A TALE OF TWO CITIES: SETTLEMENT DISPARITIES IN SCHOOL ACHIEVEMENT IN OUAGADOUGOU (BURKINA FASO)

ISMAEL MOURIFIÉ AND IDRISSA OUILI

University of Toronto and University of Montréal

Abstract. In Ouagadougou (capital of Burkina Faso), at least 33% of the 2 million inhabitants live in informal settlements. This subpopulation lacks social infrastructure and has extremely low education attainments. Through a two-step control function approach, we investigate disparities in educational attainment between formal and informal settlements in Ouagadougou. We focus on differences in families' behaviour towards schooling in the two settlement types, in particular the trade-off between child quantity and quality. We find evidence of a child quantity and quality tradeoff in both settlements. By adapting the long (2005) group comparisons multiple testing procedure we find that the trade-off is more acute in informal settlement. Moreover, we use the latter strategy to analyze the impact of a Ten-Year Basic Education Development Plan (PDDEB 2001-2011) implemented by the government of Burkina Faso, on educational attainment's disparities between the formal and informal settlements.

Keywords: School attainment, family size, regional disparities, PDDEB, Burkina Faso, policy evaluation, control function approach.

JEL subject classification: I25, I24, J13, J16, J18, C36.

Date: November 6, 2014. The authors are grateful to Marc Henry, Romuald Méango, and David Shapiro for their helpful discussions and comments. Data was drawn from the Ouagadougou Health and Demographic Surveillance System (OHDSS) supported by the Welcome Trust Grant WT081993MA. Correspondence address: Département de sciences économiques, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal QC H3C 3J7, Canada, idrissa.ouili@umontreal.ca.

1. Introduction

In developing countries, economic and population growth are concentrated within urban areas (United Nations 2012). This trend results in relatively better socio-economic indicators in urban areas than in the rest of the country. However, those relatively good indicators hide important inequalities in education between subpopulations within the capital. A broad review conducted by the National Research Council (2003) demonstrated that the urban poor in developing countries often face worse socio-economic conditions than the rural poor. Indeed, urban poverty is concentrated in slums, so the health risks faced by the urban poor might even exceed those in rural areas (see for instance African Population and Health Research Center (APHRC) 2002, Montgomery and Hewett, 2005). At least 33% of Ouagadougou's population of 2 million lives in informal settlements (Boyer and Delaunay 2009). These settlements lacks social infrastructures such as health offices, electricity and drinking water networks. Further, literacy and educational rates are very low in informal settlements relatively to formal ones and parents' occupation are very different in the two settlement types. These fundamental differences can lead to differences in household behaviour toward children's schooling between the two types of settlements.

In this paper, we study disparities in schooling between formal and informal settlements in Ouagadougou. We are interested in differences in families' behaviour toward education between the two type of settlements. In particular, we investigate a trade-off between child quantity and quality (Becker and Lewis 1973) in both settlements.

Despite the sizeable literature on economic models of children's educational attainment in the US and developed countries[1], much less is known on the determinants of these in developing countries, as argued in Glewwe (2002, 2008), and, in particular in African countries. In the case of Burkina Faso, the existing lack of data and a suitable methodology to deal with the endogeneity of family size is particularly relevant. Moreover, to the best of our knowledge, the few existing works on educational attainment in Burkina Faso neglected the presence of unobserved heterogeneity and variability among subpopulations and regions.

This paper contributes to filling this gap in the literature. First, we use a recent (2012) database (more than 80,000 individuals) collected by the Ouagadougou Health and Demographic Surveillance System (OHDSS). The objective of the OHDSS is to understand the problems of the urban poor and

[1]Among the most prominent are Manski and Wise (1983), Eckstein and Wolpin (2001), Keane and Wolpin (2001), and Hotz and Miller (1988, 1993).

to test innovative programs that promote the well-being of this population. Therefore, the dataset gathers information on the subpopulations in the two types of settlements in Ouagadougou.

Our second contribution is methodological. In fact, testing the trade-off between child quantity and quality within a family is already complicated by the endogeneity of family size. Moreover, comparing the effect of family size between two different subpopulations can be more complicated because of the presence of unobserved heterogeneity and variability among subpopulations, especially whenever the outcomes of interest are discrete. As pointed out by Long (2009), traditional tests of the equality of coefficients across groups confound the magnitude of the regression coefficients with residual variations.

To deal with those issues, we, first, use a two-stage control function approach which allows testing the endogeneity of family size in both settlements, and estimating all the structural parameters of interests, what we need for subsequent analysis. The empirical specification takes into account the existence of settlement-specific, unobserved heterogeneity, and we use the presence of twins as an instrument to deal with the endogeneity of family size. In addition of presenting suggestive evidence of the validity of the twin instrument in this study, we apply a more formal test of instrument validity, recently, proposed by Mourifié and Wan (2014), which does not invalidate the relevance of the twin instrument in our context. Second, to compare household behaviour toward children's schooling on educational attainment between the two types of settlements, we adapt the method of Long (2009). His method consists of estimating two separate probit models for each settlement type and performing a test based on the predicted probabilities to compare the two settlement types. This method assumes a probit model in which all the covariates are exogenous. Here, since we have a potential endogenous variable (family size), our constructed control variable allow us to have a valid probit in the second step, and then to compare the two settlement types using Long's multiple testing method.

In 2001, the government of Burkina Faso implemented a Ten-Year Basic Education Development Plan (PDDEB 2001-2011) with the objectives of improving access to and quality of primary schooling (less than seven years of schooling) and reducing regional and gender-related educational attainment disparities. So, finally, we use our latter strategy to analyze and compare the relative impact of the PDDEB on educational attainment between formal and informal settlements in Ouagadougou.

Our main findings are as following. First, we find a negative causal impact of family size on primary school attainment in both settlement types. Indeed, we find evidence of a child quantity and quality tradeoff in both settlement. However, we find that the trade-off is more acute in informal

settlement. This result seems to confirm the usual view where the trade-off is more pronounced in poorer and more credit-constrained regions, see for instance Li, Zhang and Zhu (2005). Although, to the best of my knowledge it is the first time such evidence is found using African Data. It is worth noting that whenever the presence of specific unobserved heterogeneity is neglected, a naive comparison of the regression coefficient in both groups suggest that the trade-off is more acute in formal settlement. This shows the importance of accounting for unobserved heterogeneity among subpopulations in explaining a variety of phenomena, and providing suitable method to tackle this issue.

Second, we observe that, for families with the same observable characteristics, the probability of reaching the post-primary school level is higher in formal settlements than in informal ones. These differences can be explained mainly by the existence of settlement-specific, unobserved heterogeneity such as school quality for which we do not have any information in our data and the existence of settlement peer effects. This result suggests that there is a comparative advantage to living in formal settlement for children's post-primary school attainment.

Concerning the impact of the PDDEB, our results show that the PDDEB has a positive and significant effect on the school enrolment rate in both settlement types. Unfortunately, we do not find evidence of a significant and positive impact from the PDDEB on post-primary school attainment in both settlement types. However, when comparing the post-primary school attainment between the two settlement types, the results suggest that exposure to the PDDEB reduces settlement disparities in the post-primary school attainment rate. Although, with lack of detailed information on the PDDEB, we can not clearly pointed out which aspects of the program is important to explain its effects.

The remainder of this paper is organized as follows. In the following section, we briefly describe the PDDEB. Sections 3 and 4 present our empirical specification and the data. In section 5, we develop suggestive evidence that the presence of twins in the family is an appropriate instrument to determine the causal effect of family size on children's educational attainment. Section 6 presents our results, and the last section our conclusions.

2. Brief description of the PDDEB

Burkina Faso is among countries with weakest education systems in the world. In 2000, it adopted an education policy statement : the 10-year Basic Education Development Plan (PDDEB 2001-2011). The PDDEB was intended to enable Burkina Faso to reach the Millennium Development Goal of

providing to every girl and boy a complete primary school education by 2015. The PDDEB had three main components. The first component of the program aimed to improve basic education supply and reduce inequalities of all kinds. To reach this objective, the program planned to construct school infrastructure, place an emphasis on girls' education and reduce education costs. The second component aimed to improve the quality of basic education by increasing the competence of teachers, improving the production and provision of textbooks and teaching materials, and developing and intensifying actions relating to health and nutrition of pupils. The third component of the program was essentially administrative. It aimed to improve the planning and coordination of administrative, financials and accounting procedures in order to increase the absorption capacities of the technical services of the Ministry of Education.

According to the statistical yearbooks of the Ministry of Education, the attendance rate in the first grade of primary school increased from 45% in 2001 to 86% in 2011. However, in the same period, we do not observe a desired increase in the success rate on the national exam at the end of primary school (CEP). Indeed, the success rate of the primary school certificate only rose from 62% in 2001 to 64% in 2011.

Our identification strategy allows us to evaluate the effect of the first two components of the program by comparing primary school enrolment and post-primary school attainment for cohorts exposed to the program with those of not so exposed.

3. An empirical specification

In this section, we present an empirical specification of primary school enrolment and post-primary school attainment, which is based on a two-step control function approach and uses a probit model in the second step.

3.1. **Formal vs informal: Settlement-specific, unobserved heterogeneity.**

As described in the next section, our database covers two types of settlements in Ouagadougou namely, formal and informal settlements. In order to measure and compare the impact of the family size and other characteristics such as PDDEB and gender on school enrolment and post-primary attainment in the two settlement types, we could estimate a single probit model for the two settlements combined, with interactions between dummy (indicator) variables for the settlements and the variables of interest. Significant interactions would indicate significant differences in coefficients across groups. However, by doing so, we would assume that the unobserved variables between the two settlements have the

same variability. Indeed, our outcome of interest is binary, so the model does not allow for different residual variation across settlements; see Allison (1999) for a formal justification of this argument.

A simple analysis of the data shows differences in observable characteristics between the formal and informal settlements, (see Table 1). If the types of settlement differ in observables characteristics, it is unlikely that they do not differ on unobserved dimensions as well (see Altonji et al. (2005) for a formal justification of this argument). Moreover, in Table 2, we perform a test of the standard deviations difference, which shows that, for several observable characteristics used in our estimation, the standard deviations are statistically different in the formal and informal settlements. This suggests that potentially, the standard deviations of unobservable characteristics are different. If so, even if we perform two separate models in each settlement, we could not directly compare the coefficients of interest between the two settlement types because, in a probit model, coefficients are identified only up to scale (see Allison (1999)).

To tackle this issue, several methodologies have been proposed in the literature, including those by Allison (1999), Williams (2010), and Long (2009). Each of those proposals has some advantages and disadvantages; see Williams, (2011) survey. The first two methods depend heavily on the assumption that the effects of at least one variable in both settlement types is equal. Unfortunately, we have not found such a possible variable in our case. Therefore, we propose to follow the Long (2009) method. This method consists of estimating two separate probit models for each settlement type and performing a test based on the predicted probabilities to compare the two settlement types. This method assumes a probit model in which all the covariates are exogenous. Here, we have a possible endogenous variable, so we propose to use, for each settlement a two-step control function approach, in which the second step is a probit model, and then to compare the two settlement types using Long's method.

3.2. **Identification and econometric specification.**

3.2.1. *Empirical strategy 1: school enrolment.* The model for school enrolment is specified as follows:

$$N_i = x'_{in}\beta_n^d + \sigma_n^d e_{in}^d \tag{1}$$

$$S_i = 1\{p_{si}\alpha_s^d + N_i\gamma_s^d + x'_{is}\beta_s^d + e_{is}^d \geq 0\}, \tag{2}$$

with the following distributional assumption

$$(e_s^d, e_n^d|w) \quad \backsim N(0, \sum(e_s^d e_n^d)),$$

with $w = (x_s, x_n)$ and

$$\sum(e_s^d e_n^d) = \begin{pmatrix} 1 & \sigma_{sn}^d \\ \sigma_{sn}^d & 1 \end{pmatrix}.$$

The first component of the model (1) is a function governing the endogenous variable N (number of children in the family). N_i represents the number of children in the family of child i. We assume that this function is linear, as in similar works[2], although the number of children is a discrete process. However, in our data, as seen in section 4, the average number of children is more than 5. It, therefore, is easier to justify the approximation of this discrete process by a continuous distribution than by the similar works on developed countries in which the average number of children is around 2. We consider N potentially endogenous because the number of children decisions would be based on the desired education level. The causal effect of the number of children on educational attainment could be identified with the use of an instrument. Two instruments are widely used in the literature to tackle the endogeneity of family size: mixed sibling-sex composition and twin births (see for instance Black-Devereux-Salvanes (2010); Angrist, Lavy, and Schlosser (2005); and Cceres-Delpiano (2006) for studies on developed countries and Rosenzweig and Wolpin (1980), Knodel and Wongsith (1991), Anh et al. (1998), Lee (2004), Li et al. (2007), and Ponczec and Souza (2012) for studies on developing countries). Here, we consider twin births to be the most appropriate instrument for this study; Indeed, since the average number of children is five in our database the mixed sibling-sex argument seems not really relevant and even if it was the probability to have the first five children from the same sex is very low therefore the instrument would be probably weak. Moreover, in a recent study, Smits and Monden (2011) pointed out that West Africa has one of the highest twinning rates in the world, more than 15 twins per 1,000 birth while the lowest rate of twinning is observed in South America and East Asia with less than 9 twins per 1,000 birth. The vector x_{in} represents the set of all observable determinants of N, including the instrument; e_n is the unobserved characteristics which affect decisions about the number of children. We think that this variable will mainly capture the unobserved expected level of education desired by the family. $d \in \{0, 1\}$ represents the settlement type.

The second component of the model (2) represents children's enrolment in primary school. This function is considered to be an index function governing the parents' decision to enrol their child in school or not. Therefore, $S_i \in \{0, 1\}$, which is an indicator function for each child i, takes the

[2]See for instance Black-Devereux-Salvanes (2010); Angrist, Lavy and Schlosser (2005); and Cceres-Delpiano (2006).

value 1 if the child is enrolled in school. It depends on exogenous and endogenous observable and unobservable variables. $p_{si} \in \{0,1\}$ indicates if a child i reached age 7 during the period of the program (PDDEB_S) because the official age for school enrolment was 7 years old at that time. Many other factors can determine child school enrolment, and we must control for all the observable determinants namely x_{is}. These factors have been widely discussed in the literature, and can be grouped into three groups: children's individual characteristics, family characteristics, and school characteristics. The vector x_{is} contains all observable child and family characteristics for the child i, such as the child's age and gender, the parents' literacy and labor-force participation, and many others (see Table 2). Following Hill and Duncan, (1987), Krein and Beller, (1988) and, Brooks-Gunn et al., (1993), we construct a proxy variable for family income using an index to summarize the combined effect of a certain number of socioeconomic resource factors. It is important to note that we interact the gender and the PDDEB variables in order to capture a possible gender difference gap; those crossed variables are included in the vector x_{is}. e_{si} is an unobservable variable which mainly captures the opportunity cost to the family of enrolling their children in school.

3.2.2. *Empirical strategy 2: Post-primary school level attainment.* Similarly for the post-primary school level attainment, we have:

$$N_i = x'_{in}\beta_n^d + \sigma_n^d e_{in}^d \tag{3}$$

$$Y_i = 1\{p_{yi}\alpha_y^d + N_i\gamma_y^d + x'_{iy}\beta_y^d + e_{iy}^d \geq 0\}, \tag{4}$$

with the following distributional assumption

$$(e_y^d, e_n^d|w) \quad \backsim N(0, \sum(e_y^d e_n^d)),$$

with $w = (p_y, x_y, x_n)$ and

$$\sum(e_y^d e_n^d) = \begin{pmatrix} 1 & \sigma_{yn}^d \\ \sigma_{yn}^d & 1 \end{pmatrix},$$

where N_i, e_{in}, and d have the same definition as in the first model. $Y_i \in \{0,1\}$ models the post-primary school level attainment for every child i who has already been enrolled in school. It takes the value 1 if the child reached the post-primary school level. p_{yi} is a discrete variable which indicates the number of years every child was potentially exposed to the PDDEB (PDDEB_Y). For instance, if a child was less than seven years old before 2002, we assume that all his primary schooling (6 years) was exposed to the PDDEB; otherwise, if the child was 10 years old in 2002, we assume that he was exposed to the PDDEB for only two years because 13 is the normal age to start post-primary

school. It is worth noting that this variable is a proxy of the real duration of exposure to the PDDEB because some children would have repeated grades[3]. This variable allows us to evaluate the impact of the PDDEB on primary school quality. Indeed, we think that, if the program had a positive effect on primary school quality, children who were more exposed to the PDDEB should be more likely to pass the primary national exam and then to be enrolled at the post-primary level. e_{yi} is an unobservable variable which could contain the school characteristics that we unfortunately do not observe, and the unobserved child ability and the level of parental support. In some recent works, Cunha and Heckman (2007, 2008); Cunha, Heckman, and Schennach (2006) and Heckman, Stixrud and Urzua (2006) found that both cognitive and non-cognitive abilities are important factors in schooling decisions. In the data on developed countries, it is possible to control for a child's ability using IQ tests as a proxy variable, but this is rarely the case with data from developing countries as argued in Glewwe (2002). As in the school enrolment equation, the vector x_{iy} contains all the observable child and family characteristics which can determine child school attainment (see Table 2).

3.2.3. *Identification and estimation: Control function approach.* To estimate each model, we performed three steps. For the sake of simplicity, we drop the indices d and i in the following formulations and restrict ourselves to Y. We likewise estimate the model for S (school enrolment).

First Step: *From (3) obtain the Ordinary Least Squares (OLS) estimates* $(\hat{\beta}_n, \hat{\sigma}_n)$ *of the first-stage equation and the standard residuals* $\hat{e}_n = (N - x_n'\hat{\beta}_n)/\hat{\sigma}_n$*, where* $\hat{\sigma}_n$ *is a household's clustered standard errors.*

Then, rewrite (4) as:

$$Y = 1\{p_y\alpha_y^* + x_y'\beta_y^* + N\gamma_y^* + \hat{e}_n\zeta_y^* + \tilde{e}_y \geq 0\}, \tag{5}$$

where

$$(\tilde{e}_y|w) \equiv (\frac{e_y - \sigma_{yn}e_n}{\sqrt{1-\sigma_{yn}^2}}|w) \quad \backsim N(0,1).$$

Second step: *Perform a simple probit model to consistently estimate* $(\alpha_y^*, \beta_y^*, \gamma_y^*,$ *and* $\zeta_y^*)$ *and to obtain all the standard deviations from all parameters.* After the second step of the control function approach, a third step is necessary to recover the structural parameters from equation (4).

[3]We think that this proxy will be very close to the real variable because during the PDDEB period, the government decided to forbid repeating grades within primary school.

Third step: *Applying the control function approach provides the following relation:* $\sigma_{yn} = \frac{\zeta_y^*}{\sqrt{1+\zeta_y^{*2}}}$, $\alpha_y = \frac{\alpha_y^*}{\sqrt{1+\zeta_y^{*2}}}, \beta_y = \frac{\beta_y^*}{\sqrt{1+\zeta_y^{*2}}}$ *and* $\gamma_y = \frac{\gamma_y^*}{\sqrt{1+\zeta_y^{*2}}}$. *We calculate the confidence interval of our structural parameters by bootstrapping the structural model to obtain the distribution of the parameters and then, compute the corresponding percentiles.*

Thourth step: *Applying group comparisons with predicted probabilities.* Let $\pi(z_d) \equiv \mathbb{P}(Y = 1|Z_d = z_d)$ for $d \in \{0, 1\}$ the school attainment predicted probability at a given value of z_d in the settlement d, where z_d is used to specify the realizations of the vector of covariates $(p_y, x_y, N, \hat{e}_n)$ for each settlement. Then test $H_0 : \pi^{Formal}(z_1) - \pi^{Informal}(z_0) = 0$ using the z-statistics:

$$z = \frac{\pi^{Formal}(z_1) - \pi^{Informal}(z_0)}{\sqrt{Var\Big[\pi^{Formal}(z_1) - \pi^{Informal}(z_0)\Big]}}.$$

The confidence region for the z-statistics can be computed using the STATA package proposed by Xu and Long (2009).

Remark 1. *If we use the fitted values approach, which consists of estimating an OLS to get* $\hat{N}$ *in the first stage and a probit estimation in a second using the following equation:* $Y = 1\{p_y\alpha_y + x_y'\beta_y + \hat{N}\gamma_y + e_y \geq 0\}$, *these two steps will provide only these parameters:* $\overline{\alpha}_y = \frac{\alpha_y}{\sigma}, \overline{\beta}_y = \frac{\beta_y}{\sigma}, \overline{\gamma}_y = \frac{\gamma_y}{\sigma}$, *where* $\sigma^2 = 1 + \gamma_y^2\sigma_n^2 + 2\gamma_y\sigma_n\sigma_{yn}$. *Note that from* $\overline{\alpha}_y, \overline{\beta}_y, \overline{\gamma}_y$, *and* σ_n, *we cannot back up estimates of* α_y, β_y, *and* γ_y *because we do not know* σ_{yn}. *Therefore, estimation of the parameters of interest through this method is problematic.*

4. Data

Data for this analysis was drawn from the Ouagadougou Health and Demographic Surveillance System (OHDSS). Launched in 2008, OHDSS is a research platform implemented in five neighbourhoods in the northern periphery of Ouagadougou. In Figure 1, a map of the city of Ouagadougou shows areas monitored by OHDSS, which are outlined in black. The first two from the bottom are the formal settlements, and the three above them are the informal settlements. The OHDSS areas were chosen to target the most vulnerable populations of the city. The main objectives are to understand the problems of the urban poor and to test innovative programs that promote the well-being of this population. The demographic surveillance consists of regularly updating data on vital events (births, deaths, unions) and migration events. This platform is constructed to contrast the two

types of settlements in Ouagadougou, namely formal and informal settlements. Since 2008, OHDSS has covered more than 80,000 individual, half from each type of settlement. OHDSS data contain considerable information on individual and family characteristics. For each individual while data on age, school attendance, literacy, labour-force participation, migrations, and unions are available, the family characteristics concern family's physical assets and living conditions[4].

Data on education were collected in 2012 for all individuals aged 6 or older. Information collected were about past and current school attendance (2011-2012 school year), current grade if still in school, and the highest grade attained by those no longer in school. This study focuses on two measures of educational attainment: the proportion of children who have ever been enrolled in school and those who attended the post-primary level.

Several exclusion criteria were applied to the sample. First, we restricted our sample to children aged at least 15 years old. This limitation avoided including individuals still in primary school in post-primary analysis. Second, we limited children's age to 30 to avoid having cohorts with too much difference. The exclusion criteria resulted in to a sample of 3,553 children from 15 to 30 years old[5].

4.1. **Formal vs informal settlement: Summary statistics of the overall OHDSS sample.** Table 1 presents the differences between the two settlements through descriptive statistics of the entire population monitored by the OHDSS in 2011. The inhabitants of the informal settlements seem to be younger than those in the formal settlements, with a greater share of children under 5 years old. Indeed, more than 20% of the population in informal areas is aged less than 5 years old compared to 15% in formal ones. The informal settlements are also characterized by a less educated population. More than 40% of this population has never been enrolled in school while 37% of the population in formal settlements has a secondary or higher level of education. In term of families' characteristics, we note that informal settlements lack of baseline socio-economic conditions. For instance, there is not an electricity network in the informal settlements, and almost all families use oil lamps or flashlights as sources of light. The national network for water does not exist in these settlements and nearly three quarters of the population uses a public fountains. The other quarter uses drilling or wells as a source of drinking water. In the formal settlements, more than half of the population is served by the national domestic water network. Regarding the material used for the wall of houses, we notice that, in the informal settlements, nearly all houses (99%) are

[4]For more details on OHDSS, see Rossier et al.(2012).
[5]We obtained similar results if we restricted the sample to those 15-25 years old.

FIGURE 1. Location of areas monitored by OHDSS

constructed of mud bricks while more than three quarters of the houses in the formal settlements are constructed of cement bricks. Even if in both settlements, families use mainly wood for cooking, in the formal settlements, more than 30% also use gas as source of energy for cooking. The main means of transportation in the informal settlement is the bike (used by 47% of this population), while the population in the formal settlements uses mainly motorcycles (65%) and cars (12%).

4.2. **Summary statistics of the 15-30-year-old sample.** Table 2 presents the descriptive statistics of the 15-30-year-old sample by settlement type. We notice that the school enrolment rate is higher in the formal settlements than the informal ones. Indeed, only 6% of the 15-30 year olds in the formal settlements have never enrolled in school compared to 18% in the informal ones. Even if there is a relatively strong enrolment in school, the school dropout rate seems to be high. For instance, among individuals enrolled in school, only 53% reached the post-primary level in the informal settlements. In the formal settlements, this percentage is approximately 72%. The average number of children per family (family corresponding to those of the 15-30-year-old sample) is around 5 in the two settlement types, with a slight higher average in the informal settlements (5.41 versus 5.26). We

notice that, in the formal settlements, 14% of children come from a family in which there are twins. In informal settlements, this proportion is approximately 10%. For school enrolment analysis, the exposure to the PDDEB is defined by the fact that, a child reached 7 years at any period of the PDDEB. The results in Table 2 show that, in the formal settlements, 32% of individuals reached age 7 (PDDEB-S) in the period of the program compared to 39% in the informal settlements. The average years of exposure to the PDDEB (PDDEB-Y) is 2.62 in formal settlements and 3.09 in the informal settlements.

5. Creating an Instrument from the presence of twins in the family

A direct comparison of school enrolment or post-primary school attainment between families of different sizes is unlikely to yield the causal effects of family size on school enrolment or post-primary school attainment. Even if a full set of controls were included in the analysis, the results would be unconvincing. If families of different sizes differ in so many observed dimensions, it is unlikely that they do not differ in unobserved dimensions as well. Table 3 shows the characteristics of the 15-30-year-old sample depending on whether individuals come from a family with more than 5 children (average number of children in the family). As we can see, individuals in the two types of families differ in several characteristics in both formal and informal settlements. Therefore, we can posit that families of different sizes also differ in unobservable characteristics, and we have to take into account the possible endogeneity of the number of children.

We can observe that this difference in observed dimensions of families with different sizes is more pronounced in the formal settlements than in the formal ones. Therefore, we would expect to have a stronger endogenous effect in the formal settlements than in the informal ones.

In this section, following Angrist, Lavy, and Schlosser (2005) and more recently Ponczec and Souza (2012), we develop suggestive evidence that the presence of twins in the family is an appropriate instrument to determine the causal effect of family size on children's school enrolment and post-primary school attainment. To be a "good exclusion" variable for the number of children, the presence of twins in the family must be correlated with the number of children in the family and uncorrelated with the unobservable variables which could affect school enrolment and post-primary school attainment.

5.1. **Presence of twins predicts family size.** To justify the correlation of the presence of twins and the number of children in the family, we first establish that the number of children is positively

correlated with the presence of twins in the family. The first column (1) in Table 4 shows that the average number of children is higher in families with twins: 6.06 vs 5.13 in the formal settlements and 6.21 vs 5.32 in the informal settlements. The presence of twins in the family leads to around 1 more child in the family. A t-test of the difference in the mean of the number of children shows that the difference between the two types of families is statistically significant at the level of 1%.

Further, we perform two types of regressions of the number of children on the instrument. The first is a linear regression on the instrument only, and in the second, we control by adding a set of family characteristics (ethnic group; religion; father and mother age, activities, level of education; and family standard of living). In Table 4, the second column (2) shows the results without the family characteristics control and the third column (3) the results whenever controlling for family characteristics. The results of the two equations show a positive and significant effect of the presence of twins on the number of children in the family in both settlement types. The results are almost the same in both cases. The F-statistic from a test of the null hypothesis that the coefficient of the instrument in regression (1) is 0 is 111.55 in the formal settlements and 33.66 in the informal ones, which suggests that the instrument is not weak in either settlement by conventional standards. Even if controlled by family characteristics, results still suggest that the instrument is not weak.

5.2. **School enrolment, post-primary school attainment, and the presence of twins in the family.** If individuals from a family with twins differed systematically from individuals in which there are no twins, then the presence of twins would be a poor instrument because it would be correlated with unobserved determinants of school enrolment or unobserved determinants of post-primary level attainment. We now present suggestive evidence that there is no such correlation.

First, we examine the possibility of differences between the two groups by regressing the instrument on individuals and family characteristics. Equation (1) in Table 5 shows the results of this regression for formal and informal settlements. We observe that all the characteristics are not significant in both settlements. Indeed, we do not reject the null hypothesis that the coefficients on all of these variables are 0. The corresponding Wald statistic is 13.59 with a p-value of 0.33 in the formal settlements, and 12.60 with a p-value of 0.40 in the informal ones.

We further examine the possibility of a direct and significant effect of the instrument on school enrolment and post-primary level attainment. Equations (2) and (3) in Table 5 show the results for a probit regression of school enrolment and post-primary level attainment respectively, controlled

by individual and family characteristics. The presence of twins in the family does not produce a significant coefficient in either formal or informal settlements.

One would think that having twins or not is not random if the family could increase the probability of twins, especially through artificial insemination. However, these practices are not common in the context of Burkina Faso because they are highly expensive and not culturally accepted.

5.3. **Mourifié and Wan (2014) LATE's Test.** Based on some insight in Heckman and Vytlacil (2005), Kitagawa (2014) and Mourifié and Wan (2014, MW) proposed, two formal testing procedure to assess the validity of an instrument. Those two testing procedure have the advantage to be the most powerful tests able to screen the violation of the instrumental variable assumption whenever the treatment response is heterogeneous. Indeed, in addition to be a "good exclusion" variable if the potential number of children is a monotone function of the instrument, Imbens and Angrist (1994) shows that the IV estimand can consistently estimate the average treatment effect (ATE) for the subpopulation of compliers, namely, the local average treatment effect (LATE). MW proposes an easy implementable test to verify the validity of those two conditions. Please see MW for more details. Although, MW's test can also be implemented for a discrete non-binary treatment, we dichotomize the number of children in order to avoid to have few observations in every subgroup. The model can therefore be written as follows: Let $Y_i = Y_{i1}D_i + Y_{i0}(1 - D_i)$, where $D_i \in \{0, 1\}$ is the observed treatment indicator of the individual i, D_i takes the value 1 if the child is in a family with more than k children. (Y_{i1}, Y_{i0}) are potential outcomes, in other terms Y_{id} is the school attainment level of a child i if he had been externally assigned to a family where $D_i = d$, $d \in \{0, 1\}$. Let $Z \in \{0, 1\}$ denote the twin instrument. It takes the value 1 if the presence of twin between the k first children. For each $z \in \{0, 1\}$, let D_z be the potential treatment had the Z been exogenously set to z. With this notation, we can also write the observed treatment $D = D_1Z + D_0(1 - Z)$.

The two well-known identification assumptions for LATE as introduced by Imbens and Angrist (1994) are restated as the following:

(1) (Independence) $Z \perp (Y_1, Y_0, D_0, D_1)$ and $\mathbb{P}(D = 1|Z = 0) \neq \mathbb{P}(D = 1|Z = 1)$.
(2) (Monotonicity) $D_0 \leq D_1$ almost surely.

We perform MW's test to assess the validity of those assumptions for $k \in \{2, 3, 4, 5\}$. We perform the test with and without controlling on the covariates. As can be seen in Table 6 the tests do not reject the two latter assumptions in both cases.

6. Results

6.1. **Family size and children's schooling.** The first variable of interest is family size. This section presents the impact of the number of children on school enrolment and post-primary school attainment for the 15-30-year-old sample in formal and informal settlements. The 95% confidence interval is computed by doing 1500 bootstrap replications. Tables 7 and 8, respectively, show the results for school enrolment and post-primary level attainment for both formal and informal settlements.

6.1.1. *Formal settlements.* In the formal settlements, the results suggest that the number of children is endogenous with respect to children's school enrolment. Indeed, the covariance between e_n and e_s is 0.46, within the 95% confidence interval [0.19 0.87] (see Table 7). We note a negative causal impact of the number of children on the school enrolment for the compliers. Indeed, the point estimate of -0.37 with a confidence interval of [-0.64 -0.20] reflects the negative effect of family size on children's primary school enrolment for families who have had more children than they otherwise would have because of twinning.

Similarly, the results suggest that, the number of children is endogenous with children's post-primary level attainment. The positive sign of the correlation indicates a positive correlation between parental support and the level of education desired by parents in this settlements. We also note a negative causal impact of the number of children on the post-primary school attainment. Please see Table 8.

This endogeneity of the number of children in both primary school enrolment and post-primary school attainment suggests that in the formal settlement, parents consider children's education when deciding on the number of children. This behaviour can be explained by the relatively high-level of parents who are educated. Indeed, 36% of mothers have at least a primary-school level education, increasing to 46% for their husbands (see Table 2).

Basically, we observe the existence of a tradeoff between the quantity and quality of children (Becker and Lewis 1973) in the formal settlements.

6.1.2. *Informal settlements.* Unlike in the formal settlements, the results indicate that there is no endogeneity in the number of children on both primary school enrolment and post-primary level attainment in the informal settlements. (see Tables 7 and 8). This result would indicate that the children's education is not a significant determinant of the family size in the informal settlements.

This behaviour can be explained by the relatively low proportion of parents who are educated in the informal settlements. Indeed, only 19% of mothers and 27% of their husbands have at least a primary school level education (see Table 2).

It is important to note that the effect of the number of children on school enrolment for the compliers is negative (-0.06) but not significant ([-0.46 0.15]), while it is negative (-0.25) and significant ([-0.57 -0.04]) for post-primary school attainment. (see Tables 7 and 8). These results can be explained by the fact that enrolment in public primary schools is almost free and that the opportunity cost to enrol children in primary school is not high because they are too young to be used in the labor force. However, enrolling them in post-primary school is more costly (direct cost and opportunity cost). First, the fees are higher than for primary school, and the transport costs increase because there are significantly fewer post primary schools than primary schools. Second, regarding opportunity cost, children are physically able (from the point of view of their parents) to help their parents in their employment activity.

6.1.3. *Settlement comparison.* In this section, we compare the impact of the family size on educational attainment in the two settlements. Notice that a naive comparison of the regression coefficients of the family size variable suggest that the trade-off quantity quality is more pronounced in the formal settlement. However, as explained previously, we cannot compare this impact by merely looking at the point estimate reported in Tables 7 and 8 due to the settlement-specific unobserved heterogeneity. Therefore, in this section, we compare predicted probabilities to reach the post-primary educational attainment by settlements. Predicted probabilities are calculated employing the child and household characteristics used in the post-primary school attainment analysis. Adapting Long's (2009) method, we make settlement comparisons for some fixed characteristics. We have multiple covariates, so we could estimate and compare multiple counterfactual probabilities. To be parsimonious, we focus on those that are more relevant to policy makers. We define two groups of individuals. First, children with a favourable background are defined as those with characteristics which have a significant, positive effect on post-primary school attainment. In this category, we include non-Muslim individuals with the highest standard of living index, whose parents have at least a primary-school education and work in an activity in which children can not participate. Children with unfavourable background are defined as individuals with characteristics which have a negative effect on post-primary school attainment. In this category, we include Muslim individuals with the lowest standard of living index, whose parents gave never been enrolled in school and work in an activity in which a child could participate. We hold continuous variables at their mean.

Figure 2 shows the predicted probabilities to reach the post-primary school level by gender and by settlement type. Above are individuals with an unfavourable background, and below are individuals with a favorable background. As expected, we observe that the probability of reaching the post-primary school level decreases with the number of children. The adverse effect of family size is more severe for children with an unfavourable background. We observe that, for families with the same observable characteristics, the probability of reaching post-primary school is higher in the formal settlements than in the informal settlements. Figures 3 and 4 provide more detailed information on this trend. In Figures 3 and 4, we calculate the difference of the probability of reaching the post-primary school level between the two settlements, for every standard of living index[6]. Therefore, Figures 3 and 4 present information on how families standard of living and number of children affect the difference in post-primary educational attainments between the two settlement types. The solid line indicates that the difference between formal and informal settlements is significant at the 95% level (Figure 3) or 90% level (Figure 4), while the dashes indicate that the difference is not significant. The main result is that this difference is consistently positive and significant at the 90% confidence level for families with high and middle standards of living (see Figure 4). This suggest that the trade-off quantity quality is more pronounced in the informal settlement. In terms of magnitude, we do not notice any significant gender differences in both favourable and unfavourable backgrounds. Overall, the settlement differences are low when the number of children in the family is low or high. These differences become larger around the average number of children per family, at which point the difference ranges from 0.15 to more than 0.30, depending on the family's standard of living. When the number of children in the family is low (less than 3), settlement differences are similar whatever the family's standard of living. When the number of children is high (more than 3), we observe that the higher the standard of living, the greater the difference between formal and informal settlements. At the 90% confidence level (see Figure 4), these differences become significant, especially for children from families with middle and high standards of living and an unfavourable background. One can ask why there are significant differences between formal and informal settlements even the family and children have the same observable characteristics. These differences can be explained mainly by a potential specific-settlement, unobserved heterogeneity. Indeed, the school quality may be extremely different in the two settlement types. The differences might also be explained by settlement peer effect. Indeed, informal settlements are composed mostly of households with low standard of living (see Tables 1 and 2), so there could be a spillover effect of educational behaviour by poor families on

[6]The standard of living is not included in the definition of favourable and unfavourable backgrounds.

the small number of wealthy families. One goal of the PDDEB was to reduce regional disparities; in the following sections, we assess whether the PDDEB has helped to reduce such disparities between the two settlement types.

6.2. **Further analysis: Impact of the PDDEB program.** The main objectives of the PDDEB were to improve the primary schooling rate, particularly for girls, and to improve primary school quality. Our empirical specification yields the following results for our sample.

6.2.1. *Formal settlements.* Table 7 column (1) shows that the PDDEB program has a positive and significant effect on the school enrolment rate. When we interact the PDDEB-S and gender (girl), the results in column (2) have the same point estimate, but it is no longer significant, which suggests that the impact of the program is the same for girls and boys. Additionally, Table 8 shows no significant impact of the exposure from PDDEB on post-primary school attainment, but even though it is not significant, the negative sign of the impact is surprising. It is worth noting that one policy of the PDDEB program was to forbid repeating all odd grades in primary school and to keep the rate of repetition in even grades below 10% (except for the last grade, called CM2). The main challenge occurs in the CM2 grade when students needs to pass a national exam to be able to access post-primary school. This policy would have spillover effects, by decreasing the level of effort exerted by children and then their academic performance, which would eventually reduce their chance of passing the national exam. This spillover effect would be severe for those who had been exposed to the PDDEB for a longer time. Indeed, over these 10 years, according to the statistical yearbooks of the Ministry of Education, the success rate on the national examination of primary level was 62.5% at the beginning of the program in 2001 and 64.0% at its end in 2011, despite the great improvement in enrolment in the first grade of primary school (from 45% to 86%, nationally). This policy should explain this negative sign. Policymakers should be aware that this policy can severely affect the long-term educational quality of the children in Burkina-Faso, even if it significantly increases the primary school enrolment rate.

6.2.2. *Informal settlements.* Table 7 column (1) shows that the PDDEB has a positive but insignificant effect on school enrolment. However, when we interact PDDEB-S and gender (girl), the results suggest that the PDDEB had a strong and significant positive effect on girl's school enrolment compared to boys. The relative gender differential in the PDDEB's effect on school enrolment rate can be explained by the fact that, before the plan, the gender differential in school enrolment was high in the informal settlements. Indeed, in this type of settlements, among those who are not exposed to

the program, 26% of girls were not enrolled in school compared to 18% for boys. Table 8 shows that, as in the formal settlements, exposure to the program had no significant impact on the post-primary school attainment, but the sign is also negative. The same explanation as in the formal settlements also holds for the informal settlements.

6.2.3. *Settlement comparison.* Although our results suggest that the PDDEB did not have a significant impact on the post-primary school attainment in either settlement type, we want to see if the PDDEB reduced disparities in post-primary school attainment between the two settlement types. Figure 5 presents information on how exposure to the PDDEB and the number of children in the family affect settlement disparities in the post-primary school attainment rate. We define three degrees of exposure to the PDDEB: individuals not exposed to the PDDEB, those exposed for 1 to 3 years, and those exposed for more than 3 years (4-6 years). The definitions of children's background are the same as used previously. Figure 5 demonstrates three main findings. First, the differences between formal and informal settlements are always positive even if they are not significant, which suggests that there is a comparative advantage to living in the formal settlements for post-primary school attainment. Second, the results suggest that exposure to the PDDEB reduces settlement disparities in the post-primary school attainment rate. This result suggests that the spillover effect of the no-repeating policy on the children's level of effort is more pronounced in the formal settlements than the informal settlements. Finally, this reduction of disparities is more pronounced for girls than boys. Indeed, Table 8 shows that girls are more likely to reach the post-primary school level than boys. Therefore, as the PDDEB significantly increases the enrolment of girls in the informal settlements, it decreases the disparities between the two settlement types.

7. Conclusion

In this paper, we investigate disparities in educational attainment between formal and informal settlements in Ouagadougou. We focus on differences in families behaviour towards schooling in the two settlement types, in particular the trade-off between child quantity and quality. Moreover we analyze the impact of a Ten-Year governmental educational program PDDEB on disparities in educational attainment between both settlements. Our first main result suggest evidence of a quantity-quality tradeoff in both settlements with more pronounced effect within the informal ones. Additional results suggest the existence of settlement-specific unobserved heterogeneity such as school quality and settlement peer effects even after the PDDEB which creates a comparative advantage from living in formal settlements. Therefore, to reduce such disparities, we recommend

prioritizing actions to improve school quality in the urban informal settlements. Also, these results show that efforts to keep children in school should not be directed only towards rural areas. In addition to reducing schooling costs in some urban areas, particularly informal areas, special emphasis should be put on raising awareness about the benefits of education for childrens' and families' welfare. Finally, in the methodological point of view, our results show the importance of accounting for unobserved heterogeneity among subpopulations in explaining a variety of phenomena, implementing policy, and analyzing the impact of governmental policy.

References

Allison, P. D. (1999): "Comparing Logit and Probit Coefficients Across Groups," *Sociological Methods and Research*, 28(2), 186–208.

Andrabi, T., J. Das, and A. I. Khwaja (2012): "What Did You Do All Day? Maternal Education and Child Outcomes," *The Journal of Human Ressources*, 47(4), 873–912.

Angrist, J., V. Lavy, and A. Schlosser (2005): "New Evidence on the Causal Link between Quantity and Quality of Children," NBER Working Paper Series.

Anh, T. S., J. Knodel, D. Lam, and J. Friedman (1998): "Family Size and Children's Education in Vietnam," *Demography*, 35(1), 57–70.

armistha Pal (2004): "Child Schooling in Peru: Evidence from a Sequential Analysis of School Progression," *Journal of Population Economics*, 17(4), 657–680.

Becker, G. S., and H. G. Lewis (1973): "On the Interaction Between the Quantity and Quality of Children," *Journal of Political Economy*, 81(2), S279–S288.

Beegle, D. N. K. (2006): "The Effect of School Type on Academic Achievement Evidence from Indonesia," *The Journal of Human Ressources*, XLI(3), 529–557.

Behrman, J., A. D. Foster, M. R. Rosenzweig, and P. Vashishtha (1999): "Women's Schooling, Home Teaching, and Economic Growth," *Journal of Political Economy*, 107(4), 682–714.

Behrman, J. R., and M. R. Rosenzweig (2002): "Does Increasing Women's Schooling Raise the Schooling of the Next Generation?," *American Economic Review*, 92(1), 323–334.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2010): "Small Family, Smart Family? Family Size and the IQ Scores of Young Men," *The Journal of Human Ressources*, 45(1), 33–58.

Brooks-Gunn, J., G. J. Duncan, P. K. Klebanov, and N. Sealand (1993): "Do Neighborhoods Influence Child and Adolescent Development?," *American Journal of Sociology*, 99(2), 353–395.

Burton-Jones, A., and J. spender (eds.) (2011): *The Oxford Handbook of Human Capital*. Oxford University Press, United States.

Cáceres-Delpiano, J. (2006): "The Impacts of Family Size on Investment in Child Quality," *Journal of Human Resources*, XLI(4), 738–754.

Castillo, M., P. J. Ferraro, J. L. Jordan, and R. Petrie (2011): "The Today and Tomorrow of Kids: Time Preferences and Educational Outcomes of Children," *Journal of Public Economics*, 95(11), 1377–1385.

Chappel, T. J. H. (1974): "The Yoruba Cult of Twins in Historical Prespective," *Journal of the International African Institute*, 44(3), 250–265.

Cunha, F., and J. J. Heckman (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43(4), 738–782.

Dusko, S., G. Massey, and R. Hodson (1994): "Who Were the Yugoslavs? Failed Sources of a Common Identity in the Former Yugoslavia," *American Sociological Review*, 59(1), 83–97.

Eckstein, Z., and K. I. Wolpin (1999): "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities," *Econometrica*, 67(6), 1295–1339.

Ermisch, J., and M. Francesconi (2012): "The Effect of Parental Employment on Child Schooling," *Journal of Applied Econometrics*, 28, 796–822.

Fryer, R. G., and J. S. D. Levitt (2006): "The Black White Test Score Gap through Third Grade," *American Law and Economics Review*, 8(2), 249–281.

Glewwe, P. (2002): "School and skills in developing countries:Education policies and socioeconomic outcomes," *Journal of Economics Literature*, XL, 436–482.

Glewwe, P. (2008): "education in developing countries," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave Macmillan, Basingstoke.

Gouxa, D., and E. Maurin (2005): "The Effect of Overcrowded Housing on Children's Performance at School," *Journal of Public Economics*, 89(5-6), 797–819.

Haveman, R., and B. Wolfe (1995): "The Determinants of Children's Attainments: Review of Methode and Findings," *Journal of Economics Literature*, XXXIII(4), 1829–1878.

Heckman, J. J., and F. Cunha (2007): "The Technology of Skill Formation," *American Economic Review*, 97(2), 31–47.

Heckman, J. J., and P. A. LaFontaine (2007): "The American High School Graduation Rate : Trends and Levels," NBER Working Paper Series.

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24(3), 411–482.

Hill, M. S., and G. J. Duncan (1987): "Parental Family Income and the Socioeconomic Attainment of Children," *Social Science Research*, 16, 39–73.

I.N.S.D (2012): "Demographic and Health survey and Multiple Indicator Cluster 2010," Discussion paper, National Institute of Statistics and Demography (Burkina Faso).

Janeen, B. (1994): "Is Husband's Class Enough? Class Location and Class Identity in the United States, Sweden, Norway, and Australia," *American Sociological Review*, 59(2), 220–235.

Jungmin, L. (2004): "Sibling Size and Investment in Children's Education : An Asian Instrument," IZA Discussion Paper Series, No. 1323.

Knodel, J., and M. Wongsith (1991): "Family Size and Children's Education in Thailand: Evidence from a National Sample," *Demography*, 28(1), 119–131.

Krein, S. F., and A. H. Beller (1988): "Educational Attainment of Children From Single-Parent Families: Differences by Exposure, Gender, and Race," *Demography*, 25(2), 221–234.

Lee, V. E., and D. T. Burkam (2003): "Dropping Out of High School: The Role of School Organization and Structure," *American Educational Research Journal*, 40, 353–393.

Li, H., J. Zhang, and Y. Zhu (2007): "The quantity-Quality tradeoff of choildren in developing country: Identification using chinese Twins," IZA Discussion Paper Series.

Long, J. S. (2009): "Group Comparisons in Logit and Probit Using Predicted Probabilities," Indiana University.

Marc, R., and W. Kenneth (1980): "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, 48(1), 227–240.

Mincer, J. (1993): *Studies in Human Capital : Collected Essays of Jacob Mincer*, vol. 1. Elgar, Edward Publishing, Inc.

Nye, B., L. V. Hedges, and S. Konstantopoulos (2000): "The Effects of Small Classes on Academic Achievement: The Results of the Tennessee Class Size Experiment," *American Educational Research Journal*, 37(1), 123–151.

Ponczek, V., and A. P. Souza (2012): "New Evidence of the Causal Effect of Family Size on Child Quality in a Developing Country," *The Journal of Human Ressources*, 47(1), 64–106.

Rivkin, S., E. Hanushek, and J. Kain (2005): "Teacher, Schools, and Academic Achievement," *Econometrica*, 73(2), 417–458.

Rossier, C., A. Soura, B. Baya, G. Compaoré, B. Dabire, S. D. Santos, G. Duthé, B. Gnoumou, J. F. Kobiane, S. Kouanda, B. Lankoande, T. Legrand, C. Mbacke, R. Millogo, N. Mondain, M. Montgomery, A. Nikiema, I. Ouili, G. Pison, S. Randall, G. Sangli, B. Schoumaker, and Y. Zourkaleini (2012): "The Ouagadougou Health and Demographic Surveillance System," *International Journal of Epidemiology*, 41(1-9), 658–666.

Rumberger, R. W. (1983): "Dropping Out of High School: The Influence of Race, Sex, and Family Background," *American Educational Research Journal*, 20(2), 199–220.

Rumberger, R. W. (1995): "Dropping Out of Middle School: A Multilevel Analysis of Students and Schools," *American Educational Research Journal*, 32(3), 583–625.

Smits, J., and C. Monden (2011): "Twinning Across the Developing World," *PLoS ONE*, 6(9).

UN (2012): "World Urbanization Prospects: The 2011 Revision," Discussion paper, United Nations.

U.N.D.P (2013): "Human Development Report 2013," Discussion paper, United Nations Development Programme.

Unesco (2012): "EFA Global Monitoring Report," Discussion paper, United Nations Educational, Scientific and Cultural Organization.

Williams, R. (2011): "Fitting Heterogeneous Choice Models With OGLM," *The Stata Journal*, 10(4), 540–567.

Wolfle, L. M. (1985): "Postsecondary Educational Attainment Among Whites and Blacks," *American Educational Research Journal*, 22(4), 501–525.

TABLE 1. Settlement characteristics

	Formal settlements		Informal settlements		All	
	Mean	SD	Mean	SD	Mean	SD
Variables						
Individual characteristics						
Age	24.79	17.07	21.82	16.2	23.26	16.69
Child under 5 years old	0.15	0.35	0.23	0.42	0.19	0.39
Education level (for people age 7 and older)						
Never enrolled in school	0.27	0.44	0.44	0.5	0.35	0.48
Primary level	0.36	0.48	0.4	0.49	0.38	0.48
Secondary level or higher	0.37	0.48	0.16	0.36	0.27	0.44
Number of individuals	40,584		42,933		83,517	
Family characteristics						
Age of the head of the household	42.33	13.36	36.97	11.72	38.99	12.63
Main means of transportation						
Bike	0.19	0.39	0.47	0.5	0.36	0.48
Motorcycle	0.65	0.48	0.42	0.49	0.51	0.5
Car	0.12	0.32	0.01	0.12	0.05	0.22
Main source of energy for cooking						
Wood	0.57	0.5	0.82	0.46	0.65	0.48
Coal	0.07	0.26	0.08	0.28	0.08	0.27
Gas	0.31	0.46	0.09	0.29	0.18	0.38
Main source of light						
Lamp	0.35	0.48	0.92	0.27	0.7	0.46
Electricity	0.65	0.48	0.07	0.25	0.29	0.45
Domestic water source						
Drilling\well	0.07	0.25	0.24	0.42	0.17	0.38
Collective public fountain	0.38	0.49	0.73	0.44	0.6	0.49
Individual subscription to the national network	0.55	0.5	0.03	0.18	0.23	0.42
Material used for walls of house						
Bricks made with mude	0.24	0.43	0.99	0.11	0.7	0.46
Bricks made with cement	0.76	0.43	0.01	0.11	0.3	0.6
Number of families	6,661		10,823		17,484	

TABLE 2. Descriptive statistics (15-30 year old group)

Variables	Formal settlements Mean	Formal settlements SD	Informal settlements Mean	Informal settlements SD	SD difference
Child's age	21.15	4.13	20.30	3.88	**
Child is a girl	0.51	0.50	0.51	0.50	NS
Child is Muslim	0.61	0.49	0.56	0.50	NS
Child belongings to the Mossi ethnic group	0.90	0.30	0.92	0.27	***
Enrolled in school at least once	0.94	0.24	0.82	0.38	***
Child reached age 7 during PDDEB	0.32	0.46	0.39	0.49	**
Years of Exposure to PDDEB	2.62	2.46	3.09	2.45	NS
Child reached post-primary level	0.72	0.45	0.53	0.50	***
Number of children in the family	5.26	1.69	5.41	1.60	**
Presence of twins in the family	0.14	0.35	0.10	0.30	***
Child's mother's age	45.21	4.01	43.75	3.85	*
Mother's husband age	54.59	7.58	52.52	8.46	***
Child's mother's has an economic activity in which a child could participate	0.90	0.30	0.95	0.21	***
Mother's husband has an economic activity in which a child could participate	0.70	0.46	0.74	0.44	NS
Child's mother's has at least a primary-school level	0.36	0.48	0.19	0.40	***
Husband has at least a primary-school level	0.46	0.50	0.27	0.44	***
Standard of living					
Low	0.08	0.27	0.47	0.50	***
Middle	0.19	0.39	0.43	0.49	***
High	0.72	0.45	0.09	0.29	***
Observations	3,024		1,207		

***p<0.01 **p<0.05 *p<0.1 NS not significant

TABLE 3. Individual characteristics by number of children in the family

	Formal settlements			Informal settlements		
	5 & -	6 & +	Difference	5 & -	6 & +	Difference
Variables						
Child's age	20.74	21.73	***	19.74	20.74	***
Child is a girl	0.50	0.53	NS	0.50	0.52	NS
Child belongings to the Mossi ethnic group	0.89	0.91	NS	0.92	0.93	NS
Child is Muslim	0.58	0.67	***	0.59	0.53	**
Enrolled in school at least once	0.96	0.92	***	0.85	0.80	***
Reached post-primary school level	0.78	0.64	***	0.60	0.45	***
Child's mother's age	44.48	46.24	***	0.43	0.45	***
Mother's husband's age	52.99	56.82	***	0.51	0.54	***
Child's mother has an economic activity in which a child could participate	0.87	0.95	***	0.95	0.96	NS
Mother's husband has an economic activity in which a child could participate	0.63	0.80	***	0.70	0.78	***
Child's mother has at least a primary-school level	0.47	0.22	***	0.24	0.14	***
Mother's husband has at least a primary-school level	0.54	0.34	***	0.29	0.24	*
Standard of living						
Low	0.08	0.08	NS	0.46	0.49	NS
Middle	0.16	0.24	***	0.42	0.43	NS
High	0.76	0.66	***	0.11	0.07	***
Observations	1,763	1,261		645	562	

***p<0.01 **p<0.05 *p<0.1 NS not significant

TABLE 4. Presence of twins and the number of children in the family

	(1)		(2)		(3)	
	Formal	Informal	Formal	Informal	Formal	Informal
Twin in the family	6.06	6.21	0.93***	0.89***	0.78***	0.92***
No twin in the family	5.13	5.32	-	-	-	-
T-test p value	0.000	0.000	-	-	-	-
F-stat	-	-	111.55	33.66	16	74.17
Family characteristics	-	-	No	No	Yes	Yes
Observations	3,024	1,207	3,024	1,207	3,024	1,207

***p<0.01 **p<0.05 *p<0.1 NS not significant
(1) Average number of children in the family
(2) Linear regression of the number of children
(3) Linear regression of the number of children
Family characteristics : ethnic group; religion; father's and mother's age economic activities, level of education; and family standard of living

TABLE 5. Presence of twins, school enrolment, and post-primary school attainment

	Formal settlement			Informal settlement		
Variables	(1)	(2)	(3)	(1)	(2)	(3)
Twins in the family	-	-0.073	0.106	-	0.002	-0.004
		(0.141)	(0.102)		(0.219)	(0.178)
Child's age	-0.006	-	-	0.007	-	-
	(0.007)			(0.015)		
Child is a girl	0.012	-	-	-0.23	-	-
	(0.061)			(0.121)		
Household belongings to the Mossi ethnic group	-0.255	-	-	0.200	-	-
	(0.194)			(0.397)		
Main religion is Islam	-0.163	-	-	0.166	-	-
	(0.125)			(0.207)		
Child's mother's age	0.004	-	-	-0.003	-	-
	(0.016)			(0.028)		
Mother's husband's age	-0.001	-	-	-0.008	-	-
	(0.008)	-	-	(0.014)	-	-
Child's mother has an economic activity in which a child could participate	0.211	-	-	-0.296	-	-
		-	-		-	-
	(0.232)			(0.340)		
		-	-		-	-
Mother's husband has an economic activity in which a child could participate	0.333	-	-	0.075	-	-
		-	-		-	-
	(0.136)			(0.227)		
		-	-		-	-
Child's mother has at least a primary-school level	-0.127	-	-	-0.004	-	-
		-	-		-	-
	(0.143)			(0.227)		
		-	-		-	-
Mother's husband has at least a primary-school level	-0.077	-	-	-0.012	-	-
		-	-		-	-
	(0.136)			(0.229)		
		-	-		-	-
Standard of living		-	-		-	-
Middle	-0.194	-	-	-0.258	-	-
	(0.218)			(0.182)		
High	-0.158	-	-	-0.476	-	-
	(0.193)			(0.414)		
Cons	-0.966	-	-	-0.708	-	-
	(0.718)			(1.430)		
Wald chi2(12)	13.59	-	-	12.60	-	-
Prob > chi2(12)	0.33	-	-	0.40	-	-
N	3,024	3,024	2,834	1,207	1,207	992

(1) Probit regression of the presence of twins in the family
(2) Probit regression of school enrolment
(3) Probit regression of post-primary school level of attainment
Robust standard error in bracket

TABLE 6. Testing the validity of the instrument

	School enrolment						Post primary attainment					
	I_n>4		I_n>5		I_n>6		I_n>4		I_n>5		I_n>6	
	[1]	[2]	[1]	[2]	[1]	[2]	[1]	[2]	[1]	[2]	[1]	[2]
Testing results	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Control variables	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES

TABLE 7. Family size, PDDEB, and school enrolment

Variables	Formal settlements				Informal settlements			
	(1)		(2)		(1)		(2)	
	Coeff	95% C I	Coeff	95% C I	Coeff	95% C I	Coeff	95% C I
σ_{sn}	0.46***	[0.18 0.88]	0.46***	[0.21 0.86]	0.11	[-0.34 0.57]	0.12	[-0.35 0.58]
Number of children	-0.37***	[-0.64 -0.20]	-0.37***	[-0.63 -0.21]	-0.14	[-0.46 0.16]	-0.14	[-0.49 0.17]
PDDEB_S	0.29**	[0.02 0.52]	0.13	[-0.19 0.39]	0.12	[-0.14 0.41]	-0.12	[-0.49 0.26]
Child is girl	-0.20***	[-0.35 -0.02]	-0.26***	[-0.44 -0.06]	-0.16*	[-0.34 0.02]	-0.31***	[-0.53 -0.10]
PDDEB_S*Girl			0.29	[-0.10 0.66]			0.49**	[0.10 0.89]
N	3,024				1,207			

***p<0.01 **p<0.05 *p<0.1
All regressions are controlled for family and child characteristics.
Child's age, religion, ethnic group
Child's parent's age, education level, and economic activity
Family standard of living level

TABLE 8. Family size, PDDEB, and post-primary school level attainment

Variables	Formal settlements				Informal settlements			
	(1)		(2)		(1)		(2)	
	Coeff	95% C I	Coeff	95% C I	Coeff	95% C I	Coeff	95% C I
σ_{yn}	0.40***	[0.21 0.78]	0.40***	[0.23 0.75]	0.22	[-0.18 0.66]	0.22	[-0.16 0.68]
Number of children	-0.32***	[-0.56 -0.14]	-0.32***	[-0.55 -0.15]	-0.25*	[-0.57 0.04]	-0.25*	[-0.56 -0.02]
PDDEB_Y	-0.01	[-0.06 0.04]	-0.02	[-0.07 0.04]	-0.01	[-0.10 0.08]	-0.01	[-0.10 0.09]
Child is girl	0.04	[-0.05 0.14]	0.01	[-0.14 0.16]	0.19***	[0.07 0.37]	0.19	[-0.07 0.45]
PDDEB_Y*Girl			0.01	[-0.03 0.05]			-0.01	[-0.08 0.05]
N	2,834				992			

***p<0.01 **p<0.05 *p<0.1
All regressions are controlled for family and child characteristics.
Child's age, religion, ethnic group
Child's parent's age, education level, and economics activity
Family standard of living level

FIGURE 2. Probability of post-primary school attainment by settlement and number of children

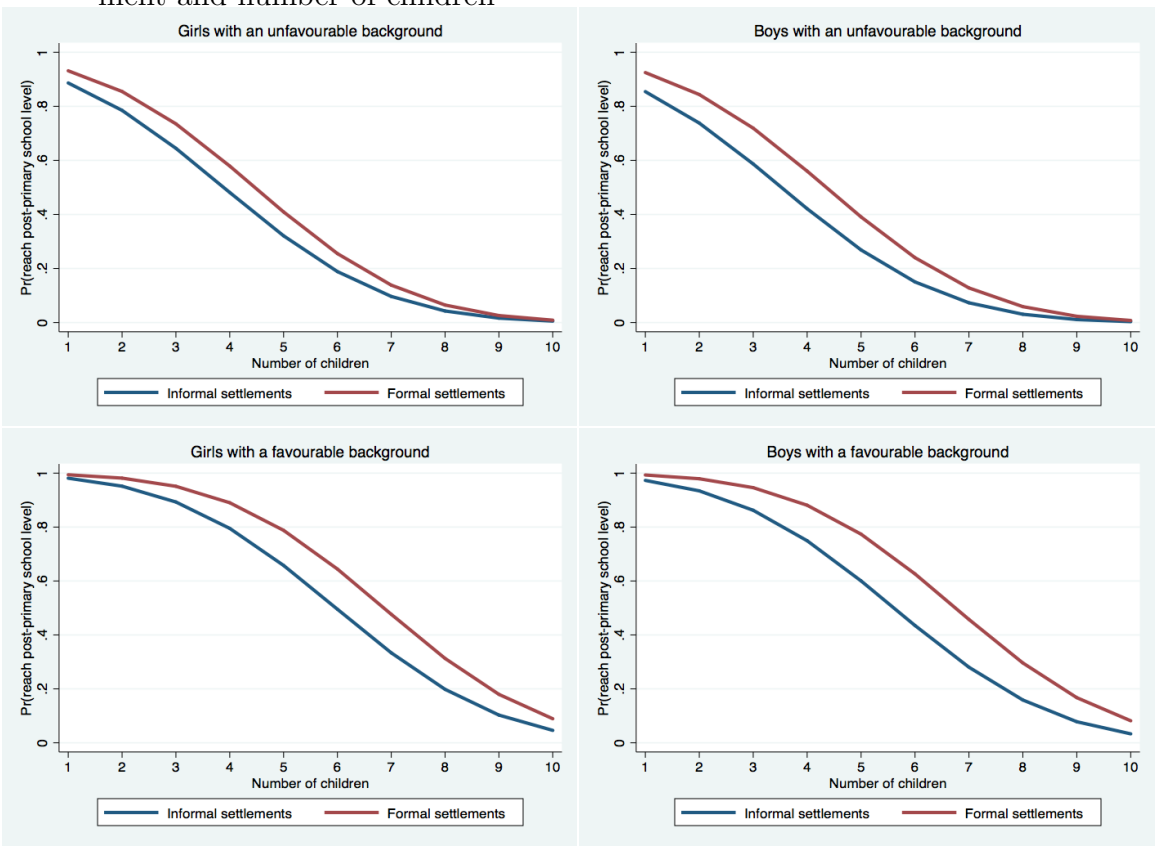

FIGURE 3. Settlement difference in probability of reaching the post-primary school level by standard of living and number of children in the family (95% confidence interval)

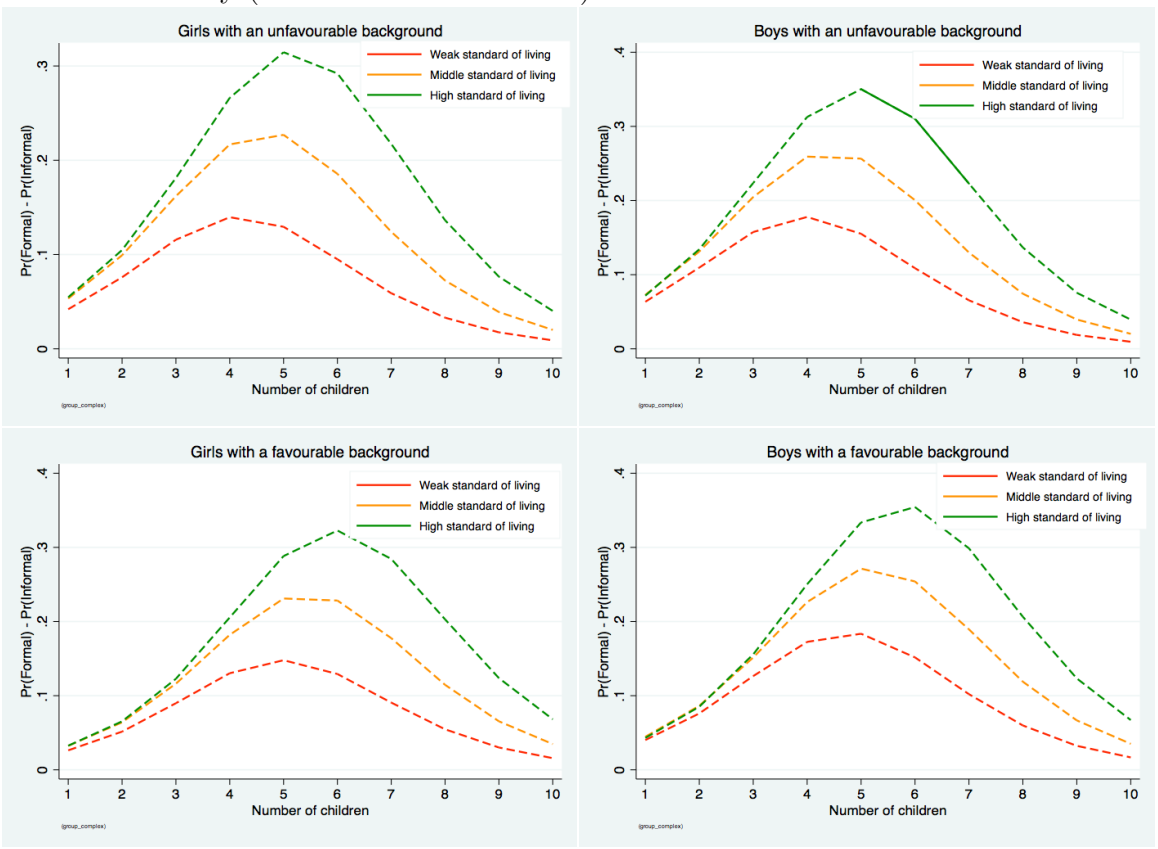

FIGURE 4. Settlement difference in probability of reaching the post-primary school level by standard of living and number of children in the family (90% confidence interval)

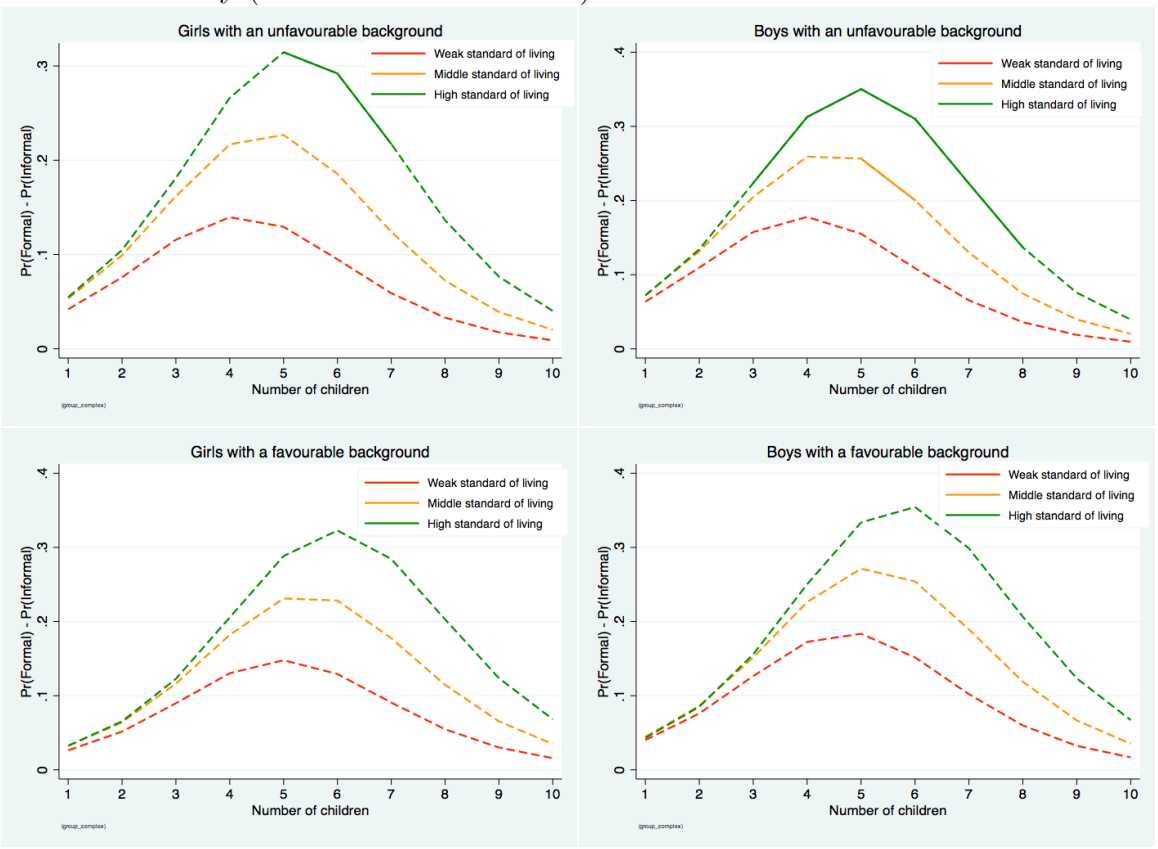

FIGURE 5. Settlement difference in probability of reaching the post-primary school level by exposure to PDDEB and number of children in the family (95% confidence interval)