



Munich Personal RePEc Archive

Skill Acquisition in the Informal Economy and Schooling Decisions: Evidence from Emerging Economies

Semih Tumen

Central Bank of the Republic of Turkey

16. March 2015

Online at <http://mpa.ub.uni-muenchen.de/62909/>

MPRA Paper No. 62909, posted 17. March 2015 05:07 UTC

Skill Acquisition in the Informal Economy and Schooling Decisions: Evidence from Emerging Economies*

Semih Tumen [†]

Central Bank of the Republic of Turkey

March 16, 2015

Abstract

Informal jobs offer skill acquisition opportunities that may facilitate a future switch to formal employment for young workers. In this sense, informal training on the job may be a viable alternative to formal schooling in an economy with a large and diverse informal sector. In this paper, I investigate if these considerations are relevant for the schooling decisions of young individuals using panel data on 17 Latin American countries as well as micro-level data for Turkey. Specifically, I ask if the prevalence of informal jobs distort schooling attainment. I concentrate on three measures of schooling outcomes: (1) secondary education enrollment rate, (2) out-of-school rate for lower secondary school, and (3) tertiary education graduation rate. I find that the secondary education enrollment rate is negatively correlated with the size of the informal economy, while the out-of-school rate is positively correlated. This means that informal training on the job may be crowding out school education in developing countries. The tertiary education graduation rate, however, is positively correlated with the size of informal sector, which implies that a large informal economy induces college attendance for those who are more likely to succeed. Policies that can potentially affect the size of the informal sector should take into consideration these second-round effects on aggregate schooling outcomes.

JEL codes: E26, I21, J24, O17.

Keywords: Informal economy; skill acquisition; schooling outcomes; Latin America; Turkey.

*I thank Franco Peracchi (the editor) and two anonymous referees for very helpful comments and suggestions. The views expressed here are of my own and do not necessarily reflect those of the Central Bank of the Republic of Turkey. All errors are mine.

[†]semih.tumen@tcmb.gov.tr. Research and Monetary Policy Department, Central Bank of the Republic of Turkey, Istiklal Cad. No:10, 06100 Ulus, Ankara, Turkey.

1 Introduction

Informal jobs offer valuable opportunities for young unskilled workers in terms of skill acquisition and career advancement.¹ To be specific, they serve as a stepping stone to formal jobs offering better conditions in terms of pay, insurance, job security, and conditions at work. Although the existence of a large informal sector may distort tax collection and reduce productivity, informal job opportunities may in fact be welfare enhancing for the young and unskilled workers in developing countries, because these jobs help workers gain expertise and build professional networks for improving their career prospects.

My starting point in this paper is the idea that informal job opportunities may induce young individuals to leave the school early. In other words, existence of a large and diverse informal sector may distort the schooling outcomes in a country. To understand the micro foundations of this argument, think of a version of the dynamic discrete schooling choice model à la [Heckman and Navarro \(2007\)](#). Each year the student decides whether to stay at school for another year or leave the school early and start working. To make this choice, he compares the expected present discounted value of staying at school versus that of start working at each decision node. For younger students, the option value of another year in school is normally large and the value of leaving school at that age is potentially low. When there is a large informal sector offering an alternative path to a good job, however, the expected present discounted value of leaving school early goes up and some students may choose to dropout or not to enroll further. In other words, informal jobs may inflate the option value of leaving the school early and, thus, may distort aggregate schooling outcomes in an economy.²

The main goal of this paper is to test whether this conjecture is supported by data. To achieve this goal, first, I perform a cross-country analysis using a panel of 17 Latin American countries and, then, I analyze micro-level labor market data from Turkey. In the cross-country analysis, I mainly estimate the sign and magnitude of the correlation between the size of the

¹See, for example, [Bosch and Maloney \(2010\)](#), [Cunningham and Salvagno \(2011\)](#), and [Tumen \(2012\)](#).

²There is another explanation along the lines of [Cunningham and Salvagno \(2011\)](#). They find that formal jobs discourage education, because once the individual enters the formal path, the probability of going back to school declines significantly. For informal jobs, however, this is not true: the informal job is likely temporary and, so, the probability of returning to school is not that much lowered for young individuals. In this sense, informal employment and school education are alternative to each other.

informal sector and several aggregate schooling outcomes controlling for per capita GDP, the level of public investment in education, the degree of income inequality as well as country and time fixed effects. I focus on three aggregate measures of schooling outcomes: (1) secondary education gross enrollment rate, (2) out-of-school rate for children of lower secondary school age, and (3) tertiary education graduation rate. I show that a percentage point increase in the size of the informal economy may lead to around 2 percentage point decline in the secondary education enrollment rate and around 1.3 percentage point increase in the out-of-school rate. This suggests that the size of the informal sector may be distorting schooling outcomes. Interestingly, I find a positive correlation between the size of the informal economy and the tertiary education graduation rates. The interpretation is as follows. The tertiary education enrollment rates are low, on average, in Latin American countries. But, the college graduation rates increase with the size of the informal sector, because those who enroll are mostly the ones who are more likely to do well in college. This is the analogue of the survival of the fittest idea in college education [see, e.g., [Cameron and Heckman \(1998\)](#)]. In the micro-level analysis for Turkey, I perform the microeconomic counterparts of the cross-country regressions. I confirm that the results obtained from the cross-country analysis are robust.

This is the first paper in the literature arguing that the availability of informal job opportunities—which may lead to a potential transition to formal jobs—can diffuse into the *ex ante* option value of leaving the school and, therefore, distort the schooling outcomes. This paper is closely related to several papers in the literature. [Monk, Sandefur, and Teal \(2008\)](#) show using data from Ghana that informal apprenticeship is an important institution providing training and it is undertaken by those with junior high school or lower levels of education. They find that for those who did apprenticeships but have no formal advanced degree, the apprenticeship training increases earnings by almost 50 percent. This supports my main hypothesis that training opportunities in the informal sector may lead to an increase in the option value of leaving the school early and, therefore, induce dropout or non-enrollment behavior. Another closely related paper is [Gunther and Launov \(2012\)](#), who argue using data from Ivory Coast that returns to schooling and experience is high for formal jobs and low for informal jobs. If

the individual is less likely to succeed in school, then he may choose to dropout early and start accumulating experience in the informal sector, which will potentially be rewarding in case of a switch to formal sector. In this sense, an informal job raises the earnings potential of those who are less likely to succeed in school and may let them choose to leave school early. Similarly, [Cano-Urbina \(2014\)](#) shows using survey data from Mexico that informal jobs generate extra value, while waiting to find a formal sector job. This also justifies the validity of the hypothesis proposed in this paper.³

This paper is different from the others in that it investigates the role of informal job availability on the schooling decisions of young-unskilled workers. For those who are less likely to succeed in school, the option value of receiving an additional year of schooling may be lower than the option value of dropping out and start working at an informal job providing skill acquisition opportunities. Such a mechanism may distort secondary school attendance rates and the continuation rates to college education. College graduation rates, in turn, may go up with the size of the informal sector, because those who are more likely to succeed in college will tend to receive education beyond high school.

Section 2 motivates the theoretical background of the paper. I demonstrate that the main hypothesis of the paper can be placed into a version of the [Ben-Porath \(1967\)](#) model of life cycle human capital accumulation and earnings. Specifically, I show that the existence of informal job opportunities allowing workers to accumulate skills and experience may lead the young individuals to receive less schooling. In terms of the terminology of the Ben-Porath model, such opportunities will lead to a shorter “period of specialization.” It is worthwhile to note that, in the Ben-Porath model, the option value of leaving the school is reflected in the initial (after-school) human capital level (i.e., the ability to earn). Availability of informal jobs offering skill accumulation opportunities enhances the ability to earn and induce dropping out of school.

³The academic interest in the issue of informal employment is not specific to economists. In other disciplines—such as sociology, psychology, and industrial relations—there are several studies that are related to the hypothesis posed in the current paper. For example, [Paternoster, Bushway, Brame, and Apel \(2003\)](#) and [Apel, Paternoster, Bushway, and Brame \(2006\)](#) investigate whether work intensity in formal versus informal jobs is associated with problem behavior. [Zapata, Contreras, and Kruger \(2011\)](#) and [Rammohan \(2012\)](#) investigate the work-school tradeoff within the context of child labor. [Dancer and Rammohan \(2007\)](#) try to answer the question whether maternal education has any effect on the work-school tradeoff. [Amuedo-Dorantes \(2004\)](#) analyzes the poverty implications of work-school tradeoff with a particular emphasis on informal employment.

The rest of the paper is organized as follows. Section 3 describes the data used, explains the econometric methods employed, and presents a detailed discussion of the results. Section 4 concludes.

2 Theoretical Motivation

The purpose of this section is to formally establish the view that any factor—other than school education—affecting the “earning capacity” should also diffuse into schooling decisions and alter schooling attainment. Increased training opportunities in the informal sector may lead to a rise in individuals’ expectations on their earnings potentials and may induce them to leave the school early. In this sense, informal on-the-job training can crowd out formal school education. I would like to explicitly mention at this stage that the theoretical formulation presented in this section is just for motivation purposes and should neither be perceived as a complete theoretical assessment of the problem nor it should be expected to map into the empirical analyses that Section 3 presents. I start with a baseline model, which is simple and stylistic. Then, I discuss the possibility of extending this baseline model toward several directions.

2.1 Baseline Model

The basic theoretical framework is a version of the [Ben-Porath \(1967\)](#) model of life cycle human capital accumulation and earnings. Time is continuous. Risk-neutral individuals maximize the present discounted value of their life-time earnings given by

$$\int_0^T e^{-rt} y(t) dt, \tag{2.1}$$

where $y(t)$ is the current period earnings, $r > 0$ is the interest rate, and T is the finite life-length. The maximization is subject to the following law of motion for human capital:

$$\dot{h}(t) = f(z, k(t), h(t)) - \delta h(t), \tag{2.2}$$

given the initial stock of human capital $h(0) = x$, where the notation \dot{h} describes the time derivative of the human capital stock, $0 \leq \delta < 1$ is the depreciation rate of the human capital stock, and $f(z, k(t), h(t))$ is the production of human capital as a function of the effectiveness ($z > 0$) in the production of human capital, the fraction of time ($0 \leq k(t) \leq 1$) devoted to human capital investment in each period, and the current stock of human capital ($h(t)$). I assume that $f(z, k(t), h(t)) = z [k(t)h(t)]^\gamma$, where $\gamma > 0$ is the returns to scale parameter. For algebraic simplicity, I ignore depreciation and set $\delta = 0$ in what follows.

The initial human capital stock (x) sets the initial conditions on the earning capacity of the individual. It is possible to think that x is an object that the individual forms expectations on. Formally, one can think that

$$x = \int x(q_s) dF(q_s), \quad (2.3)$$

where q_s is the capacity (or ability) to earn and $F(q_s)$ is the cumulative density describing the agent's beliefs on his earning capacity. Any factor that leads to a change in the individual's beliefs on his earning capacity will lead to a change in x . This will, in turn, affect the individual's human capital accumulation path. For example, existence of an established informal sector that can serve as a quick and effective on-the-job training path to a formal-permanent job may update these beliefs and may lead to an increase in x . I will come back to this point later.

The formula for the current-period earnings is given by

$$y(t) = Rh(t) [1 - k(t)], \quad (2.4)$$

where R is the rental rate of human capital. So, each individual splits his effort between the human capital investment and market work. Notice that the prices, r and R , are taken as given by the individuals. Thus, in this sense, I take a partial equilibrium stance. Next, I solve the maximization problem of the individual. The current-value Hamiltonian can be

constructed as

$$\mathcal{H}(k, h, \mu, t) = Rh(t) [1 - k(t)] + \mu(t)z [k(t)h(t)]^\gamma, \quad (2.5)$$

where $\mu(t)$ is the shadow value of human capital investment. The first-order condition for $k(t)$ is

$$R = \gamma\mu(t)z [k(t)h(t)]^{\gamma-1}. \quad (2.6)$$

Therefore, at the equilibrium, the marginal cost of human capital investment, R , equals the marginal return given by the right hand side of Equation (2.6). The law of motion for the shadow value of human capital investment, $\mu(t)$, is

$$\dot{\mu}(t) = r\mu(t) - R[1 - k(t)] - \gamma\mu(t)zk(t)^\gamma h(t)^{\gamma-1}, \quad (2.7)$$

which yields, after combining with Equation (2.6), that

$$\dot{\mu}(t) = r\mu(t) - R. \quad (2.8)$$

Equation (2.8) is a differential equation describing the evolution of the shadow value of human capital investment over time. The standard transversality condition,

$$\lim_{t \rightarrow T} e^{-rt} \mu(t)h(t) = 0, \quad (2.9)$$

has to hold, which obviously implies that $\mu(T) = 0$, for large T (i.e., for a long working life). Solving for this differential equation, I get

$$\mu(t) = \frac{R(1 - e^{r(t-T)})}{r}. \quad (2.10)$$

Following [Brown \(1976\)](#) and [Heckman, Lochner, and Taber \(1998\)](#), I assume that the working

life is long, i.e., $T \rightarrow \infty$, so, at the limit,

$$\mu(t) = \mu = \frac{R}{r}. \quad (2.11)$$

In words, when the working life is long, the shadow value of the investment in human capital is a constant and is given by R/r . Suppose r goes up exogenously. Then marginal cost of human capital investment exceeds the marginal return to it. To increase marginal returns, from Equation (2.6), investment in human capital $k(t)h(t)$ should go down. This must be the case since a higher interest rate means that future earnings will be discounted heavily and therefore investment should be worth less, justifying the decline in μ .

Next I derive the schooling choice in this framework. It is well known that a “period of specialization” arises in the Ben-Porath model, if the inequality

$$\mu(0)\gamma zx^{\gamma-1} > R \quad (2.12)$$

holds. The condition $\mu(t) = R/r$ holds for all t when T is large, so it must also hold for $t = 0$. Thus, this expression becomes

$$\gamma zx^{\gamma-1} > r. \quad (2.13)$$

Notice that this inequality comes from the first-order condition (2.6). In other words, if the marginal return to investing full time (i.e., $k(t) = 1$) in human capital exceeds the marginal cost of it at the beginning of life, then it is optimal to set $k(t) = 1$ until the equality is reached. Observe that $\gamma zx^{\gamma-1}$ declines over time since $h(t)$ goes up which makes the left hand side decline monotonically over time. To prove that $h(t)$ goes up monotonically, one can plug (2.11) into (2.6) which yields

$$k(t)h(t) = \left(\frac{\gamma z}{r}\right)^{1/(1-\gamma)}. \quad (2.14)$$

Then, I insert this into the law of motion for human capital accumulation (2.2) to obtain

$$\dot{h} = z \left(\frac{\gamma z}{r} \right)^{\gamma/(1-\gamma)}, \quad (2.15)$$

which is a positive constant. This completes the argument that h constantly increases over time. Thus, there exists a period of specialization, the length of which is denoted with s . At period s , the first order condition for $k(s)$ holds with equality. I interpret the interval $[0, s]$ as the period of “schooling” and s as the total years of schooling. Thus, at the end of the schooling period, by continuity, one should have

$$z\gamma h(s)^{\gamma-1} = r. \quad (2.16)$$

Next I derive an explicit formula for s . Observe that when $k(t) = 1$,

$$\dot{h}(t) = zh(t)^\gamma, \quad (2.17)$$

for $t \leq s$. The solution to this differential equation is

$$h(t) = [z(1-\gamma)t + x^{1-\gamma}]^{1/(1-\gamma)}. \quad (2.18)$$

Since this formula holds for $0 \leq t \leq s$, it must hold for $t = s$. Solving Equation (2.16) for $h(s)$ and substituting it into (2.18), I get

$$\gamma z [z(1-\gamma)s + x^{1-\gamma}]^{-1} = r. \quad (2.19)$$

Solving this equation for s gives

$$s = \frac{\gamma}{(1-\gamma)r} - \frac{x^{1-\gamma}}{(1-\gamma)z}. \quad (2.20)$$

Thus, total years of schooling, s , arises in the Ben-Porath model as a function of the interest rate (r), the degree of returns to scale in the production of human capital (γ), the ability to earn (x), and the ability to learn (z).

If the availability of informal jobs raises the earning potential x , then the years of schooling goes down. It is easy to see this algebraically. Clearly,

$$\frac{\partial s}{\partial x} = -\frac{1}{x^\gamma z} < 0. \quad (2.21)$$

In words, regardless of the magnitudes of the returns to scale and efficiency parameters, the optimal years of schooling goes down with the earning potential that training in informal sector would provide.

2.2 Possible Extensions

In this sub-section, I discuss the possibility of extending the baseline model, which is mostly stylistic, toward several directions. These extensions consist of incorporating (i) income shocks, (ii) family labor supply, (iii) two-sector labor market, and (iv) multidimensional skill accumulation to the baseline specification. Although I do not present explicit formulations of these extensions, I discuss in detail the potential implications and consequences of them.

2.2.1 Income shocks

The baseline model ignores the existence of both aggregate and idiosyncratic income shocks. It is well known that income shocks affect labor demand and wages.⁴ Income shocks may also have an impact on the tradeoff between schooling decisions and labor market participation of young individuals. When a negative (positive) aggregate shock hits the economy, the level aggregate labor demand falls (increases) and, thus, wage offers decline (rise). Since borrowing is costly, households—especially the low-income households—may choose to cut costs, including the schooling expenditures, which may increase the dropout rates. In addition, to compensate for diminished labor market returns for the adult members of the households, these dropouts might be pushed into the informal labor market.

There is also an effect operating in the opposite direction. A decline in aggregate wages also reduces the opportunity cost of schooling, which would lead to a decline in the school

⁴See, e.g., [Bertola \(1999\)](#) for an extensive review of the literature.

dropout rates. So, the net effect of an income shock on the tradeoff between schooling and work is largely ambiguous. The case of idiosyncratic shocks produces less ambiguity. Since these shocks are, by definition, idiosyncratic (i.e., independent of aggregate labor market conditions), the change in children’s time use is more predictable if the adult members of the household experience idiosyncratic shocks—say, an idiosyncratic job loss. Normally, the offsetting opportunity cost effect does not exist in such a case. However, the main problem is that idiosyncratic and aggregate shocks are generally hard to distinguish empirically, especially during recessions.

Separately, there is a large literature on the potential impact of income shocks on human capital accumulation patterns.⁵ This literature suggests that the uncertain nature of future income affects human capital investment decisions over the life cycle. These include schooling decisions as well as post-school human capital investment decisions. Thus, the existence of income shocks affect both the timing and type of human capital investment. It is well-known that life-cycle models with multiple layers of uncertainty cannot be solved analytically [see [Auerbach and Kotlikoff \(1987\)](#)]. Instead, simulation methods should be used to understand the predictions of this class of models, which is out of the scope of this paper.

Although I recognize that introducing income shocks may introduce some insights that are not present in the baseline specifications, the nature of such an extension would be mostly ambiguous and complicate the model in terms of analytical tractability. However, this is still a useful thought experiment and the effect of income shocks will be discussed further in the empirical analysis section.

2.2.2 Family labor supply

The baseline model assumes a representative-family framework; that is, the decisions are made as if there is a benevolent family planner optimally choosing all household-level variables including the schooling levels of the children. The representative-family framework is useful since (*i*) it yields analytically tractable solutions for the family-level maximization problem and

⁵See, e.g., [Baker \(1997\)](#), [Lillard and Reville \(1999\)](#), [Meghir and Pistaferri \(2004\)](#), [Cunha, Heckman, and Navarro \(2005\)](#), and [Kim \(2010\)](#).

(*ii*) it is a good approximation for many problems studied in the family economics literature [Becker (1993)].

However, there is also a large literature documenting the limitations of the representative-family approach. When individual-level concerns are introduced into a family-level model, the predictions of such a model may largely deviate from the predictions of the representative-family model. One potential channel that would introduce individual-level conflicts is intra-household bargaining [see, e.g., Lundberg and Pollak (1993)]. Another is collective labor supply [see, e.g., Chiappori (1992)].

Even in the absence of these rather complicated channels, decisions made by a household unit will not be the same as those made by individuals. Suppose, for example, that the household consists of two individuals: old (adult) and young (child). Suppose also that the child's decisions are also made by the adult. It could be possible that the size of the informal sector in an economy affects directly the type of the jobs that the adult can have. If the earnings of the adult member of the household are vulnerable to income shocks, then a negative income shock may induce the adult to withdraw the child from school and push into the labor force. Schafer (2006) documents the existence of such patterns using data from Kenya and Malawi.

When such a scenario is in effect, the size of the informal economy can potentially alter schooling decisions not because of a change in the ability to earn but because of a negative shock to parents' income. However, as I explain in Section 2.2.1, the predictions of a human capital accumulation model with income shocks will be ambiguous. The potential effects of income shocks will be discussed in much more detail in Section 3.

2.2.3 Two-sector model

The baseline model does not explicitly model formal and informal employment separately. Instead, it assumes that the underlying parameters for formal and informal employment are potentially different. Making such an assumption is equivalent to assuming that the labor market has a dual structure along the formal/informal divide. The duality argument is based

on the view that the fundamentals driving the labor market outcomes in both markets are substantially different and, thus, the two markets can be separately analyzed. Although the empirical tests of the duality hypothesis yield mixed results, it is generally accepted in the literature that the duality (or segmentation) assumption is not a bad approximation to reality.⁶

It is still possible to relax the segmented markets assumption and allow for employment in two sectors as formal employment and informal employment. When these two sectors are jointly modeled, one has to allow for unlimited transitions between formal and informal employment along the model horizon. This means that there will be two types of earnings: earnings in formal jobs and earnings in informal jobs. It is not plausible, however, to think that formal and informal employment are the sole labor market outcomes and workers freely travel from formality to informality and vice versa. Unemployment should also be incorporated into such a setup, because, for example, transitions from formal employment toward informality would only be justified if there is unemployment in between. The ideal setup for such a model will of course be a random search model.

In a full-fledged model, the two-sector structure will naturally arise due to the existence of taxes and regulations in the formal sector as in [Meghir, Narita, and Robin \(2014\)](#). Besides schooling and work decisions, the workers will choose a life-cycle trajectory for their sector of employment. The economy will be exposed to search frictions and search will be allowed both when unemployed and employed. The workers will receive competing offers from both sectors. The steady-state worker flows in each sector will characterize the long-run equilibrium solution for the model. Given these equilibrium worker flows, years of schooling could be solved for using backward induction. The main problem here is that the solution for the search model alone (without the inclusion of schooling decisions) can only be characterized using simulation methods [see, e.g., [Meghir, Narita, and Robin \(2014\)](#)]. Incorporating schooling decisions along with on-the-job human capital accumulation into such a structure will further complicate the model and will quickly take the focus away from an intuitive basis. Although I believe that a two-sector model can produce additional insights, it will bring enormous computational costs

⁶Papers including [Stiglitz \(1976\)](#), [Dickens and Lang \(1985\)](#), and [Heckman and Hotz \(1986\)](#) argue that the duality argument holds, at least partially. See [Magnac \(1991\)](#) for an opposite view.

that would go well beyond the objectives of this paper. It is also suspicious that the qualitative predictions of the baseline model will be substantially altered.

2.2.4 Multidimensional skills

The baseline specification does not explicitly model skill acquisition in the informal economy. The segmented markets assumption explained above simply presumes that the skills acquired in an informal job is perfectly substitutable to the skills accumulated in a formal job. So, once the individual drops out of school—say, without receiving a high school degree—to start working as an informal worker, he will accumulate human capital on the informal job as if he is employed in a formal job. As the degree of substitutability between skills accumulated in formal versus informal jobs goes down, the incentives to dropout early would also go down. In other words, if skills are multidimensional (i.e., formal and informal jobs require different types of skills), then the strength of the baseline model’s predictions would diminish, although the qualitative nature of the results would stand still.

In a companion paper, [Tumen \(2012\)](#) shows that the demand for informal jobs depends on the degree of skill accumulation intensity offered by informal jobs. If informal jobs help workers acquire skills, gain expertise, and build professional networks for boosting the chances to switch to a formal job, then the option value of a job in informal sector will be high since the probability of switch to a high-pay formal job will also be high. If, on the other hand, informal sector does not provide satisfactory training opportunities (i.e., the skills accumulated in informal jobs are not good substitutes with the skills accumulated in formal jobs), then demand for informal jobs will be low. Data suggest, however, that the degree of substitutability will likely be high in emerging economies—e.g., most of the Latin American countries, Turkey, and some of the South East Asian countries—because informal jobs are densely located in urban areas or regions with capital-intensive sub-sectors. If the informal jobs are heavily concentrated in rural or agriculture-dependent areas and sub-sectors with less physical capital requirements, then potential for advancement for an informal worker is slim [[Wahba \(2009\)](#)].

Next, I test the main hypothesis proposed by the baseline model first using cross-country

panel data on selected Latin American countries and then using micro-level labor market data from Turkey.

3 Empirical Analysis

3.1 Cross-Country Analysis: Evidence from Latin America

3.1.1 Data

I focus on 17 Latin American countries in my cross-country empirical analysis: Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, El Salvador, Ecuador, Guatemala, Mexico, Nicaragua, Panama, Paraguay, Peru, Uruguay, and Venezuela.⁷ The data period is from 1999 to 2007 (due to data availability concerns). I focus on the Latin American countries for two reasons: (1) the size of the informal sector is typically large in Latin American countries and (2) countries in the Latin American region are similar to each other in many respects, thus, concentrating on these countries minimizes (not totally eliminates) the need to control for the effects of regional, social, and institutional factors in the regressions.

The empirical analysis in this study relies on different data sources. The data on the size of the informal sector are taken directly from [Schneider, Buehn, and Montenegro \(2010\)](#), who estimate the share of the informal sector in GDP for 162 countries between 1999 and 2007. Their estimation strategy can be summarized as follows. At the first stage, they provide a list of the main causal determinants of the size of informal sector based on the empirical results presented in the previous literature. Broadly speaking, these factors are (1) tax and social security contribution burdens, (2) intensity of regulations, (3) quality and quantity of public services, (4) opportunities in the formal (or official) economy, (5) monetary indicators, and (6) labor market indicators related to informal employment. At the second stage, they construct indices based on model-based estimations of informal sector size, which use these causal factors as independent variables. After appropriate normalizations, these indices are interpreted as

⁷I exclude Haiti for data availability reasons and Honduras for the reason that there has been a profound institutional change (the duration of compulsory primary education is increased from 5 years to 6 years) in the sample period.

the size of informal sector. In particular, their estimates are readily interpretable as the share of informal economy in the GDP. Cross-country data on the size of informal sector is scarce and this is a recent but a widely used data set in the literature.

The schooling attainment data is taken from the EdStats database of the World Bank. I use three distinct measures of aggregate schooling achievement. The first one is the secondary education gross enrollment rates for all programs. This indicator measures the total enrollment in secondary education, regardless of age, expressed as a percentage of the population of official secondary education age. Note that the gross enrollment rate can exceed 100 percent due to the inclusion of over-aged and under-aged students because of early or late school entrance and grade repetition. The second indicator is the out-of-school rate for children of lower secondary school age. It refers to the number of children of official lower secondary school age who are not enrolled in lower secondary school expressed as a percentage of the population of official lower secondary school age. Finally, I use the tertiary education graduation rate. This corresponds to the total number of graduates in tertiary ISCED 5A programs (i.e., college degree) expressed as a percentage of the total population of the age where they typically finish the most common college-equivalent program in a given country. Data on the current public expenditure on education as percentage of gross national income are also taken from the same database. Data on GDP per capita in current US dollars are taken from the Economic Policy and External Debt database of the World Bank. Data on the Gini coefficient comes from the Poverty database of the World Bank. Tables (1) and (2) summarize the means of all these variables over the sample period.⁸

Data on schooling attainment indicators, public expenditures on education, and the Gini coefficient are missing for some countries for certain time periods; but, the number of empty cells is limited. In this sense, I deal with an unbalanced panel data, which can still produce unbiased estimates with small efficiency losses.

⁸Further information can be obtained from the official EdStats documentation.

3.1.2 Cross-country model

The empirical strategy is based on a standard panel data regression framework with country and time fixed effects. Countries are indexed by j and time is indexed by t . The model is simply as follows:

$$s_{jt} = \beta_0 + \beta_1 i_{jt} + \beta_2 \mathcal{G}_{jt} + \beta_2 e_{jt} + \beta_3 g_{jt} + \nu_j + \theta_t + \epsilon_{jt}, \quad (3.1)$$

where s_{jt} is the measure of aggregate schooling attainment, i_{jt} is the share of informal sector in GDP, \mathcal{G}_{jt} is the Gini coefficient for income distribution, e_{jt} is the public expenditures on education as a fraction of gross national income, g_{jt} is the natural logarithms of per capita GDP, ν_j is the country-level fixed effect that is potentially correlated with the observables, θ_t is the time fixed effect, and ϵ_{jt} is a random error term. The main purpose is to estimate the sign and magnitude of β_1 , controlling for observed and unobserved country-level variation. Equation (3.1) is estimated using three alternative definitions for s_{jt} . First, secondary education enrollment rate; second, out-of-school rate for children of lower secondary school age; and, finally, tertiary education graduation rate.

The main idea is as follows. The size of the informal sector as a percentage of GDP proxies the availability of job opportunities in the informal sector. The larger the informal sector is, the greater the job options that young-unskilled workers have. In this setting, β_1 measures the extent to which the aggregate schooling outcomes may be related to the availability of informal job opportunities—controlling for observed and unobserved variation at the country level. Next, I report and discuss the results of the fixed-effect panel data regressions that I perform for each of the aggregate schooling indicators I focus on, controlling for both country and time fixed effects as well as observed variation. I cluster standard errors at the country level and report these robust standard errors. Tables (3), (4), and (5) summarize the regression outcomes.

In the first set of regressions [Table (3)], the dependent variable is the secondary education gross enrollment rate. There is a clear negative correlation between the secondary school

enrollment rates and the size of the informal sector. This result is quite robust to the inclusion of the country-level variables for per capita GDP, Gini coefficient, and public investment on education. Among the other control variables, per capita GDP seems to be the one with highest predictive power, while the public investment in education with the lowest. The results state that a percentage point increase in the size of the informal sector (as share in GDP) may lead to approximately two percentage points decline in the secondary-school enrollment rates.

The results of the second set of regressions are provided in Table (4). The dependent variable is the out-of-school rate for children of lower secondary school age. The regression outcomes suggest that there is a stable positive association between the out-of-school rate and the size of the informal sector. The coefficient of the size of informal sector is slightly above 1 (around 1.3), meaning that a percentage point increase in the size of informal sector leads to an around 1.3 percentage point increase in the out-of-school rate. Inclusion of the control variables does not seem to improve the fit.

These two sets of regressions jointly yield the result that the availability of job opportunities in the informal sector tend to affect the school enrollment decisions of young individuals negatively. To be concrete, Latin American countries with larger informal sectors tend to have lower school enrollment rates and higher out-of-school rates. An interesting side note is that the coefficient of the school enrollment rates is larger than the out-of-school rate, which perhaps implies that the informal job opportunities are more likely to deter enrollment rather than to induce dropout.

The final set of regressions focuses on the tertiary education graduation rates as the dependent variable [Table (5)]. I find a positive correlation between the size of the informal sector and the tertiary education (i.e., college-equivalent and above) graduation rates. The fit improves with the inclusion of the other regressors. I show that a percentage point increase in the size of informal sector is associated with an approximately 2 percentage points increase in the tertiary school graduation rates. At the first instance, this result sounds to be counter-intuitive, because it suggests that college graduation tendency is higher in countries with larger

informal sectors. However, by standard selectivity arguments, the correct interpretation is along the lines of [Cameron and Heckman \(1998\)](#); that is, this result suggests that when informal sector is large, those who are more likely to succeed in college choose to go to college. Thus, college graduation rates tend to be higher in countries with sizable informal sectors.

These results are also consistent with the findings in the growth literature. Growth models in the RBC tradition predict that large informal sectors are phenomena associated with developing countries, the size of the informal sector shrinks along the development path, and, at the steady state, the informal sector vanishes [see, for example, [Ihrig and Moe \(2004\)](#)]. Moreover, schooling attainment also improves along the development path. The findings of this study rationalizes this parallel movement between the size of the informal sector and schooling achievement along the development path.

In terms of policy implications, these results suggest that any policy intervention that can potentially affect the size of the informal sector may have large second-round (and rather long-term) effects on aggregate schooling outcomes in a given country. Because, for example, cutting the size of the informal sector may lead to less employment opportunities in the informal sector and, therefore, may induce the young individuals to receive more education at the school. The downside is that it may restrict the skill accumulation options for the unskilled.

3.1.3 Potential Limitations

Although the results of the cross-country panel regressions tell a coherent story, the empirical analysis has some potential limitations. *First*, the empirical specification given in Equation (3.1) can only imperfectly represent the nature of income shocks that would affect the relationship between school achievement and informality. The natural logarithm of per capita GDP, country fixed effects, and time fixed effects can only partially capture the effect of income. In particular, there is no indicator for informal income that can provide a measure of the opportunity cost of dropping out of school.

Second, the cross-country regression setup may be subject to a standard self-selection criticism in the sense that those who drop out of school for an informal job might be the ones who cannot get a formal job. Also, those who have a college degree are the ones who are more likely to succeed in formal jobs. This is another statement of the standard “ability bias” argument. Although I mention that the results should be interpreted as suggestive correlations rather than causal effects, it should be noted once again that there might be several layers of observed and unobserved heterogeneity issues that might be driving the results of the cross-country analysis.

Third, the share of informal employment used in Equation (3.1) is perhaps an inadequate measure of job availability in the informal sector. The country-level differences in this variable might reflect other institutional differences rather than informal job availability. An analysis with micro-level data, however, would allow for region level variation in the share of informality within the same country; thus, microeconomic analysis could provide a better environment to capture variation in informal job availability. *Finally*, the variable representing public expenditures on education (as a fraction of gross national income) might also be an imperfect measure of school investment, because cross-country differences in education systems might distort the link between expenditures and investment intensity. There are also other cross-country differences in labor market institutions that cannot be captured by Equation (3.1).

Most of these potential problems can be, at least partially, remedied with micro-level data. In the next section, I use data from Turkey—another developing country with extensive informal employment—to test the main hypothesis of this paper. Turkey is out of the cross-country sample; so, such an exercise would be an appropriate robustness check. Most importantly, the Turkish data will allow us to control for factors such as regional informal wage differences, regional unemployment differences, and unobserved heterogeneity. Moreover, since I deal with only one country, the problems related to cross-country differences that cannot be captured by country fixed effects will not spoil the results.

3.2 Micro-Level Analysis: Evidence from Turkey

3.2.1 Data

To test the main hypothesis of the paper with micro-level data, I use the 2012 and 2013 waves of the Turkish Household Labor Force Survey (THLFS), which is a nationally-representative and publicly available data set. The Turkish labor force statistics are calculated from the THLFS; so, it is a large-scale data set with extensive micro-level details on labor market outcomes. Employed workers whose jobs are not registered with the Social Security Authority are marked as “informal” workers in THLFS. The fraction of informally-employed workers in the pool of employed workers is around 20 percent, which means that informality is extensive in Turkey—albeit being slightly lower than the informal employment rates in the Latin American countries.

For labor market income, I use the sum of monthly wage income and other non-wage monthly earnings such as bonuses and overtime payments. I use the Consumer Price Index for 2013 (which is 7.4 percent) to deflate earnings observations in 2013—i.e., 2013 earnings are converted into 2012 prices. This means that the earnings variable is defined as “real monthly earnings.” Consistent with the official definition of the labor force, I restrict the sample to the individuals of age 16 and above. I also restrict the sample to the individuals in the labor force (i.e., employed workers and unemployed individuals who are actively seeking a job) to focus on those who are attached to the labor market—thus, whose schooling decisions are more likely to interact with labor market outcomes. I focus on four educational outcomes: (1) no degree (ND), (2) high school dropouts (HSD), (3) high school graduates (HSG), and (4) college graduates (COL).⁹ I also include other control variables such as year fixed effects, cohort fixed effects, region fixed effects, gender, age (as a quadratic polynomial), informal employment, and urban versus rural location. Table (6) presents the summary statistics.

There is significant region-level variation both in terms of demographic characteristics and labor market outcomes in Turkey. Since there are no regional federations or states in Turkey, the

⁹Note that the HSG category also includes the ND category.

formal labor market institutions are almost homogeneous across the regions—except perhaps a few regional subsidy programs, which are rather minor.¹⁰ To control for region-level variation in earnings and unemployment rates, I construct the following region-specific variables: regional unemployment rates, regional average informal earnings, regional average formal earnings, and regional rates of informal employment. In constructing these variables, I use the most detailed (NUTS2-level) regional classification—which divides Turkey into 26 regions—that is available in the THLFS.

3.2.2 Microeconomic model

In this section, I test whether the predictions of the cross-country regression described in Equation (3.1) are holds with micro-level labor market data from Turkey. In other words, I use Turkish micro data to check if informal employment opportunities can potentially distort schooling outcomes. The micro-level analysis improves upon the limitations of the cross-country analysis in several ways. First, the cross-country analysis does not allow controlling for the relative attractiveness of informal work in an appropriate way. The best way to control for this factor is to include a measure of average informal labor earnings. Such an extension is possible using micro-level data. Second, cross-country differences in the labor market conditions may not be fully controlled for by including country fixed effects. With micro-level data, regional unemployment rates and regional rates of informal employment can do this task. Finally, cross-country institutional differences also may not be fully captured by country fixed effects. Using micro data from a single country—preferably with country-wide institutional homogeneity, as Turkey—will, at least partially, circumvent this problem.

To perform this task, I construct the following econometric model:

$$s_{i,r,t} = \alpha_0 + \alpha_1 \ln(w_{r,t}^{\text{inf}}) + \alpha_2 \ln(w_{r,t}^{\text{for}}) + \alpha_3 u_{r,t} + \alpha_4 e_{r,t}^{\text{inf}} + \boldsymbol{\theta}' \mathbf{X}_{i,r,t} + f_{r,t} + \epsilon_{i,r,t}, \quad (3.2)$$

where i , r , and t index individuals, regions, and time periods, respectively, $s_{i,r,t}$ is the schooling

¹⁰For example, in the United States, state-level labor market institutions, such as minimum wage laws, are different than federal labor market institutions. Such differences do not exist in Turkey. The only remaining regional differences in Turkey come from regional subsidies as well as differences in cultural norms. These minor differences can be more easily captured by region-specific fixed effects.

outcome, $w_{r,t}^{\text{inf}}$ is the region-specific average informal earnings, $w_{r,t}^{\text{for}}$ is the region-specific average formal earnings, $u_{r,t}$ is the region-specific unemployment rate, $e_{r,t}^{\text{inf}}$ is the region-specific informal employment rate, $\mathbf{X}_{i,r,t}$ is a vector of individual level controls, $f_{r,t} = f_r + f_t$ denote region¹¹ and year fixed effects, and $\epsilon_{i,r,t}$ is an error term.

As I describe in Section 3.2.1, I use four different schooling attainment variables: no degree (ND), high school dropout (HSD), high school graduate (HSG), and college graduate (COL). These are all constructed as separate dummy variables. ND and HSD correspond to low-achievement outcomes, while HSG and COL can be classified as high-achievement ones. The vector of control variables $\mathbf{X}_{i,r,t}$ include gender, cohort dummies (as 5-year age groups), age as a quadratic polynomial, and a dummy variable describing the location of the individual along the urban-rural divide. The coefficients α_1 and α_2 jointly capture the relative attractiveness of informal versus formal labor market earnings measured in terms of the averages calculated within local labor markets. α_3 captures the state of the local labor market and α_4 measures the prevalence (or availability) of informal job opportunities in the local labor market. Although all of the coefficients are of interest, I will focus on interpreting the sign, magnitude, and the degree of statistical significance of α_4 .

Table (7) presents the estimates. The results of the micro-level analysis confirm those of the cross-country analysis. Specifically, I find that the secondary school dropout rates (i.e., ND and HSD) decline with the prevalence of informal employment opportunities in the local labor market. In particular, 1 percentage point increase in the rate of informal employment in the region increases the probability of not receiving any school degree by 0.23 percentage points and increases high school dropout rates by around 0.17 percentage points. The story is reversed for the HSD categories. I find that 1 percentage point increase in the rate of informal employment in the local labor market reduces high school graduation probability by 0.25 percentage points. For COL (i.e., tertiary education completion) category, however, I observe that the selectivity argument still holds; that is, the college graduation rates are higher in regions with higher informal employment opportunities. In other words, those who

¹¹Since the region-level average wages and employment variables are calculated with respect to the NUTS2 regional classification, region dummy variables are included in the NUTS1 classification (a coarser classification relative to NUTS2).

are more likely to succeed in college tend to attend college in these regions. The coefficients of the region level real earnings variables also support these results. I find that an increase in region-level real informal earnings increases the secondary school dropout rates and reduces the high school (and above) graduation rates. I also document that an increase in region-level real formal earnings reduces the secondary school dropout rates and strongly increases higher education graduation rates.

To control for unobserved heterogeneity, I re-estimated Equation (3.2) using the propensity score matching (PSM) method. The PSM is performed conditioning on one percentage point intervals on predicted school attainment probabilities. The results of the PSM regression suggest that the qualitative nature of the estimates presented in Table (7) remains mostly unchanged, although the magnitudes of some of the coefficients are slightly altered. This means that the baseline results are robust with respect to the unobserved heterogeneity concerns.

4 Concluding Remarks

This paper contributes the literature by introducing the idea that availability of informal jobs may provide an alternative skill accumulation path that might induce a tendency to leave the school early for those who are less likely to succeed in school. After constructing a stylized theoretical model, I test the empirical relevance of this idea in two steps. First, I use panel data on 17 Latin American countries to estimate the sign and the magnitude of the correlation between the size (i.e., share in GDP) of the informal economy and several school attainment measures, controlling for observed and unobserved heterogeneity at the country level. Second, I use micro-level data from Turkey to perform the same task.

I show that the existence of a large informal economy providing skill acquisition opportunities for young-unskilled workers reduces the secondary education enrollment rates and increases the out-of-school rate for children of lower secondary school age in emerging economies. I also show that there is a positive correlation between the size of the informal economy and the college graduation rates, which may imply that those who choose to attend college are the

ones who are more likely to be successful in economies with larger informal sectors. The policy implication is that any government intervention that would reduce that size of the informal sector may lead to a greater flow into further education and may increase the average level of schooling within the country in the long-run. But, whether this is a welfare enhancing outcome or not is not clear *ex ante*, because a smaller informal sector restricts the options for the unskilled and especially for those who are less likely to succeed in school. The costs and returns will most likely be different for different segments of the population.

References

- AMUEDO-DORANTES, C. (2004): “Determinants and Poverty Implications of Informal Sector Work in Chile,” *Economic Development and Cultural Change*, 52, 347–368.
- APEL, R., R. PATERNOSTER, S. BUSHWAY, AND R. BRAME (2006): “A Job isn’t Just a Job: The Differential Impact of Formal versus Informal Work on Adolescent Problem Behavior,” *Crime & Delinquency*, 52, 333–369.
- AUERBACH, A. J. AND L. J. KOTLIKOFF (1987): *Dynamic Fiscal Policy*, Cambridge, UK: Cambridge University Press.
- BAKER, M. (1997): “Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings,” *Journal of Labor Economics*, 15, 338–375.
- BECKER, G. S. (1993): *A Treatise on the Family*, Cambridge, MA: Harvard University Press.
- BEN-PORATH, Y. (1967): “The Production of Human Capital and the Life Cycle of Earnings,” *Journal of Political Economy*, 75, 352–365.
- BERTOLA, G. (1999): “Microeconomic Perspectives on Aggregate Labor Markets,” in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, vol. 3, chap. 45, 2985–3028.
- BOSCH, M. AND W. F. MALONEY (2010): “Comparative Analysis of Labor Market Dynamics Using Markov Processes: An Application to Informality,” *Labour Economics*, 17, 621–631.
- BROWN, C. (1976): “A Model of Optimal Human Capital Accumulation and the Wages of Young High School Graduates,” *Journal of Political Economy*, 84, 299–316.
- CAMERON, S. V. AND J. J. HECKMAN (1998): “Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males,” *Journal of Political Economy*, 106, 262–333.
- CANO-URBINA, J. (2014): “The Role of the Informal Sector in the Early Careers of Less-Educated Workers,” Forthcoming, *Journal of Development Economics*.

- CHIAPPORI, P.-A. (1992): “Collective Labor Supply and Welfare,” *Journal of Political Economy*, 100, 437–467.
- CUNHA, F., J. J. HECKMAN, AND S. NAVARRO (2005): “Separating Uncertainty from Heterogeneity in Life Cycle Earnings,” *Oxford Economic Papers*, 57, 191–261.
- CUNNINGHAM, W. AND J. B. SALVAGNO (2011): “Youth Employment Transitions in Latin America,” World Bank: Policy Research Working Paper #5521.
- DANCER, D. AND A. RAMMOHAN (2007): “Maternal Education and Child Schooling Outcomes in Nepal,” in *Research in Labor Economics*, ed. by S. Polachek and O. Bargain, Emerald Group Publishing Limited, vol. 26, 365–396.
- DICKENS, W. T. AND K. LANG (1985): “A Test of Dual Labor Market Theory,” *American Economic Review*, 75, 792–805.
- GUNTHER, I. AND A. LAUNOV (2012): “Informal Employment in Developing Countries: Opportunity or Last Resort?” *Journal of Development Economics*, 97, 88–98.
- HECKMAN, J. J. AND J. V. HOTZ (1986): “An Investigation of the Labor Market Earnings of Panamanian Males Evaluating the Sources of Inequality,” *Journal of Human Resources*, 21, 507–542.
- HECKMAN, J. J., L. J. LOCHNER, AND C. TABER (1998): “Explaining Rising Wage Inequality: Explorations With A Dynamic General Equilibrium Model of Labor Earnings With Heterogeneous Agents,” *Review of Economic Dynamics*, 1, 1–58.
- HECKMAN, J. J. AND S. NAVARRO (2007): “Dynamic Discrete Choice and Dynamic Treatment Effects,” *Journal of Econometrics*, 136, 341–396.
- IHRIG, J. AND K. S. MOE (2004): “Lurking in the Shadows: The Informal Sector and Government Policy,” *Journal of Development Economics*, 73, 541–557.
- KIM, S. (2010): “Uncertainty in Human Capital Investment and Earnings Dynamics,” *Journal of Human Capital*, 4, 62–83.

- LILLARD, E. L. AND R. T. REVILLE (1999): “Life Cycle Human Capital Investment: New Evidence on an Old Issue,” Unpublished manuscript, *University of Michigan*.
- LUNDBERG, S. AND R. A. POLLAK (1993): “Separate Sphere Bargaining and the Marriage Market,” *Journal of Political Economy*, 101, 988–1010.
- MAGNAC, T. (1991): “Segmented or Competitive Labor Markets,” *Econometrica*, 59, 165–187.
- MEGHIR, C., R. NARITA, AND J.-M. ROBIN (2014): “Wages and Informality in Developing Countries,” Forthcoming, *American Economic Review*.
- MEGHIR, C. AND L. PISTAFERRI (2004): “Income Variance Dynamics and Heterogeneity,” *Econometrica*, 72, 1–32.
- MONK, C., J. SANDEFUR, AND F. J. TEAL (2008): “Does Doing and Apprenticeship Pay off? Evidence from Ghana,” Unpublished manuscript, University of Oxford.
- PATERNOSTER, R., S. BUSHWAY, R. BRAME, AND R. APEL (2003): “The Effect of Teenage Employment on Delinquency and Problem Behaviors,” *Social Forces*, 82, 297–335.
- RAMMOHAN, A. (2012): “The Trade-off Between Child Labour and Schooling in India,” *Education Economics*, 22, 484–510.
- SCHAFFER, M. J. (2006): “Household Change and Rural School Enrollment in Malawi and Kenya,” *Sociological Quarterly*, 47, 665–691.
- SCHNEIDER, F., A. BUEHN, AND C. E. MONTENEGRO (2010): “New Estimates for the Shadow Economies all over the World,” *International Economic Journal*, 24, 443–461.
- STIGLITZ, J. E. (1976): “The Efficiency Wage Hypothesis, Surplus Labor, and the Distribution of Labour in LDCs,” *Oxford Economic Papers*, 28, 185–207.
- TUMEN, S. (2012): “Informality as a Stepping Stone: A Search-Theoretical Assessment of Informal Sector and Government Policy,” Unpublished manuscript, Central Bank of the Republic of Turkey.

WAHBA, J. (2009): “Informality in Egypt: A Stepping Stone or a Dead End?” Economic Research Forum Working Paper #456.

ZAPATA, D., D. CONTRERAS, AND D. KRUGER (2011): “Child Labor and Schooling in Bolivia: Who’s Falling Behind? The Roles of Domestic Work, Gender, and Ethnicity,” *World Development*, 39, 588–599.

Average Schooling Attainment by Country

Country	[1]	[2]	[3]
Argentina	85.84	-	10.62
Bolivia	82.72	3.09	-
Brazil	105.00	-	14.44
Chile	87.14	5.54	16.08
Colombia	79.21	12.15	5.31
Costa Rica	73.38	-	25.21
Dominican Republic	66.84	11.22	-
Ecuador	61.25	23.93	-
El Salvador	58.63	18.00	8.61
Guatemala	45.79	34.44	1.61
Mexico	78.55	9.27	16.04
Nicaragua	61.55	27.66	3.03
Panama	69.16	17.36	22.89
Paraguay	64.82	13.92	6.75
Peru	86.47	2.45	-
Uruguay	101.03	10.68	7.03
Venezuela	68.97	16.41	10.19

Table 1: INDICATORS OF SCHOOLING ATTAINMENT (%). The first column gives the secondary education gross enrollment rates, the second gives the out-of-school rates for children of lower secondary school age, and the third describes tertiary graduation rates. The cells describe the average values of the respective indicators over the sample period (1999–2007). See Section 3.1 for a detailed description of the indicators.

Country-level Observables

Country	i	g	e	\mathcal{G}
Argentina	25.5	8.55	4.42	50.8
Bolivia	68.1	6.95	5.48	58.8
Brazil	40.5	8.29	4.14	58.4
Chile	20.3	8.69	3.64	53.9
Colombia	41.0	7.95	3.97	58.4
Costa Rica	26.6	8.41	4.66	48.9
Dominican Republic	32.3	8.02	1.95	51.3
Ecuador	36.6	7.66	1.60	55.4
El Salvador	47.4	7.85	2.74	50.6
Guatemala	52.5	7.56	2.86	55.8
Mexico	30.2	8.84	4.88	49.2
Nicaragua	45.8	6.73	2.96	44.9
Panama	65.1	8.40	4.42	56.1
Paraguay	41.1	7.15	4.41	55.4
Peru	61.8	7.83	2.75	52.9
Uruguay	51.5	8.61	2.27	46.3
Venezuela	33.4	8.50	3.48	47.6

Table 2: COUNTRY-LEVEL OBSERVABLES. The columns describe the average values of the covariates, where the averaging is performed over the sample period (1999–2007). To summarize, i is the size of the informal sector (% of GDP), g is the natural logarithm of the per capita GDP, e is the public expenditures on education (share in gross national income), and \mathcal{G} is the Gini coefficient. See Section 3.1 for a detailed description of the variables.

Dependent variable: Secondary Education Gross Enrollment Rate					
	[1]	[2]	[3]	[4]	[5]
Size of informal sector	-1.753*** (0.388)	-2.214*** (0.407)	-2.246*** (0.399)	-0.866* (0.420)	-1.167** (0.509)
Log per capita GDP		6.542*** (2.205)	8.188*** (2.262)		6.919*** (2.383)
Gini coefficient			0.509** (0.208)		0.295 (0.261)
Investment (% of GNI)				0.259 (1.144)	-0.067 (1.148)
Year dummies	Yes	Yes	Yes	Yes	Yes
Constant term	146.283*** (15.908)	114.067*** (18.849)	75.343*** (24.313)	118.521*** (22.760)	58.926* (31.976)
# of observations	144	144	144	78	78

Table 3: ESTIMATION RESULTS I. *, **, *** indicate the 10%, 5%, and 1% significance levels, respectively. Robust standard errors are reported in parentheses.

Dependent variable: Out-of-School Rate

	[1]	[2]	[3]	[4]	[5]
Size of informal sector	1.066** (0.496)	1.366** (0.573)	1.477*** (0.498)	0.929 (0.698)	1.343** (0.651)
Log per capita GDP		-4.579 (4.367)	-6.704* (3.893)		-7.304 (4.914)
Gini coefficient			-0.505* (0.256)		-0.437 (0.326)
Investment (% of GNI)				-1.208 (2.110)	-0.169 (1.730)
Year dummies	Yes	Yes	Yes	Yes	Yes
Constant term	-39.481* (22.079)	-16.445 (32.584)	34.955 (28.565)	-25.498 (33.582)	36.797 (37.606)
# of observations	94	94	94	74	74

Table 4: ESTIMATION RESULTS II. *, **, *** indicate the 10%, 5%, and 1% significance levels, respectively. Robust standard errors are reported in parentheses.

Dependent variable: Tertiary Education Graduation Rate

	[1]	[2]	[3]	[4]	[5]
Size of informal sector	0.455 (0.649)	0.506 (0.641)	0.561 (0.498)	2.799* (0.655)	1.718** (0.811)
Log per capita GDP		-4.579 (4.367)	-2.918 (2.920)		-3.767* (2.086)
Gini coefficient			0.233 (0.386)		0.374 (0.320)
Investment (% of GNI)				-3.525 (3.076)	0.287 (0.401)
Year dummies	Yes	Yes	Yes	Yes	Yes
Constant term	-4.733 (24.397)	23.617 (32.163)	3.363 (46.795)	-65.680* (36.191)	-42.092 (38.313)
# of observations	89	89	89	66	66

Table 5: ESTIMATION RESULTS III. *, **, *** indicate the 10%, 5%, and 1% significance levels, respectively. Robust standard errors are reported in parentheses.

Means of Micro-level Covariates					
	ND	HSD	HSG	COL	Total
Informal worker	0.617	0.311	0.108	0.026	0.182
Male	0.630	0.794	0.765	0.619	0.738
Age	37.275	36.381	33.737	35.957	35.614
Urban	0.738	0.795	0.866	0.909	0.844
Monthly real earnings (log)	6.460	6.657	6.948	7.532	6.971
Informal monthly real earnings (log)	6.308	6.323	6.398	6.704	6.349
Formal monthly real earnings (log)	6.704	6.808	7.015	7.554	7.109
Unemployment rate	0.071	0.086	0.113	0.095	0.093
ND	1	0	0	0	0.040
HSD	0	1	0	0	0.478
HSG	0	0	1	0	0.246
COL	0	0	0	1	0.277
# of observations	7,602	91,833	47,207	53,222	192,262

Table 6: SUMMARY STATISTICS – TURKISH MICRO DATA. This table provides the sample averages of the main variables used in micro-level analysis with Turkish data. Note that the ND category is a subset of the HSD category.

Dependent variable: School attainment				
	ND	HSD	HSG	COL
Regional rate of informality	0.232*** (0.011)	0.169*** (0.020)	-0.248*** (0.019)	0.079*** (0.018)
Regional unemployment	0.239*** (0.026)	-0.013 (0.048)	0.044 (0.044)	-0.031 (0.043)
Regional log real informal earnings	0.160*** (0.010)	0.052*** (0.019)	-0.032* (0.018)	-0.020 (0.017)
Regional log real formal earnings	-0.087*** (0.011)	-0.407*** (0.026)	0.017 (0.024)	0.390*** (0.024)
Male	-0.026 (0.001)	0.131*** (0.002)	0.047*** (0.002)	-0.177*** (0.002)
Urban	-0.012*** (0.001)	-0.124*** (0.003)	0.031*** (0.003)	0.093*** (0.003)
Age + Age ² /100	Yes	Yes	Yes	Yes
Cohort dummies	Yes	Yes	Yes	Yes
Survey year dummies	Yes	Yes	Yes	Yes
Region dummies (NUTS1)	Yes	Yes	Yes	Yes
Constant term	-0.302*** (0.077)	3.433*** (0.154)	0.309** (0.143)	-2.742*** (0.141)
R^2	0.09	0.16	0.04	0.12
# of observations	192,262	192,262	192,262	192,262

Table 7: ESTIMATION RESULTS – MICRO-LEVEL ANALYSIS. *, **, *** indicate the 10%, 5%, and 1% significance levels, respectively. Robust standard errors are reported in parentheses. “Male” is a dummy variable taking 1 if male and 0 if female. “Urban” is another dummy variable taking 1 if urban and 0 if rural.