Why are urban areas more educated?

Rowan Shi

2019-01-15

When asked, most people might guess that the population in urban areas is more educated than the population in rural areas. There are two possible source of this discrepancy:

- for some reason, those born in urban areas achieve higher levels of education; or
- those who achieve higher levels of education move to urban areas (or those who achieve lower levels move to rural areas).

In other words, the difference could be attributable to those native to the urban (or rural) area, or to those who move there later in life. The distinction is important: common wisdom and policy prescribe education as a way out for those living in rural areas, which tend to be poorer.

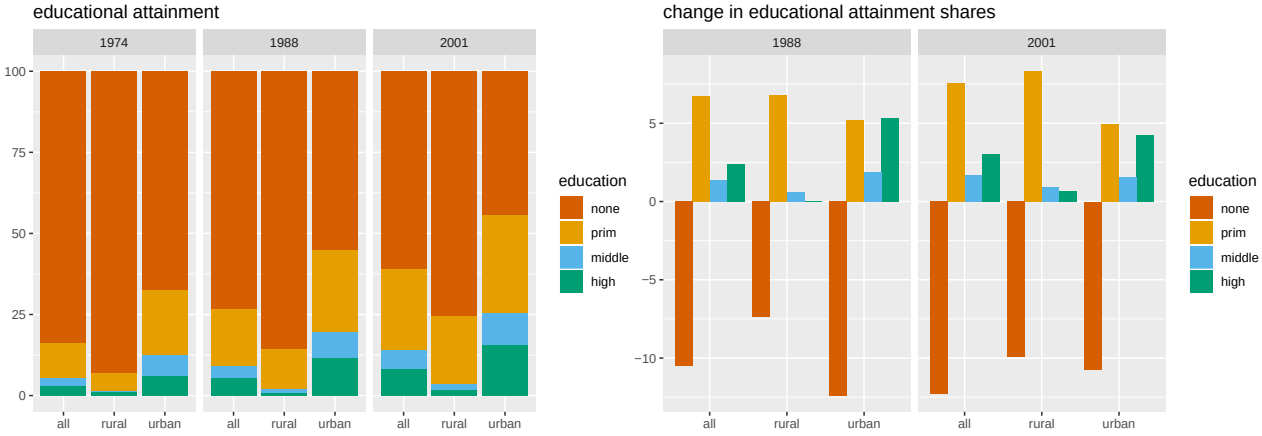
In this exercise, I use Honduras census data (taken in 1974, 1988, and 2001) to investigate this question.

A first descriptive look

Educational attainment

To begin, let's take a look at educational attainment of those living in urban and rural areas – where educational attainment is the highest degree a person has completed.

For each census year, I calculate the percentage of people who have no diploma, at least a primary school degree, at least a middle school degree, or at least a high school degree. (Since very few people in the sample have completed higher education, I do not include those degree categories here.) The breakdown is down the whole population, as well as for rural and urban subsamples. For clarity, the changes of these shares for 1974-1988 and 1988-2001 are also pictured on the right.



There are several patterns to notice immediately:

- in all years, educational attainment is generally low, with most people having no diploma at all;
- in all years, the population of rural areas is less educated than the population of urban areas; and
- over time, educational attainment has been increasing for both rural and urban areas.

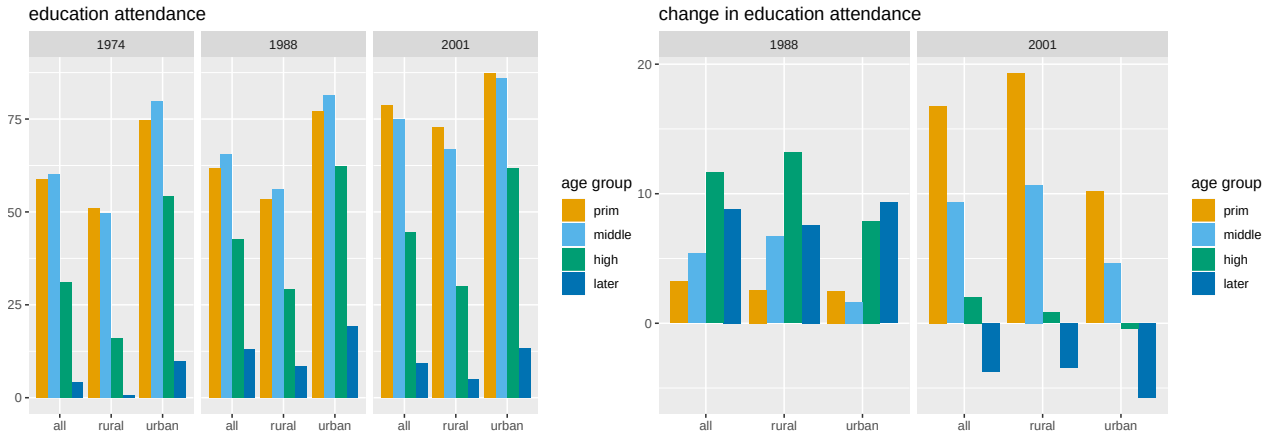
A closer look reveals that, while both types of areas have seen increasing educational attainment, there is a difference in the types of degrees. *The main change in rural areas has been primary school certification. On the other hand, urban areas have also seen increases in share of population holding middle and high school diplomas.*

Educational attendance

Working towards the goal of separating the two explanations, let's now turn to educational attendance, calculated as the percentage of people who report they are in school.

The census unfortunately does not ask what level of school the respondant is attending, so instead I divide the population into age categories. The Honduras educational system consists of 6 years for primary school, 3 years for middle school, and 3 years for high school. Schooling generally starts at age 5. Consequently, the age categories are defined based on a child who begins school at age 5 and completes each degree in sequence without skipping or repeating years. For example, a respondant who is 7 years old would be classified as "primary school aged".

Observations are divided into age groups, and the share of each age group that is attending school is shown in the graphs below. Again, the changes over time are displayed on the right.



Here, we see that attendance is also rather low (especially after primary school), but that it's higher in urban areas and also over time. Similarly to educational attainment, the composition of the changes over time for educational attendance vary across urban and rural areas. *In particular, from 1974 to 1988, the share of middle and high school aged children increased by much more in rural areas than in urban areas.*

The first clue

As uncovered above, there was a increase in rural middle school and high school attendance over the 1974 to 1988 period, which was more pronounced than the change in urban areas. However, from 1988 to 2001, urban areas experienced an increase in residents holding middle and high school diplomas. This descriptive analysis is suggestive that the second explanation could have some bite – educated people move to urban areas. If that is the case, common sense and policy are more justified: more education can help those born in rural areas move to urban areas.

Regression analysis

While the descriptive analysis is provocative, it doesn't guarantee that implementing better education in rural areas means more migration towards urban areas. For example, those who stay in school could be different from the rest of the population, and they would have gone to school to move to the city no matter how tough the educational conditions. Put another way, those who stay in school longer are a special self-selected population. In this case, it could be possible that improving education will have no impact on outcomes for others in the population.

Experimental design

The broad sample I consider is the set of individuals questioned in the 2001 census, while the outcome of interest is whether they live in an urban or rural area. In particular, the experiment should examine whether

education (measured as years of schooling) has an effect on the outcome, especially for those born in rural areas.

To try to establish a causal link, I will use the educational conditions of a person's birthplace. In particular, the graphs above belie considerable variation at the municipality level of the change in educational attendance.

I begin by calculating the same statistic as before, changes in educational attendance by age group, but at the municipality level. The censes provide 96 municipality areas, most containing both rural and urban areas. The table below shows the standard deviation in these changes, by age group.

```
     prim    middle      high     later
11.150858 14.023124 11.512653  3.197009
```

Across municipalities, the standard deviation in the change over 1974 to 1988 of school attendance in primary school aged children is 11.1508581 percentage points. It is similarly high for middle and high school. In other words, the educational climate in some municipality areas was improving much faster than others.

I will use these municipality-level changes as instruments for the individual education variable. Since individuals cannot affect where they were born, or the conditions of the area, they serve as quasi-random components of an individual's eventual years of schooling.

Due to this choice of instrument, the sample is restricted to exclude anyone for whom the 1974-1988 period occurred after their educational period (ages 5 - 17). The sample also excludes anyone who is still in their educational period. The final sample therefore includes all respondants aged 18 - 44.

Regression

The final regression is a linear probability model with 1 indicating the person lives in an urban area, as below:

$$\text{urban} = \alpha_0 + \alpha_{a_1}\text{age} + \alpha_{a_2}\text{age}^2 + \alpha_{bpr}\text{bpr} + \alpha_e\text{school} + \alpha_{bpr,e}(\text{bpr} * \text{school}) + \varepsilon$$

where bpr measures the "rural-ness" of the person's birthplace. The censes only provides the municipality area in which respondant was born, not whether they were born in an urban or rural area. This information is therefore proxied with the share of population in each municipality block living in a rural area.

The constant α_0 varies with gender, maritial status, and the interaction of the two.

Again, the variable school is instrumented with the four variables representing changes in educational attendance, one for each education age group. The interaction term $\text{bpr} * \text{school}$ is instrumented with bpr interacted with the four instrument variables.

Results

The results of the regression are as follows:

	live_urban	
school_years	0.0507***	(14.47)
c.bp_ruralness × c.school_years	0.0596***	(17.14)
age	-0.00440**	(-2.81)
c.age × c.age	0.000153***	(5.85)
1.female	0.0140**	(2.67)
1.married	-0.00234	(-0.61)
1.female × 1.married	0.00316	(0.51)
bp_ruralness	-0.687***	(-25.45)
_cons	0.438***	(14.12)
N	139663	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The first row shows that each additional school year adds about a 5.4 percentage point chance the individual lives in an urban area in 2001. The second line addresses the original question: if the person was born in a rural area, how much more does education increase this likelihood?

How do we interpret the 0.055 coefficient? Consider two people, one whose birth municipality had a 70-30 urban-rural breakdown (A), and the other whose birth municipality had a 60-40 split (B). If A and B are identical along the other dimensions in the regression (years of schooling, age, gender, and maritial status), then each year of schooling increases B's likelihood of living in an urban area by 0.55 percentage points more than it does for A.

The remaining rows show coefficient values for the controls. Women are more likely to live in urban areas than men, while maritial status does not seem to have much of an effect. Finally, those both in a more rural area have a lower chance of living in an urban area – consistent with the idea that rural-born individuals "need" something like education to make their way out.

The final take

From this quick dive into the Honduras census data, we observed that urban areas tend to have more educated residents than rural areas, that this pattern is strengthening over time, but that rural areas seem to be catching up in terms of educational attendance. After a closer look, there seems to be support that rural-born educated people are moving to urban areas more than their rural-born uneducated counterparts.

Since we used variation in a person's birthplace (and that birthplace's educational environment) as the source of variation in our experimental design, there is more confidence in the interpretation that this link is causal – that is, more education *causes* or helps rural-born people move to urban areas.

Notes on limitations

This exercise is limited by data availability. Ideally, the sample should consist of people for whom the 1974-1988 period occurred before they became school age; that is, before they were 5 years old. However, with the current data, it is not possible to use such a sample since it must also be that those included in the sample are over 17. These two restrictions cannot both be achieved using the 2001 census. Honduras conducted its next census in 2013; this data would be ideal, but the microdata has been elusive.

In addition, if the respondants provided details on whether they were born in a rural or urban area, then the "rural-ness" of birthplace municipality would not be needed to proxy for this measure. Likewise, providing information on which level of school those in school are attending would do away with the education age group approximations.

Finally, the regression estimated is a linear probability model. While a model like logit or probit could be estimated, these are more sensitive to specification and suffer from inconsistency in these situations. The linear probability model performs worse near values of 0 or 1, but does well near the center. The coefficient results are therefore better interpreted as "average effects" over the sample population.

References

Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.1 [dataset]. Minneapolis, MN: IPUMS, 2018. https://doi.org/10.18128/D020.V7.1