social origins of individuals in these countries.

For the analysis, we use data from the *Skills Towards Employability and Productivity (STEP)* surveys of urban male and female labour force participants in southeastern Europe (Armenia, Georgia, and Macedonia), Latin America (Bolivia and Colombia), sub-Saharan Africa (Ghana and Kenya), and southeast Asia (Laos and Vietnam). The STEP surveys contain data on social origins, including retrospective data on parental socio-economic status, parental human capital, parental engagement, economic shocks suffered, early childhood education, and siblings. We compare male and female labour force participants holding higher education degrees to those who have completed an upper-secondary education or less.

We find that social origins, primarily through parental education, are strongly associated with higher education degree attainment in these countries. The results show that there are sizable differences in labour force and work participation between individuals with and without degrees in eight of the nine countries we studied. Likewise, the findings from our regression and propensity score models reveal that there are large earnings premia for individuals with degrees in eight of the nine countries. Finally, we find little evidence that the payoff to a degree varied with social origins, nor with the likelihood of the person earning a degree. Overall, we conclude that higher education systems may promote social mobility by providing the disadvantaged with comparable benefits to those from more advantaged backgrounds.

This study contributes to the scarce empirical literature on equity in higher education in low- and middle-income countries (for example Hu & Hibel, 2014). From a policy perspective, we highlight the extent to which higher education supports social mobility in each of the nine countries. By addressing issues of higher education equity, our study has implications on whether the governments of nine countries have to intervene to make higher education systems more accessible and beneficial for individuals from disadvantaged origins.

2. Literature review

2.1. Social origins and higher educational attainment

The empirical relationship between social origins and higher educational attainment has been extensively studied in the social sciences since at least 1967 (Blau and Duncan, 1967). According to the economic theories developed by Becker (1975, 1993), families with higher socio-economic status have more financial resources to fund their children's higher education (Belley & Lochner, 2007). Many studies have shown that there is a connection between the choices that individuals make about higher education and the educational attainment of their parents (Pascarella & Terenzini, 2005). Social and cultural capital theories (Bourdieu, 1977; Coleman, 1988) suggest that individuals with highly-educated or engaged parents are more likely to learn about higher education from them and are thus more likely to consider going to university themselves. In addition, students with highly-educated parents may have more interactions with other adults who have also gone to university, which in turn may assist them in forming their own educational aspirations (Sandefur, Meier, & Campbell, 2006). Finally, economic theories on the quality and quality of children predict that birth order and the number and composition of siblings may affect higher educational attainment (Becker, 1993); for instance, holding all else constant, parents may invest more in older children because they provide greater benefits over the parental life cycle.

2.2. Higher education degree attainment and labour market earnings premia

There is a sizable literature documenting the association between higher education degree attainment and earnings. Studies have been conducted using aggregate-level data to compare the average earnings of those with and without a higher education degree (Psacharopoulos & Patrinos, 2018; Toutkoushian, Shafiq, & Trivette, 2013; Toutkoushian & Paulsen, 2016). These studies have shown that, on average, those who have enrolled in and/or completed a higher education degree achieve significantly higher lifetime earnings.

Likewise, other studies (for example Heckman, Humphries, & Veramendi, 2016) have used individual-level data to estimate Mincerian earnings equations (Mincer, 1974) to examine the association between the natural log of weekly earnings and degree attainment. These Mincerian studies have likewise found that earnings are higher for individuals with higher education degrees, even after controlling for other observable factors that could affect earnings, such as years of potential work experience, ability, race, and gender.[1] Though most studies interpret the coefficient on education as 'economic returns,' we prefer the term 'earnings premia' because, strictly speaking, estimates of returns should include the direct costs of obtaining degrees, such as tuition and fees; our data source, however, lacks direct educational cost data.

2.3. Heterogeneous earnings premia by social origins

It is possible that the effect of education on earnings is not uniform across individuals but, rather, heterogeneous. As Brand and Xie (2010) articulate, economists and sociologists typically emphasise different explanations for heterogeneous effects of holding higher education degrees. Economists attribute differences to 'positive selection,' meaning the individuals most likely to select into university are also those who may benefit the most from the degree in their labour market earnings. This behaviour is consistent with rational choice theory, such that individuals will choose to invest their time and money in attaining a degree only if it offers higher net benefits. In comparison, while individuals with lower ability and motivation may obtain degrees, they are likely to earn less in the labour market.

Sociologists have focused more on the social origins explanations of heterogeneous higher education degree effects, such as explanations positing parental socio-economic status, parental engagement, or parental human capital as influential. Brand and Xie (2010) theorise that *net* of observed economic and noneconomic factors influencing degree attainment, there is a 'negative selection,' meaning the individuals who are least likely to obtain a degree are the ones who would benefit the most in their labour market earnings. Brand and Xie's intuition is that the connection between social origins and occupational outcomes, such as earnings, is stronger for less-educated workers than it is for more-educated workers, and thus individuals in lower social origin groups would benefit more from additional education.

To illustrate the Brand and Xie (2010) framework, let us consider two individuals: the first has a highly disadvantaged social origin (wealth quintile one, or poorest), and the second individual has a privileged social origin (wealth quintile four, or upper-middle class). Suppose that the higher education degree treatment improves the position of the first individual from wealth quintile 1=one to wealth quintile four, and improves the position of the second individual from wealth quintile four to wealth quintile five (richest). In this case, we can conclude that the higher education degree treatment yields greater improvements in both relative social mobility and absolute social mobility for the disadvantaged individual than for the privileged individual. In other words, individuals from disadvantaged origins benefit more from a higher education degree.

3. Context and data

3.1. Context

The nine countries were selected to illustrate the regional and economic diversity among countries that are considered low- and middle-income. These countries represent several world regions:

southeastern Europe (Armenia, Georgia, and Macedonia), Latin America (Bolivia and Colombia), sub-Saharan Africa (Ghana and Kenya), and southeast Asia (Laos and Vietnam). The Supplementary Materials include economic metrics for the nine countries such as the annual per-capita income, economic growth rate, and Gini coefficient to provide context for understanding the strength of the relationships between social origins, higher education degree attainment, and earnings. For example, in the lower income countries, having disadvantaged social origins may make it especially unlikely for one to obtain a higher education degree. In higher income countries, there may be more opportunities and benefits for individuals with higher education degrees.

3.2. Data

The STEP Skills Measurement surveys are a World Bank initiative for documenting the skills of urban adults in select low- and middle-income countries (Pierre, Sanchez, Valerio, & Rajadel, 2014). Country participation in the STEP survey is a function of several factors including interest, logistics, and funding. The selected countries were intended to illustrate the economic and regional diversity among countries that are considered low- and middle-income. In this study, we use countries that were included in the initial wave between March 2012 and July 2014.[2]

The STEP Survey sampling strategy was designed to ensure that the target population represents at least 95 per cent of the urban working-age population (ages 15 to 64) in each country. To allow comparability of the data collected with other country surveys and to account for country contexts, the STEP surveys used each country's official definition of 'urban'. This was also essential to the quality of the sample frames. To ensure consistency of the sampling strategies across all countries, all survey firms designed their sampling strategies in close cooperation with the STEP survey methodologist, who approved all sampling plans and drew the sample files used in each country.

The surveys share some similarities with the OECD's Programme for International Assessment of Adult Competencies (PIAAC) surveys but conducted in low- and lower-middle-income countries. The data include information on household characteristics and in-depth information from a randomly selected member of each household on his or her education, skills, and employment (Pierre et al., 2014). We use the sampling weights provided by STEP in all of our analyses to ensure that the members of the sample contribute to the population estimates only in the proportion in which they exist in the population.

For our study, we considered male and female labour force participants in the 18–65 age-group who are no longer enrolled in a formal educational institution. Given the relative prevalence of unemployment in low- and middle-income countries, we included all individuals who were identified as being in the labour force (employed or seeking employment) in our primary analyses. Appendix Table A1 shows the variable names and descriptions.

Table 1 presents the means of the explanatory variables (excluding region) in the nine countries. Our primary variable of interest is a dummy variable for whether the respondent completed a higher education degree (*ED*). Following UNESCO (2011) International Standard Classification of Education (ISCED), we consider individuals at education level six (bachelor's), seven (master's), or eight (doctorate) to be college educated ($ED = 1$). The comparison group consists of individuals who are level zero (no formal education or less than primary), one (primary completed), two (lower secondary completed), four (post-secondary specialist technical/vocational), or five (two-year professional); for individuals in this category, we assigned $ED = 0$. According to the data, the three southeastern European countries have relatively high percentages of individuals with at least a bachelor's degree.

The main outcome variable of interest in our study is labour market earnings. STEP permits us to measure benefits using weekly earnings in each nation's currency. To ensure a better fit with the data, ease interpretation of the coefficients and retain those without earnings in the study, we used logged weekly earnings as our dependent variable (*ln(wage)*) after adding one unit of currency to each respondent's weekly wage.

Table 1. Descriptive statistics for selected variables for each country

	Southeastern Europe			Latin America		Sub-Saharan Africa		Southeast Asia	
Variable	Armenia	Georgia	Macedonia	Bolivia	Colombia	Ghana	Kenya	Laos	Vietnam
Higher Education (ED)	46%	54%	27%	20%	9%	7%	8%	4%	23%
Weekly Wage	27.90	72.41	2899.59	542.81	188,959.50	127.21	3969.90	238,381.20	1317.25
ln(Wage)	1.73	2.46	5.55	5.13	10.20	3.33	5.95	7.95	6.11
Female	63%	60%	49%	56%	54%	56%	49%	50%	57%
Yrs Experience	19.58	18.22	20.85	16.76	17.26	14.11	13.01	21.01	20.86
(Yrs Experience) 2	535.76	471.64	580.76	455.58	490.56	323.42	283.75	608.50	596.51
Attend Preschool	61%	71%	52%	51%	49%	76%	72%	8%	70%
Birth Order	2.11	1.82	1.93	3.13	3.01	3.67	3.97	3.15	3.15
Number of Siblings	2.22	1.77	1.70	4.16	3.94	4.76	5.05	4.54	3.94
Lower Family SES	8%	9%	11%	36%	33%	18%	25%	42%	33%
Middle Family SES	47%	51%	61%	56%	56%	60%	63%	53%	60%
Upper Family SES	45%	41%	28%	7%	11%	22%	12%	5%	7%
Number Economic Shocks	0.32	0.27	0.14	1.57	0.96	0.73	1.16	1.33	0.52
Parent Ed: Primary	4%	2%	26%	11%	56%	4%	24%	16%	46%
Parent Ed: Secondary	40%	29%	45%	27%	28%	71%	33%	10%	21%
Parent Ed: Higher	56%	69%	18%	17%	8%	18%	17%	3%	10%
Parents Engaged in Ed	78%	82%	82%	78%	81%	64%	57%	85%	89%
Total Sample Size	2,354	2,566	3,238	1,410	1,921	1,183	2,865	1,638	2,578
Sample in Labour Force	1,342	1,498	2,138	1,249	1,582	1,031	2,474	1,487	2,046

Notes: Sample restricted to respondents ages 18 to 65 who are in the labour force, not currently enrolled in school and do not have missing data on the variables used in this study. Data are weighted to account for the stratified sampling design. ED = 1 if earned a bachelor's degree or higher, else 0. Weekly wages are shown in home currencies. Parents engaged in educ = 1 if parents were reported to have been 'highly engaged' in *i*'s education during childhood, else 0. More complete variable definitions are provided in the Appendix.

The explanatory variables that we used in the study include the person's gender, whether the person attended preschool, their years of labour market experience, number of siblings, and their birth order in the family. The explanatory variables also include family characteristics including family socio-economic status (three categories), parental education, number of economic shocks the family experienced prior to the respondent turning 15, and whether the parents were reported to have been 'highly engaged' in the respondent's education.[3] When available in the STEP survey data, we also relied on a series of dummy variables for the district/region/province within each country where a person resides.[4]

We acknowledge several data limitations. We lack important determinants such as a person's academic ability and motivation. We also acknowledge that dummy variables are a crude measure of some of the underlying factors in our models.

4. Analysis: the role of social origins on higher education degree attainment

To assess the role of social origins and other selected characteristics on higher education degree attainment, we estimated the following logistic regression model for each country:

$$ED = \alpha_0 + \boldsymbol{\alpha}_1 \mathbf{X} + \varepsilon \tag{1}$$

where $ED = 1$ if the individual has a higher education degree and = 0 otherwise, and $\mathbf{X}$ includes the set of personal and family variables defined earlier excluding years of labour market experience. Table 2 presents the results from the logistic regression models. For ease of interpretation, we convert the coefficients in Equation (1), which represent the effect of each variable on the log odds of earning a higher education degree, to average marginal effects. Of particular interest in this model is the relationship between social origin (as represented by parental education and family SES) and the educational attainment of the respondent.

The results show that several of the social origins variables are statistically associated with higher education degree attainment. We found that parental education is the most prominent and consistent factor associated with obtaining a higher education degree. Across all nine countries, individuals whose parents have earned a higher education degree were themselves more likely to do the same. The marginal effects range from a low of 3.6 per cent in Laos to a high of 30.2 per cent in Armenia. In five of the nine countries we studied, the marginal effects of this variable exceeded 20 per cent. There was evidence in Macedonia, Bolivia, Ghana, Kenya, and Laos that females are only between 0.3 per cent and 0.7 per cent less likely than males to earn a degree. Family socio-economic status and number of economic shocks also exhibited small but positive and significant associations with degree attainment in half of the countries, even after controlling for parental education. There is little evidence that birth order or the number of siblings are related to degree attainment.

The results on higher education attainment show some patterns. We observe a gender gap favouring males in the three poorest counties (Ghana, Kenya, and Laos). As expected, the magnitudes of the marginal effects are larger in the countries with bigger shares of individuals with higher education degrees (see Table 1); in particular, the marginal effects for parental education are largest for countries in southeastern Europe and Latin America. In the earlier section, we raised the possibility that having disadvantaged social origins should make it especially unlikely for one to obtain a higher education degree in the poorest countries or regions. Yet our results show statistically significant social origins variables in all nine countries, regardless of per-capita incomes or region. Thus, if we only focus on degree attainment, we can conclude that individuals from disadvantaged social origins are less likely to benefit from higher education systems.

5. Analysis: the role of higher education degree attainment on labour market participation

In this section, we focus on whether individuals with a higher education degree are more likely than those without a higher education degree to be able to participate in the labour market. Conceptually, labour force participation and employment make it possible for individuals to earn wages. Thus, having a higher education degree is beneficial if it improves the possibility of labour force participation and employment. To do this, we specified the following three logistic regression models:

$$LF = \beta_0 + \beta_1 ED + \beta_2 \mathbf{X} + \varepsilon \tag{2.1}$$

Table 2. Logistic regression models of educational attainment

	Southeastern Europe			Latin America		Sub-Saharan Africa		Southeast Asia	
Variable	Armenia	Georgia	Macedonia	Bolivia	Colombia	Ghana	Kenya	Laos	Vietnam
Female	−0.032	0.010	0.069***	−0.071+	−0.019	−0.041*	−0.036**	−0.026**	−0.015
	(0.030)	(0.029)	(0.021)	(0.038)	(0.018)	(0.018)	(0.011)	(0.010)	(0.020)
Attended Preschool	0.039	0.062*	0.058**	0.024	0.029+	0.023	0.077***	0.047***	0.075**
	(0.030)	(0.028)	(0.022)	(0.029)	(0.017)	(0.022)	(0.021)	(0.013)	(0.024)
Family SES – Middle	0.076	0.045	0.024	0.056	0.046+	0.049	−0.018	0.014	0.069**
	(0.061)	(0.056)	(0.041)	(0.035)	(0.024)	(0.032)	(0.016)	(0.011)	(0.022)
Family SES – High	0.101	0.175**	0.047	0.048	0.076**	0.018	0.040*	0.026	0.082*
	(0.062)	(0.059)	(0.047)	(0.055)	(0.029)	(0.038)	(0.018)	(0.021)	(0.037)
Number Econ Shocks	−0.046*	−0.044*	−0.047+	−0.015	−0.027**	−0.030*	−0.008	−0.010*	−0.046***
	(0.020)	(0.022)	(0.024)	(0.013)	(0.010)	(0.013)	(0.005)	(0.004)	(0.012)
Parent Ed: Primary	−0.063	−0.290+	−0.113***	0.039	0.026	−0.111	−0.052**	−0.001	−0.045*
	(0.086)	(0.154)	(0.030)	(0.048)	(0.019)	(0.069)	(0.018)	(0.014)	(0.022)
Parent Ed: Higher	0.302***	0.244***	0.246***	0.241***	0.148***	0.063**	0.101***	0.036+	0.223***
	(0.027)	(0.024)	(0.023)	(0.032)	(0.025)	(0.019)	(0.013)	(0.019)	(0.027)
Parents Engaged in Ed	0.099*	0.150***	0.094**	0.007	0.016	0.033+	0.034+	−0.006	−0.019
	(0.040)	(0.038)	(0.035)	(0.038)	(0.022)	(0.019)	(0.018)	(0.016)	(0.028)
Birth Order	−0.014	−0.037*	0.013	0.024*	0.007	0.005	−0.001	0.007*	0.005
	(0.017)	(0.018)	(0.015)	(0.011)	(0.009)	(0.004)	(0.004)	(0.003)	(0.008)
Total Siblings	−0.006	−0.014	−0.044**	−0.017+	−0.001	−0.002	−0.000	−0.007+	−0.013+
	(0.015)	(0.012)	(0.014)	(0.009)	(0.008)	(0.002)	(0.003)	(0.004)	(0.008)
Sample Size	1345	1503	2138	1251	1584	1032	2483	1487	2051
F-Statistic	7.21	8.23	20.46	6.79	5.84	3.65	18.80	7.88	18.12
P-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Dependent variable in each model = ED such that ED = 1 if earned a bachelor's degree or higher, else 0. Standard errors are shown in parentheses. Coefficients are reported as average marginal effects. Data are weighted to account for the stratified sampling design. Sample restricted to respondents ages 18 to 65 who are in the labour force, not currently enrolled in school and do not have missing data on the variables used in this study. Weekly wages are shown in home currencies. Parents engaged in ed = 1 if parents were reported to have been 'highly engaged' in *i*'s education during childhood, else 0. *** p < .001, ** p < .01, * p < .05, + p < .10 (two-tailed tests).

$$EMPL = \beta_0 + \beta_1 ED + \beta_2 \mathbf{X} + \varepsilon \tag{2.2}$$

$$EMPL = \beta_0 + \beta_1 ED + \beta_2 \mathbf{X} + \varepsilon \quad \text{if} LF = 1 \tag{2.3}$$

where $LF = 1$ currently if in the labour force, else 0; and $EMPL = 1$ if currently employed, else 0. Equations (2.1) and (2.2) are estimated for the full sample (that is those in and out of the labour force), whereas Equation (2.3) only applied to those who were in the labour force. Table 3 summarises the results for the educational attainment variable (*ED*) in each model and country. The first three columns show the differences in means for the respective labour force measures for individuals with and without higher education degrees. The last three columns report the regression coefficients for the *ED* variable, which denote the differences in labour force measures after taking into account the variables in Equations (2.1) to (2.3). The variables **X** in these models also include years of labour market experience (linear and quadratic). All of the results in the last three columns are expressed as marginal effects.

The mean differences in columns 1–3 reveal that individuals with higher education degrees are significantly more likely than their counterparts to be in the labour force or be employed. The advantages in labour force participation rates for individuals with degrees range from 7 per cent to almost 25 per cent. After controlling for other personal and family characteristics, individuals with degrees in all countries except Colombia are still more likely to be in the labour force. In the last column, where we compare the employment rates for individuals with and without higher education degrees, we find that having a degree is associated with a higher likelihood of employment in all countries except Colombia and Laos.

Table 3. Effects of higher education degree completion on labour market participation

	Difference in Means			Regression Analysis		
Country	Labour Force	Employed	Emp | LF	Labour Force	Employed	Emp | LF
Armenia	0.213***	0.258***	0.194***	0.177***	0.224***	0.204***
	(0.020)	(0.020)	(0.025)	(0.026)	(0.022)	(0.030)
Georgia	0.139***	0.181***	0.169***	0.086***	0.142***	0.161***
	(0.019)	(0.018)	(0.025)	(0.025)	(0.023)	(0.033)
Macedonia	0.246***	0.248***	0.073***	0.116***	0.153***	0.103***
	(0.021)	(0.022)	(0.021)	(0.026)	(0.029)	(0.031)
Bolivia	0.073***	0.093***	0.026+	0.146***	0.161***	0.047+
	(0.020)	(0.023)	(0.015)	(0.038)	(0.040)	(0.026)
Colombia	0.075*	0.110**	0.049+	−0.003	0.049	0.060
	(0.030)	(0.035)	(0.027)	(0.057)	(0.064)	(0.042)
Ghana	0.093**	0.108*	0.025	0.171*	0.263***	0.140**
	(0.035)	(0.043)	(0.031)	(0.076)	(0.068)	(0.046)
Kenya	0.096***	0.151***	0.078**	0.119*	0.200***	0.142***
	(0.023)	(0.030)	(0.027)	(0.047)	(0.049)	(0.041)
Laos	0.080**	0.065*	0.013+	0.138**	0.098*	−0.001
	(0.028)	(0.029)	(0.008)	(0.052)	(0.038)	(0.009)
Vietnam	0.095***	0.105***	0.016*	0.041	0.070*	0.041**
	(0.019)	(0.019)	(0.007)	(0.029)	(0.031)	(0.013)

Notes: Standard errors for all statistics are shown in parentheses. Columns 1–3 report the mean differences in log wages for those with and without a bachelor's degree. Columns 4–6 show the estimated coefficients for the higher education variable from the logistic regression models. Labour Force = 1 if in the labour force, else 0. Employed = 1 if employed, else 0 (includes both labour force participants and non-participants). Emp | LF = 1 if employed, else 0 (conditional on being in the labour force. Logistic regression models control for the same variables as used in Table 3, as well as years of experience and squared experience. *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .10$ (two-tailed tests).

6. Analysis: average earnings premia from a higher education degree

We now turn to estimating the *average* benefits from a higher education degree in order to establish a baseline or set of reference points. For instance, if we found statistical evidence of earnings premia from obtaining a degree, but subsequent analysis showed that the benefits only accrue to those most likely to obtain a degree, such results would suggest that higher education does not promote social mobility. We recognise the possible sensitivity of our estimates to alternative statistical models, and therefore estimate the average earnings premia from a higher education degree using four methods: Ordinary Least Squares (OLS), Tobit, hurdle, and propensity score methods.

6.1. Earnings premia: ordinary least squares (OLS)

To estimate the average earnings premia, we consider the Mincerian earnings equation:

$$\ln(y) = \delta_0 + \delta_1 ED + \delta_2 \mathbf{X} + u \tag{3}$$

where ln(y) is logged weekly earnings, *ED* is again a dummy representing whether the individual has a bachelor's degree or higher, and $\boldsymbol{X}$ is a vector of social origin and potential labour market experience covariates. The parameters $\boldsymbol{\delta_2}$ are regression coefficients measuring the changes in log earnings associated with changes in the covariates. The key parameter of interest (δ_1) shows the relationship between higher education degree attainment and log wages after taking into account the factors in $\boldsymbol{X}$. Typically, OLS earnings equations are only run on data from the employed (that is, those with positive wages). In contrast, we expand the sample to also include unemployed individuals to take into account the effect of education on both earnings and being in the labour force. The results from the analysis are shown in Table 4. For reference, the first row provides the difference in mean (log) wages between individuals with and without higher education degrees.

We find that there are statistically significant coefficients for the higher education degree indicator coefficient in all of these countries except for Bolivia. The coefficients suggest large associations between higher education and earnings. At one extreme, a person with a higher education degree in Laos earns almost three times (δ_1 = 2.957) more than the amount of someone with an upper-secondary education or less, controlling for the full set of covariates. Other estimates of this earnings advantage range from 56 per cent in Macedonia to 213 per cent more in Kenya. The only country without a statistically significant effect of education on log wages was Bolivia.

Of the remaining regressors in the models, there are quadratic effects of experience on log wages in all nine countries. We further find that in six of the nine countries, females earned less than comparable men, with the gender pay gaps being particularly large in Latin America and sub-Saharan Africa. The vast majority of the remaining social origin variables, however, are not significantly associated with earnings in the regression models. Of particular interest to our study is the lack of significance for the socio-economic measures in our model. This suggests that individuals from disadvantaged social origins do not face larger deficiencies in labour market earnings than individuals from more privileged origins.

6.2. Robustness checks on OLS

In the previous model, we restricted the analysis to those individuals who indicated that they were currently in the labour force. The sample used is potentially important for our study, however, because the earnings premia for workers with higher education degrees could be sensitive to the sample selected. For example, some of the financial benefits received by people with higher education degrees could be due to their greater likelihood of being in the labour force and finding work. In Table 5, we explore whether the average earnings premia varies according to the set of individuals examined in the analysis. Column 1 shows the average earnings premia from Table 5

Table 4. OLS estimates of determinants of log weekly wages

	Southeastern Europe			Latin America		Sub-Saharan Africa		Southeast Asia	
Variable	Armenia	Georgia	Macedonia	Bolivia	Colombia	Ghana	Kenya	Laos	Vietnam
Mean Diff in ln(Wage)	0.775***	0.920***	1.819***	0.599***	1.846***	1.415***	2.228***	3.277***	1.526***
By Education	(0.058)	(0.083)	(0.178)	(0.167)	(0.434)	(0.213)	(0.254)	(0.549)	(0.146)
Regression Coefficients:									
Higher Education	0.705***	0.840***	0.559*	0.285	1.202**	1.422***	2.131***	2.957***	1.125***
	(0.097)	(0.162)	(0.254)	(0.337)	(0.399)	(0.323)	(0.314)	(0.709)	(0.142)
Female	−0.532***	−0.303+	−0.154	−1.051***	−1.299***	−0.744***	−1.443***	0.163	−0.155
	(0.088)	(0.159)	(0.212)	(0.160)	(0.279)	(0.157)	(0.170)	(0.453)	(0.137)
Yrs Experience	0.033*	0.056**	0.159***	0.062*	0.108**	0.138***	0.191***	0.163*	0.103***
	(0.013)	(0.021)	(0.034)	(0.026)	(0.037)	(0.020)	(0.027)	(0.065)	(0.019)
Yrs Experience Squared	−0.001*	−0.001*	−0.003***	−0.002**	−0.002**	−0.003***	−0.004***	−0.003+	−0.002***
	0.000	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Attended Preschool	0.118	0.102	−0.060	0.193	−0.065	−0.103	−0.170	1.279*	−0.425**
	(0.097)	(0.151)	(0.231)	(0.181)	(0.344)	(0.148)	(0.176)	(0.575)	(0.145)
Family SES – Middle	−0.143	0.538+	0.508	0.096	−0.222	0.385**	0.539**	0.990+	−0.078
	(0.160)	(0.280)	(0.316)	(0.136)	(0.326)	(0.145)	(0.206)	(0.538)	(0.114)
Family SES – High	−0.192	0.439	0.216	−0.436	−1.007+	0.304	0.048	1.046	−0.127
	(0.173)	(0.289)	(0.394)	(0.373)	(0.583)	(0.227)	(0.385)	(0.994)	(0.274)
# Economic Shocks	−0.036	−0.030	−0.574*	−0.008	−0.144	−0.043	−0.133*	−0.022	−0.139*
	(0.069)	(0.128)	(0.284)	(0.061)	(0.135)	(0.061)	(0.062)	(0.144)	(0.069)
Parent Ed: Primary	−0.013	0.091	−0.038	0.523**	0.257	0.030	0.152	0.609	−0.079
	(0.239)	(0.522)	(0.236)	(0.170)	(0.316)	(0.368)	(0.203)	(0.648)	(0.150)
Parent Ed: Higher	−0.038	0.416*	0.373	−0.529	1.362**	0.271	0.462	−0.975	0.282
	(0.097)	(0.168)	(0.310)	(0.349)	(0.476)	(0.188)	(0.288)	(1.122)	(0.209)
Parents Engaged in Ed	−0.101	0.022	−0.022	−0.127	0.442	−0.106	−0.299	0.435	0.456*
	(0.116)	(0.201)	(0.262)	(0.188)	(0.368)	(0.138)	(0.185)	(0.791)	(0.194)
Birth Order	0.049	0.119	−0.238*	−0.055	−0.080	0.013	−0.004	0.179	−0.033

(*continued*)

Table 4. (*Continued*)

	Southeastern Europe			Latin America		Sub-Saharan Africa		Southeast Asia	
Variable	Armenia	Georgia	Macedonia	Bolivia	Colombia	Ghana	Kenya	Laos	Vietnam
	(0.045)	(0.084)	(0.110)	(0.050)	(0.094)	(0.034)	(0.058)	(0.163)	(0.043)
Number of Siblings	0.003	−0.033	0.080	0.004	0.059	0.002	−0.042	−0.183	0.033
	(0.046)	(0.048)	(0.113)	(0.050)	(0.087)	(0.024)	(0.046)	(0.145)	(0.037)
Constant	1.290***	0.501	3.801***	5.421***	11.073***	1.963***	5.334***	3.690+	5.109***
	(0.365)	(0.392)	(0.541)	(0.296)	(0.682)	(0.396)	(0.392)	(2.142)	(0.337)
Sample Size	1,342	1,498	2,138	1,249	1,582	1,031	2,474	1,487	2,046
R2	0.11	0.06	0.04	0.11	0.07	0.16	0.11	0.12	0.08

Notes: First row shows results from two-sample t-tests of mean differences in log wages by education. Dependent variable in each regression model = log of weekly wages in home country's currency. Standard errors are shown in parentheses. Data are weighted to account for the stratified sampling design. Sample restricted to respondents ages 18 to 65 who are in the labour force, not currently enrolled in school, and do not have missing data on the variables used in this study. Weekly wages are shown in home currencies. ED = 1 if earned a bachelor's degree or higher, else 0. Parents engaged in educ = 1 if parents were reported to have been 'highly engaged' in *i*'s education during childhood, else 0. *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .10$ (two-tailed tests).

Table 5. OLS robustness checks for higher education on log of weekly wages

	← – – – – – OLS Samples – – – – – →				← – Hurdle Samples – →	
Country	Labour Force	All	Employed	Tobit Model	Labour Force	All
Armenia	0.705***	0.730***	0.264***	1.100***	0.696***	0.678***
	(0.097)	(0.075)	(0.078)	(0.162)	(0.095)	(0.069)
	1,342	2,354	887	1,342	1,342	2,354
Georgia	0.840***	0.690***	0.269*	1.431***	0.831***	0.673***
	(0.162)	(0.108)	(0.119)	(0.295)	(0.157)	(0.105)
	1,498	2,566	899	1,498	1,498	
Macedonia	0.559*	1.005***	−0.148	0.695+	0.568***	0.987***
	(0.254)	(0.242)	(0.239)	(0.371)	(0.254)	(0.229)
	2,138	3,238	1,652	2,138	2,138	3,238
Bolivia	0.285	0.694*	0.063	0.284	0.286	0.659**
	(0.337)	(0.319)	(0.336)	(0.389)	(0.282)	(0.316)
	1,249	1,410	1,177	1,249	1,249	1,410
Colombia	1.202**	0.964	0.689**	1.287**	1.314**	1.066
	(0.399)	(0.631)	(0.214)	(0.457)	(0.466)	(0.692)
	1,582	1,921	1,401	1,582	1,582	1,921
Ghana	1.422***	1.702***	1.023**	1.618***	1.339***	1.555***
	(0.323)	(0.345)	(0.313)	(0.377)	(0.297)	(0.326)
	1,031	1,183	941	1,031	1,031	1,183
Kenya	2.131***	2.408***	1.311***	2.556***	2.073***	2.307***
	(0.314)	(0.334)	(0.224)	(0.402)	(0.326)	(0.345)
	2,474	2,865	2,001	2,474	2,474	2,865
Laos	2.957***	3.674***	3.023***	4.082***	3.672***	4.545***
	(0.709)	(0.645)	(0.713)	(1.054)	(1.120)	(1.018)
	1,487	1,638	1,478	1,487	1,487	1,683
Vietnam	1.125***	1.093***	0.889***	1.209***	1.150***	1.125***
	(0.142)	(0.195)	(0.132)	(0.159)	(0.164)	(0.206)
	2,046	2,578	2,001	2,046	2,046	2,578

Notes: Dependent variable in each model = log of weekly wages in home country's currency. Coefficients in first row represent effect of higher education on log of weekly earnings. Standard errors are shown in parentheses in the second row. Sample sizes are shown in the third row. Data are weighted to account for the stratified sampling design. First and fourth columns show results for respondents who were in the labour force. Second column reports results for the full sample. Third column shows results for only those respondents who were employed. *** p < .001, ** p < .01, * p < .05, + p < .10 (two-tailed tests).

when the sample consists of those respondents who were in the labour force (employed or unemployed). The second column expands the sample to also include those who were not in the labour force. A narrower sample definition was used in the third column, where we restricted the sample to only those who were employed at the time of the survey.

Table 5 also includes two analyses that serve as robustness checks on how the OLS models treat individuals without earnings. We first use a Tobit model that jointly modelled the process of being employed and having earnings for those who were employed. In labour economics, Tobit models (originally developed by Tobin, 1958) are better for fitting data with sizeable shares of zero and positive wages. Tobit models typically produce more consistent estimates than OLS models (pp. 1988–90, Long, 1997). The Tobit model can be expressed in terms of a latent variable:

$$\ln(y^*) = \alpha + \delta ED + \mathbf{\beta X} + u \tag{4}$$

where $\ln(y^*) = \begin{cases} 0 & if\ \ln(y) \leq 0 \\ \ln(y) & if\ \ln(y) > 0 \end{cases}$ and $\ln(y)$ is zero for the unemployed and positive monetary amounts for those who are employed. The Tobit model results are statistically significant in all

countries except in Bolivia and Macedonia. In terms of magnitude, the higher education degree coefficients are consistently larger than the coefficients from the OLS models.

The last two columns in Table 6 show the findings when we used hurdle models, which are also referred to as two-part models. Cameron and Trivedi (2010, pp. 552–553) explained that the Tobit model makes a strong assumption that the same probability mechanism generates both the zeros and the positive values. In contrast, they argue that the hurdle model is more flexible to permit the possibility of different mechanisms in the two parts. In our hurdle model, the first part is a logistic regression equation that models the probability of positive earnings, and the

Table 6. Effects of higher education on log wages – propensity score methods

Country	Method	Measure	Treatment Effect	Standard Error	T-ratio	P-value
Armenia	PSM	ATE	0.723	0.101	7.13	0.000
	PSM	ATT	0.739	0.108	6.85	0.000
	PSW	ATE	0.709	0.101	7.02	0.000
	PSW	ATT	0.685	0.103	6.65	0.000
Georgia	PSM	ATE	0.887	0.154	5.76	0.000
	PSM	ATT	0.829	0.188	4.42	0.000
	PSW	ATE	0.799	0.163	4.90	0.000
	PSW	ATT	0.806	0.170	4.74	0.000
Macedonia	PSM	ATE	0.543	0.300	1.81	0.071
	PSM	ATT	0.248	0.236	1.05	0.294
	PSW	ATE	0.488	0.273	1.79	0.074
	PSW	ATT	0.442	0.268	1.65	0.099
Bolivia	PSM	ATE	0.277	0.165	1.68	0.093
	PSM	ATT	0.156	0.247	0.63	0.529
	PSW	ATE	0.257	0.303	0.85	0.397
	PSW	ATT	0.294	0.319	0.92	0.357
Colombia	PSM	ATE	0.996	0.690	1.44	0.150
	PSM	ATT	0.871	0.457	1.91	0.056
	PSW	ATE	1.295	0.479	2.70	0.007
	PSW	ATT	1.137	0.403	2.82	0.005
Ghana	PSM	ATE	0.824	0.208	3.97	0.000
	PSM	ATT	0.871	0.321	2.71	0.007
	PSW	ATE	1.105	0.319	3.46	0.001
	PSW	ATT	1.074	0.362	2.97	0.003
Kenya	PSM	ATE	1.243	0.453	2.74	0.006
	PSM	ATT	1.922	0.410	4.69	0.000
	PSW	ATE	1.754	0.358	4.90	0.000
	PSW	ATT	2.334	0.339	6.88	0.000
Laos	PSM	ATE	2.925	0.368	7.94	0.000
	PSM	ATT	2.097	0.623	3.37	0.001
	PSW	ATE	4.674	0.484	9.66	0.000
	PSW	ATT	3.085	0.675	4.57	0.000
Vietnam	PSM	ATE	0.912	0.117	7.77	0.000
	PSM	ATT	0.927	0.132	7.00	0.000
	PSW	ATE	0.973	0.161	6.04	0.000
	PSW	ATT	0.962	0.155	6.21	0.000

Notes: Dependent variable in each model = log of weekly wages in home country's currency. Coefficients represent effect of higher education on log of weekly earnings, standard errors are shown in parentheses. Data for PSW models are weighted to account for the stratified sampling design. Sample restricted to respondents ages 18 to 65 who are not currently enrolled in school and do not have missing data on the variables used in this study. PSW = propensity score weighting. PSM = propensity score matching using 'teffects psmatch' command in Stata 15 with nearest neighbour matching (n = 1 or more) and calliper = 5 per cent. Propensity scores were determined from logit regression model, and were used to construct weights for PSW models. ATE = average treatment effect. ATT = average treatment of the treated cases. *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .10$ (two-tailed tests).

second part uses linear expenditure to model log of earnings conditional on having positive earnings:

$$y^{**} = \alpha_1 + \delta_1 ED + \beta_{j1}\mathbf{X} + u \tag{5.1}$$

$$\ln(y) = \alpha_2 + \delta_2 ED + \beta_{j2}\mathbf{X} + u \quad \text{iffln(y)>0} \tag{5.2}$$

where y** = 1 if ln(y) > 0 and 0 otherwise. The results for the variable *ED* are expressed as average marginal effects which combined the results from Equations (5.1) and (5.2) to make them consistent with prior models. In the absence of obvious exclusion restrictions, we use the same covariates in both parts of the model. The hurdle model results in Table 6 are consistent with the OLS models in terms of statistical significance across the samples of labour force participants and all individuals. The magnitude of the hurdle model coefficients are generally smaller than the Tobit model results, but comparable to the OLS coefficients.

Overall, the results from the OLS, Tobit, and hurdle methods in Table 6 show that the earnings premia for individuals with higher education are positive and sizable in most of the nine countries regardless of the sample that we considered or the method used to model labour market participation. Although the earnings premia are smaller in all countries when we only considered those who are employed, there are still significant earnings differentials in seven of the nine countries. A few cases are worth noting. We see the largest differences between individuals with and without higher education degrees in Laos. Growth in higher education has excelled rapidly in the past decade. In particular, regions with substantial expansion in higher education expansion have also seen substantial growth in wages (UNESCO, 2008). We see the smallest benefits between individuals with and without degrees in Bolivia and Macedonia. The main issue facing Macedonia's higher education system, as well as the labour market is structural, and is related to national recovery and reorganisation (European Commission, 2009). In the case of Bolivia, the absence of earnings premia may reflect a generation of individuals with degrees who did not find adequate labour market opportunities during the country's economic struggles prior to 2004 (Morales, 2010).

6.3. Propensity score models of earnings premia

Although informative, the OLS approach to estimating the relationship between higher education and earnings has two potential limitations. The first is that OLS is a fully parametric model for estimating the relationship between higher education and earnings, in that the coefficient is dependent on the specific functional form used in OLS. The second concern is that there may be omitted or unobservable factors that are correlated with both educational attainment and earnings, and thus bias the estimated coefficient on the education variable (Heckman et al., 2016). Notably, the decision to obtain a higher education degree may be endogenous, such that individuals who have chosen to obtain a degree may differ from those who have chosen not to do so. If these pre-existing differences affect earnings in ways other than the effects of the degree itself, a comparison in outcomes between the two groups will produce a biased estimate of the causal effect of higher education degree attainment.

Accordingly, we turn to propensity score methods – specifically propensity score weighting (PSW) and propensity score matching (PSM) – as robustness checks on the parametric OLS results and to possibly help reduce the potential bias in higher education degree attainment (Rosenbaum & Rubin, 1983). Within each of these methods, we focus on both the average treatment effect (ATE) and the average treatment effect on the treated cases (ATT). All of these methods begin by estimating the probability of an individual having a higher education degree, which is referred to as a propensity score. Propensity scores ($\widehat{PS}$) are obtained by finding the predicted probability of each person having a degree in the logistic regression model shown in Equation (1).

Propensity score weighting is equivalent to weighted least squares, where the weights (w_{ATE}, w_{ATT}) used in the multiple regression model are derived from the propensity scores so that the means for the explanatory variables for higher education and non-higher education cases are balanced. The weights are derived as follows for the ATE and ATT in propensity score weighting:

$$w_{ATE} = \begin{cases} 1/\widehat{PS} \; if \; ED = 1 \\ 1/\left(1 - \widehat{PS}\right) \; if \; ED = 0 \end{cases} \tag{6.1}$$

$$w_{ATT} = \begin{cases} 1 \; if \; ED = 1 \\ \widehat{PS}/\left(1 - \widehat{PS}\right) \; if \; ED = 0 \end{cases} \tag{6.2}$$

We checked the balancing assumption that after applying weights the expectations of each covariate were the same for the individuals with and without higher education degrees:

$$\mathrm{E}(X|ED = 1) = \mathrm{E}(X|ED = 0) \tag{7}$$

where conditional on the propensity score, higher education degree attainment (*ED*) is independent of the characteristics (*X*), which results in both groups having similar post-matching observed characteristics. If the balancing assumption holds, then we can interpret the coefficient on the variable *ED* as being the treatment effect (ATE or ATT) of having earned a degree. In the supplementary files, we show the balancing tests for the ATE using PSW. For propensity score matching, we matched individuals with and without a degree based on having similar propensity scores, and then compared the mean log of wages for the two matched groups (treated versus non-treated for ATE, treated cases with and without treatment for ATT).[5]

We acknowledge that the propensity score methods are not a perfect identification strategy for making strong causal statements about the relationship between higher education degree attainment and earnings. Causality in these methods rely on the conditional independence assumption that all factors relevant for selection into treatment assignment are observed and taken into account in the formation of propensity scores. Since matching based on unobservable factors is by definition not possible, to the extent that there exist individual characteristics that influence both educational attainment and wages that we cannot control for, our estimates may still be affected by omitted variable bias. The richness of the STEP data, however, allows us to consider some traditionally unobserved characteristics and reduce this possible bias. Furthermore, the propensity score methods offer a robustness check on the fully parametric OLS approach to estimating the relationship between higher education and wages. As a result, the propensity score models offer valuable information in our study and may improve over the OLS model.

Table 6 presents summaries of the results for the higher education variable from the propensity score models for each of the nine countries. The sample consists of all individuals in the labour force. The second column indicates whether we used PSW or PSM technique. The third column denotes whether the treatment effect shown is the average treatment effect or the average treatment for only the treated cases. The estimated treatment effect (ATE or ATE) and its corresponding standard error, calculated t-ratio, and (two-tailed) p-value are then provided in the last four columns.

We find that in general the estimated earnings premia using propensity score methods are very similar to what we obtained with the OLS models. As before, all of these countries except for Bolivia exhibited positive and significant treatment effects for higher education degree completion. The treatment effects, however, are slightly smaller in Ghana and Vietnam when we use propensity score methods relative to OLS. Nonetheless, our main conclusion from the table is that the sizable earnings premia for higher education degree completion in most of these countries is not due to the parametric approach used in OLS. We tested for robustness of the propensity score methods to different

matching algorithms (such as kernel and increasing the nearest neighbours) and found that the results were very similar, and thus only report these in this study.

7. Analysis: social heterogeneity in the earnings premia from higher education degree

In the approaches presented in the previous sections, the effect of completing a higher education degree is assumed to be a constant parameter and invariant across all labour force participants. In this section, we test whether the earnings premia vary according to social origin and the likelihood of obtaining a degree. By doing so, we are able to examine whether the benefits from higher education are related to social mobility in each country.

We used two different approaches to examine heterogeneous effects in our models. First, we interacted the higher education degree variable with the family socio-economic status variables and added the interaction terms to the earnings Equation (3):

$$\ln(y) = \delta_0 + \delta_1 ED + \delta_2 EDxSES + \delta_3 \mathbf{X} + u \tag{8}$$

where *EDxSES* = interactions of *ED* with the two dummy variables for family socio-economic status. If either of the two interaction variables were statistically significant, then it would suggest that the earnings premia from a higher education degree vary by socio-economic status of the family. Likewise, insignificant interaction effects would show that the financial benefits from completing a degree are relatively uniform across social origins.[6]

In our second approach to examining social heterogeneity, we estimate heterogeneous treatment effects using the stratification-multilevel methodology introduced by Xie, Brand, and Jann (2012) to determine if the average treatment effect varied by propensity score. This method stratifies each sample based on the propensity scores and estimated separate average treatment effects using propensity score matching within each strata. For consistency across the nine countries, we used the same propensity score groupings for each country.[7] Because individuals with higher propensity scores for higher education are, on average, expected to come from more privileged social backgrounds, the measures of average treatment effects across propensity score levels provides more evidence about social mobility in these nine countries. If people from the highest social origins benefit the most from higher education degrees, we should observe the largest coefficients for individuals in the highest propensity score ranges. The findings from these two approaches are summarised in Table 7.

In eight of the nine countries that we considered, the interaction effects between higher education degree completion and family socio-economic status are statistically insignificant. The only exception to this pattern was for Laos, where the negative interactions for middle- and upper-socio-economic status categories show that the earnings premia are actually higher for those from disadvantaged family backgrounds. In all of the other instances, we find no evidence that social origin as reflected in family socio-economic status influenced the earnings premia from a degree.

Similar to the interaction effects analysis, the heterogeneous treatment effect analysis do not reveal any statistically significant pattern in the earnings premia by the person's likelihood of earning a degree (reflected in their propensity scores). In other words, we find no evidence of negative selection, such that individuals less likely to obtain a degree go on to benefit most from a degree. This conclusion holds for Laos, where we no longer find evidence that individuals from disadvantaged origins benefit more from higher education. Taken together, the benefits from higher education appear to be fairly consistent across social origins in the nine countries.

Table 7. Selected tests of heterogeneous effects of higher education on earnings

	Southeastern Europe			Latin America		Sub-Saharan Africa		Southeast Asia	
	Armenia	Georgia	Macedonia	Bolivia	Colombia	Ghana	Kenya	Laos	Vietnam
Heterogeneity: SES									
ED	0.855**	1.361*	−0.187	0.752**	0.170	2.098***	2.624***	5.032***	0.805**
	(0.297)	(0.538)	(0.844)	(0.265)	(0.969)	(0.600)	(0.643)	(0.935)	(0.262)
ED x SES2	−0.207	−0.400	0.792	−0.604	1.439	−0.889	−0.920	−1.992+	0.446
	(0.313)	(0.556)	(0.890)	(0.418)	(1.028)	(0.695)	(0.674)	(1.032)	(0.280)
ED x SES3	−0.117	−0.799	0.832	−0.640	0.441	−0.285	0.123	−5.916**	0.281
	(0.316)	(0.536)	(0.912)	(0.930)	(1.501)	(0.718)	(0.840)	(2.203)	(0.584)
Heterogeneity: Propensity Scores									
PS ≤ 0.20	0.677	0.915	0.611	0.084	0.949+	1.242***	1.572***	3.538***	0.805**
	(0.470)	(0.790)	(0.482)	(0.459)	(0.555)	(0.258)	(0.405)	(0.645)	(0.211)
0.20 < PS ≤ 0.40	0.814***	0.997***	−0.271	0.110	1.887**	0.300	1.925**	1.519	1.387**
	(0.161)	(0.335)	(0.384)	(0.412)	(0.677)	(0.896)	(0.577)	(1.296)	(0.209)
0.40 < PS ≤ 0.60	0.360+	0.611*	1.691+	0.652	1.649+	– –	2.547+	7.174*	0.151
	(0.211)	(0.289)	(0.870)	(0.758)	(0.857)		(1.409)	(3.641)	(0.365)
0.60 < PS ≤ 0.80	0.816***	0.890***	−0.628	0.728	– –	– –	– –	– –	0.861+
	(0.150)	(0.219)	(0.540)	(0.772)					(0.468)
Slope	0.010	−0.017	−0.267	0.235	0.437	−0.943	0.417	−0.598	−0.074
Intercept	0.686*	0.889	0.655	−0.233	0.637	2.185*	1.142	3.980*	1.095***

Notes: Dependent variable in each model = log of weekly wages in home country's currency. Coefficients represent effect of higher education on log of weekly earnings, standard errors are shown in parentheses. Data are weighted to account for the stratified sampling design. ED = 1 if completed bachelor's degree or higher, else 0. ED x SES2 = interaction of higher education with middle SES. ED x SES3 = interaction of higher education with upper SES. Propensity score heterogeneity tests conducted with '*hte*' command in Stata 15. Propensity scores were stratified by increments of 0.20. *** p < .001, ** p < .01, * p < .05, + p < .10 (two-tailed tests).

8. Discussion and conclusion

In this article, we examined higher education access and outcomes by observable social origin characteristics in nine low- and middle-income countries. Based on our analyses, we can comment on the extent to which higher education facilitates social mobility in each of the countries. As we have noted throughout this article, higher education systems can promote social mobility by making it accessible for people of all social origins and can provide the same benefits regardless of those social origins.

We find evidence that higher education systems hinder social mobility because individuals from disadvantaged origins are less likely to go to university and earn a higher education degree. In particular, we find that parental education is strongly associated with degree attainment in all nine countries. Parental socio-economic status and number of economic shocks are also associated with degree attainment in most countries. These findings are consistent across countries with different economic and regional profiles. These findings draw attention to policy accessibility-enhancing interventions that target upper-secondary school students and university students from disadvantaged backgrounds. Recent policy innovations in the United States, include targeting marginalised student groups with intensive student support and counselling in upper-secondary school, targeted financial aid, and programmatic support once on university campus (Pascarella & Terenzini, 2005).

In terms of labour market benefits from a higher education degree, we find that the average person with a higher education degree in these countries benefits substantially from larger earnings. In our analysis of heterogeneous benefits across social origins, we do not find support for the negative selection hypothesis that individuals from disadvantaged origins are most likely to benefit from a higher education degree. Rather, we find that the earnings premia from a degree are comparable within a country across individuals of different social origins and likelihood of obtaining a degree. These findings suggest that higher education contributes to social mobility by providing similar benefits.

In conclusion, we can say that higher education systems deter social mobility by being less accessible to disadvantaged individuals but promote social mobility by providing similar financial benefits regardless of social origins. We recommend further country-specific research on higher education interventions that support individuals from disadvantaged social origins.

Acknowledgments

Two anonymous reviewers, Javier Botero, Sangeeta Goyal, Toby Linden, and Francisco Marmolejo provided detailed comments on earlier drafts. Ji Liu provided research support. We are grateful to Sean Kelly and Marc DeFrancis for detailed edits and comments on an earlier draft. We also thank Luis Benveniste, Harry Patrinos, Halsey Rogers, Sebastian Taborda, Namrata Tognatta, and seminar participants at the World Bank for comments and discussions. The findings, interpretations, and conclusions expressed in this study are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent. The original data are freely downloadable from the World Bank's STEP Skills and Measurement Program website (http://microdata.worldbank.org/index.php/catalog/step). The authors are happy to share the data and code with bona fide researchers. Please email the authors for access to the data and code.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was financially supported by the World Bank as a part of the background papers commissioned for the World Development Report 2018.

Notes

1. The economic literature on 'sorting' may consider both the economic and sociological explanations (Fernandez, 2001). That is, sorting models consider that ability and motivation affect sorting into higher education by reducing the cost of education, *and* that social origins affect the preferences and likelihood of sorting into higher education. Sorting models may also consider heterogeneous returns to education.
2. We did not include three STEP participants: Sri Lanka, Ukraine, and Yunnan province in China. Sri Lanka is not included because we would have to consider a region (South Asia) with only one country. Ukraine is not considered because we ran into issues with small shares of male bachelor's degree holders in Ukraine. The Yunnan province was not included because the complex interplay of Yunnan with other provinces in China is beyond the scope of this article.
3. We construct a parental engagement variable that measures the quality of time spent using the STEP question: 'When you were attending primary school, did either of your parents/guardians actively keep themselves informed of your exam/test results or grades?' Response choices include 'Yes, always or almost always', 'Yes, sometimes', and 'No, never or almost never'. Previous studies using similar measures of quality of time as an indicator for parental engagement and investment in education include Sandberg and Hofferth (2001). We acknowledge the possibility of reference bias arising because of retrospective reporting. We have compared the distribution of socio-economic status at age 15 to the distribution of current assets as part of our preliminary analysis. Using information on dwelling characteristics and types of assets, factor analysis was used to create an asset index for each of the countries in the sample. Measures of assets and dwellings with extremely skewed distributions, agricultural assets, and those showing low factor loadings were excluded from the asset index. We find no evidence that the retrospective data are upwardly biased.
4. For Ghana, Kenya, and Macedonia, we did not create regional dummy variables because the STEP survey data did not support them.
5. PSM models were implemented using the '*teffects*' command in Stata 15. We used a nearest neighbour matching (one or more) and a caliper of 0.05. It should be noted that the teffects command for PSM in Stata 15 jointly estimates the propensity scores and the ATE/ATT, and thus does not reduce the sample for common support as is true for the Stata command psmatch2 in earlier versions of the programme.
6. Consistent with the sorting idea, our results show that individuals with more highly-educated parents were themselves more likely to obtain a degree. But despite the sorting, we found no strong evidence that those from lower social origins have different earnings premia than others, so once they have been sorted the labour market gains are the same.
7. The heterogeneous treatment effects were estimated in Stata 15 using the '*hte sm*' routine developed by Jann, Brand, and Xie (2010).

References

Altbach, P. (2016). *Global perspectives on higher education*. Baltimore: Johns Hopkins University Press.

Becker, G. (1975). *Human capital: A theoretical and empirical analysis, with special reference to education*. Chicago: University of Chicago Press.

Becker, G. (1993). *A treatise on the family*. Cambridge, MA: Harvard University Press.

Belley, P., & Lochner, L. (2007). The changing role of family income and ability in determining educational achievement. *Journal of Human Capital*, *1*, 37–89.

Blau, P., & Dudley, O. (1967). *The American occupational structure*. New York: Wiley and Sons.

Bourdieu, P. (1977). Cultural reproduction and social reproduction. In J. Karabel & A. Halsey (Eds.), *Power and ideology in education* (pp. 487–511). Oxford: Oxford University Press.

Brand, J., & Xie, Y. (2010). Who benefits most from college? Evidence for negative selection in heterogeneous earnings premia from higher education. *American Sociological Review*, *75*, 273–302.

Cameron, C., & Trivedi, P. (2010). *Microeconometrics using stata*. (Revised ed.) College Station, TX: Stata Press.

Carnoy, M., Froumin, I., Loyalka, P. K., & Tilak, J. B. (2014). The concept of public goods, the state, and higher education finance: A view from the BRICs. *Higher Education*, *68*, 359–378.

Carnoy, M., Loyalaka, P., Dobryakova, M., Dossani, R., Froumin, I., Kuhns, K., … Wang, R. (2013). *University expansion in a changing global economy: Triumph of the BRICs?* Stanford, CA: Stanford University Press.

Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, *94*, S95–S120.

European Commission. (2009). *Labour market in the former Yugoslav Republic of Macedonia*. Brussels: Author.

Fernandez, R. (2001). *Sorting, education and inequality* (NBER Working Paper 8101). Cambridge, MA: National Bureau of Economic Research.

Ferreyra, M., Ciro, A., Javier, B., Paz, H., & Urzúa, S. (2017). *At a crossroads: Higher education in Latin America and the Caribbean*. D.C.: World Bank Group.

Gerber, T., & Cheung, S. Y. (2008). Horizontal stratification in postsecondary education: Forms, explanations, and implications. *Annual Review of Sociology*, *34*, 299–318.

Heckman, J., Humphries, J., & Veramendi, G. (2016). *Returns to education: The causal effects of education on earnings, health, and smoking* (NBER Working Paper 22291). Cambridge, MA.

Hu, A., & Hibel, J. (2014). Changes in college attainment and the economic returns to a college degree in urban China, 2003–2010: Implications for social equality. *Social Science Research*, *44*, 173–186.

Husén, T. (1987). *Higher education and social stratification: An international comparative study*. Paris: UNESCO, International Institute for Educational Planning.

Jann, B., Brand, J., & Xie, Y. (2010). HTE: Stata module to perform heterogeneous treatment Effect Analysis. Retrieved from http://ideas.repec.org/c/boc/bocode/s457129.html.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Majgaard, K., & Mingat, A. (2012). *Education in Sub-Saharan Africa: A comparative analysis*. Washington, DC: World Bank.

McMahon, W. (2009). *Higher learning, greater good: The private and social benefits of higher education*. Baltimore: Johns Hopkins Press.

Mincer, J. (1974). *Schooling, experience, and earnings*. New York: National Bureau of Economic Research.

Morales, G. C. (2010). The challenge of progressive change under Evo Morales. In K. Weyland, R. Madrid, & W. Hunter (Eds.), *Leftist governments in Latin America: Success and shortcomings* (pp. 57–76). New York: Cambridge University Press.

Pascarella, E., & Terenzini, P. (2005). *How college affects students: A third decade of research*. San Francisco: Jossey-Bass.

Pierre, G., Sanchez, M. L., Valerio, A., & Rajadel, T. (2014). *STEP Skills Measurement Surveys: Innovative tools for assessingskills*. Social Protection and Labor Discussion Paper 1421. Washington, DC: The World Bank.

Post, D., Clipper, L., Enkhbaatar, D., Manning, A., Riley, T., & Zaman, H. (2004). World Bank okays public interest in higher education. *Higher Education*, *48*, 213–229.

Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. *Education Economics*, *26*, 445–458.

Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55.

Sandberg, J., & Hofferth, S. (2001). Changes in children's time with parents: United States, 1981-1997. *Demography*, *38*, 423–436.

Sandefur, G., Meier, A., & Campbell, M. (2006). Family resources, social capital, and college attendance. *Social Science Research*, *35*, 525–553.

Schofer, E., & Meyer, J. W. (2005). The worldwide expansion of higher education in the twentieth century. *American Sociological Review*, *70*, 898–920.

Shah, M., & McKay, J. (2018). *Achieving equity and quality in higher education: Global perspectives in an era of widening participation*. London: Palgrave MacMillan.

Shavit, Y., Arum, R., & Gamoran, A. (2010). *Stratification in higher education: A comparative study*. Stanford, CA: Stanford University Press.

Sullivan, A., Parsons, S., Wiggins, R., Heath, A., & Green, F. (2014). Social origins, school type and higher education destinations. *Oxford Review of Education*, *40*, 739–763.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, *26*, 24–36.

Toutkoushian, R., & Paulsen, M. (2016). *Economics of higher education: Background, concepts, and applications*. The Netherlands: Springer.

Toutkoushian, R., Shafiq, M. N., & Trivette, M. (2013). Accounting for risk of non-completion in private and social rates of return to higher education. *Journal of Education Finance*, *39*, 73–95.

UNESCO. (2008). *The current situation of higher education in Lao PDR*. Macao SAR: Author.

UNESCO. (2011). *International Standard Classification of Education ISCED 2011*. Montreal: Author.

Xie, Y., Brand, J., & Jann, B. (2012). Estimating heterogeneous treatment effects with observational data. *Sociological Methodology*, *42*, 314–347.

Appendix

Table A1. Variable definitions

Variable Name	Definition
Wage	Average reported earnings per week in domestic currency (plus 1).
Ln(Wage)	Natural log of weekly wage
Higher Education (ED)	1 if highest level of education is bachelor's degree or higher, else 0. Respondents who were still in school were omitted.
Yrs Experience	Years of potential labour market experience, defined as age minus approximate age at the end of the respondent's education.
Birth Order	Order of respondent's birth among siblings (1 = first born, 2 = second born, and so forth).
Number of Siblings	Number of older and younger brothers and sisters
Female	1 if female, else 0
Lower Family SES	Family socio-economic status at age 15 is categorised as 'Low'
Middle Family SES	Family socio-economic status at age 15 is categorised as 'Middle'
High Family SES	Family socio-economic status at age 15 is categorised as 'High'
Economic Shocks	Number of economic shocks experienced by family before age 15
Parent Ed: Primary	1 if maximum education of either parent is primary, else 0
Parent Ed: Secondary	1 if maximum education of either parent is secondary, else 0
Parent Ed: Higher	1 if maximum education of either parent is higher education degree, else 0
Attended Preschool	1 if respondent attended preschool, else 0
Parents Engaged Ed	1 if parents were 'highly engaged' in respondent's education, else 0
Region	Series of dummy variables based on region or district

Notes: All variable except Higher Education (ED) were defined and constructed by the STEP Skills Measurement survey team. The Higher Education variable was derived from the STEP survey questions regarding educational attainment.