Courant Research Centre
'Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis'

Georg-August-Universität Göttingen
(founded in 1737)

Discussion Papers

No. 183

The Nutrition-Learning Nexus: Evidence from Indonesia

Maria C. Lo Bue

August 2015

Wilhelm-Weber-Str. 2 · 37073 Goettingen · Germany
Phone: +49-(0)551-3914066 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: http://www.uni-goettingen.de/crc-peg

The Nutrition-Learning Nexus: Evidence from Indonesia

Maria C. Lo Bue[±]

Abstract: This paper investigates the effect of nutritional status on subsequent educational achievements for a large sample of Indonesians children. I use a long term panel data set and apply a maternal fixed effect plus an instrumental variables estimator in order to control for possible correlation between some of the components of the error term and the main independent variable which will likely to cause a bias in the estimates. Differences in nutritional status between siblings are identified by using exposure in the earliest months of life to the drought associated with the Indonesian wildfires of late 1997. Estimation results show that health capital (measured by height-for-age z-scores at childhood) significantly and positively affects the number of completed grades of schooling and the score on cognitive test. Nevertheless, I only find little robust evidence of an effect on the readiness to enter school.

JEL Classification: I12, I20, O15, O53

Keywords: Educational achievement, child nutrition, siblings' difference models, environmental shocks, Indonesia.

[±] Department of Development Economics, University of Göttingen.
mlobue@uni-goettingen.de

1. Introduction

Over the last 20 years, a growing number of studies have documented the existence of positive complementarities between schooling outcomes and childhood nutrition. There are two main channels that -interacting with each other- relate these two important human development aspects.

First, as nutritional intake in early childhood contributes to determine the health capital of an individual, poorly nourished children will be likely to be more vulnerable to disease or simply physically weaker and this in turn, will affect the development of cognitive skills, their ability to learn and to regularly attend classes and thereby their education performance. Given that cognitive development and school achievements are two important components of human capital, the long term consequences of poor nutritional status will also likely be reflected in worse outcomes on labor productivity and on life time earnings.

Second, children with poor nutritional status are exposed to higher risks of morbidity and therefore may enroll later at school. This is especially true in a developing countries context where the rule enforcement on compulsory school may be relatively weak and the economic returns to investments in health capital are relatively large. Although delayed enrolment decisions are rational caregivers' responses to early childhood malnutrition (Glewwe and Jacobi, 1995), late entry is never optimal as it will result in fewer years of earnings because in order to maintain total years of schooling with delayed enrolment, an individual will have to enter the job market later[1].

In this paper, I conduct a micro-econometric analysis on a long term panel dataset collected in 13 Indonesian provinces over the period 1997 to 2007 in order to gauge the impact of child nutritional status on subsequent educational attainments. The econometric strategy followed here is based on an instrumental variable-mother-fixed-effects estimator where differences in nutritional status among siblings are identified by the exposure to an exogenous shock, namely to the drought associated with the late 1997's Indonesian forest fires.

This study therefore seeks to contribute to previous literature in a number of ways. First, by empirically assessing the causal relationship between nutrition and learning, it extends the literature on human capital formation and overcomes the weaknesses of cross sectional studies. The bulk of empirical evidence on this field has indeed been built on cross sectional data and/or relied on econometric strategies with low accuracy, and so it has only been able to provide evidence of strong associations between child nutrition and school attainments, yet the causal relationship has so far only been investigated in few studies (Glewwe and Miguel, 2008).

[1] This implies that delayed enrolment doesn't necessarily lead to fewer years of education; despite this may be an expected outcome if one supposes that the opportunity cost of schooling increases with age.

Second, this paper aims at improving current knowledge on the impact of shocks at the individual level. These kinds of adverse events can drastically affect households' welfare, by generating substantial reductions in their levels of income and consumption (Morduch, 1995; Townsed, 1995) but the magnitude and the duration of such shocks may vary substantially among households' members. By considering the effect of the drought on children nutritional status and –through this- on later educational achievements, this paper contributes to current knowledge of both short and long term consequences of exposure to transitory environmental shocks

Third, this analysis will also extend current understanding on the consequences of the Indonesian wildfires. To date, only two studies have investigated the effects of this massive environmental shock on the Indonesian population by focusing on the impact of haze on adult respiratory problems (Frankenberg et al. 2005) and of air pollution on under-3 mortality rates (Jayachandran, 2009).

Lastly, this study contributes to build knowledge on the strength of the nutrition-learning nexus in Indonesia: a country which has been growing remarkably over the last 20 years and which is recently experiencing large reductions in poverty rates. The Indonesian education sector has benefited from massive supply side interventions which boosted school enrolment rates (Duflo, 2001). Yet, despite these gains, there are still some challenges that the country needs to face in terms of disparities within and between provinces and regions in many quantitative and qualitative indicators of school achievement (World Bank, 2011).

The remainder of this paper is organized as follows: I present the theoretical framework and its econometric implications in Section 2 and then describe the data in Section 3. Section 4 deals with the empirical approach and instrumental validity and illustrates main findings and robustness checks. The final section concludes.

2. Theoretical framework and econometric implications

I start with a simple model that shows the relationship existing between child health and educational achievements.

Following Alderman et al. (2001a, 2006); Glewwe et al. (2001); Cunha et al. (2006) and Yamauchi (2008), I begin with a dynamic process of human capital accumulation where the efficiency in the production of educational outcomes realized in a given time period is determined by physical and mental abilities and by the health and nutritional inputs which a person has cumulated up to that time.

In other words, there is an input-output relationship between health and education which can be formally represented by the following achievement function:

$$E_{it} = f(H_{it-1};\ X_{it};\ \alpha;\ \rho;\ \tau) \qquad (1)$$

where E_{it} is the educational achievement of child i (realized in time t); H_{it-1} is child health formed in the preceding period, which is meant to represent the child health history up until the beginning of time t[2] and X_{it} is a set of household and community characteristics that influence educational performance (e.g. the availability of schools and learning facilities, teachers' and parents' levels of education; household wealth).

Moreover, this function shows that achievement also depends on the child endowment (α) which is represented by her innate ability and motivation; by a vector ρ that, encompassing school fees and prices of consumption goods and of schooling materials, determine the household budget constraint and by a τ component which is common across siblings and includes parental tastes and attitudes for child education and health. Both ρ and τ are supposed to influence educational performance indirectly though the effects that they exert on the supply of educational inputs and, more generally, on schooling investment.

An important consideration underlying the achievement function is that there is dynamic household behaviour which contributes to shape the simple input-output relationship between health and school and possibly interacts with the degree of complementarity or substitutability between health capital and schooling inputs (Yamauchi, 2008).[3]

The above-specified achievement function can be transformed into a schooling equation which can be empirically estimated:

$$E_{it} = \beta_0 + \beta_1 H_{it-1} + \beta_2 C_{it} + \varepsilon_i \qquad (2)$$

where H_{it-1} can be proxied by the nutritional status of the child in the preceding period (e.g. given by her height-for-age z-scores) and C_{it} is a vector including the household and community

[2] This health capital is partly determined by individual genetic endowment and by time invariant parental and environmental factors as well as by a vector of prices and household characteristics which may vary across siblings and determine the level and efficacy of parental investments in child health.

[3] As noted by Yamauchi (2008), the optimal level of schooling investment is also affected by whether health capital substitutes for schooling inputs or it accrues their productivity. Assuming perfect substitutability implies that parents will make more schooling investments in unhealthier children. On the other hand, if health capital and schooling inputs are complementary, only healthier children will attract more schooling investment.

characteristics (X_{it}) and ρ prices. Lastly, ε_i is a disturbance term which represents the sum of the α component (i.e. child genetic potential, innate ability and motivation); the τ home-invariant component and the ξ_{it} white-noise error term component.

The main interest of this paper is, of course, to assess the magnitude and significance of the coefficient β_1 but there are a number of econometric problems entailed in the estimation of (2) which should first be addressed.

First, there may be an omitted variable bias problem, i.e. there may be other factors which relate to both height and the outcome variable that drive the association. The consumption of certain micronutrients, such as iron, vitamin A and iodized or fortified salt, for example, may affect nutritional status as well as learning abilities.

Second, parents observing the weak health status of their child, might engage in compensatory actions by increasing the amount of food or other health resources devoted to her or by diverting more resources for her educational achievement (given that the assumption of substitutability between health capital and educational inputs holds).

Third, as suggested by the medical and biology literature, not only the individual genetic endowment highly correlates with health status (see, *inter alia*, Preece, 1996 and Weedon et al., 2008), but also there is evidence of a shared genetic architecture between height and intelligence (Marioni et al., 2014; Keller et al., 2013; Posthuma et al., 2010; Silventoinen et al., 2006; Sundet et al., 2005; Van Dam et al., 2005). Therefore, over subsequent time periods children with higher genetic potential will be healthier than their peers while -on the other hand- less endowed children will be more likely to experience worst health conditions and might even die before the educational outcome is realized, leaving us with a “biased” sample of selected healthier individuals (Alderman et al., 2006; Yamauchi, 2008).

These considerations imply that simple ordinary least squares (OLS) estimates of β_1 are likely to be either upward or downward biased because the main independent variable can be not orthogonal to the error term. In other words, there is an endogeneity problem due to the possible correlation existing between H_{it-1} and τ and/or between H_{it-1} and α (Behrman, 1996; Alderman et al. 2006).

As suggested by Glewwe et al. (2001), the econometric approach that can best sweep out these two forms of correlation combines a sibling difference model with instrumental variable techniques: maternal fixed effects will indeed remove the bias caused by the correlation between the endogenous variable and the siblings-invariant error term component, while the use of a relevant and exogenous instrument will purge the remaining correlation with the child specific error term component.

Lastly, as it is clear from the model, the estimation of (2) requires data measured at different points in an individual's life. Cross-sectional datasets can include such variables only if they are based on questions relying on the respondent's memory but this would easily entail problems of data reliability and measurement error. As noted by Glewwe and Miguel (2008), the vast majority of studies which were conducted using cross sectional data is indeed flawed as these rely on very strong and often untestable assumptions in order to make inference.

To date, the literature on this research field includes only four published studies (i.e. Alderman et al., 2001a; Glewwe et al., 2001; Alderman et al., 2006; Yamauchi, 2008) that have used panel data to estimate the impact of child health on later education achievements.

Table 1 gives an overview on these studies by summarizing the main information on the country on which the research was based, the variables used for educational achievement, the type of estimation approach, the endogenous health status variable and the variable that was chosen to instrument for it.

Table 1 Review of previous panel data based studies on the nutrition-learning nexus

Authors	Country	Educational outcome variable(s)	Estimation strategy	Endogenous health status variable	Instrument(s)
Alderman et al. (2001a)	Pakistan	Dummy =1 if enrolled in time	IV	Height for age z-scores	Food Price shocks
Glewwe et al. (2001)	Philippines	Test scores	HHFE-IV	Height for age z-scores	Height of the older sibling by the age of 24 months
Alderman et al. (2006)	Zimbabwe	Adolescent height; grades attained; age starting school	MFE-IV	Height for age z-scores	Exposure to civil war and to drought
Yamauchi (2008)	South Africa	Age starting school; grades completed; Mathematics test scores	HHFE-AFE-IV	Height for age z-scores	Community health facilities; Weight for age z scores

Note: IV stands for Instrumental Variable; HHFE is household-fixed effects; MFE is mother fixed effects and AFE is age fixed effects.

All these studies have found a strong and often statistically significant effect of child nutritional status on later academic achievement. However, it should be noted that-except for the Zimbabwe study- the validity or the relevance of the chosen instrument can be subject to criticism. As indeed noted by Glewwe and Miguel (2008) and Alderman et al. (2006), the use of food price shocks in the Pakistan study may be problematic because these shocks can easily affect household savings and therefore also schooling outcomes (via the effect on schooling investment). The Philippines study, on the other hand, is built on the questionable assumption that growth in children's height after age two is not correlated with height up to the age of two. Lastly, in the South Africa study, although weight for age was used with the purpose of sweeping out measurement error, its exogeneity may be easily subject to criticism as weight for age represents itself a facet of the (endogenous) nutritional status. Moreover, the estimation strategy followed in the majority of these studies was not always ideal: while the Pakistan study is only based on the instrumental variable (IV) strategy, the Philippines and the South Africa studies follow a household fixed effect-instrumental variable estimation approach. Household fixed effects, nevertheless, are not exactly the same as differencing across siblings of the same mother, especially in a developing country context where more than one family units share the same house.

3. Data and Sample

3.1 The Indonesian Family Life Survey

Our main source of data is the Indonesian Family Life Survey (IFLS) which is an ongoing longitudinal survey of individuals, households, communities and facilities which was conducted in 13 Indonesian provinces spread out in the islands of Sumatra, Java, Kalimantan, Sulawesi, Bali and West Nusa Tenggarra. The first wave (IFLS1) was conducted in late 1993 and surveyed 7,224 households and 22,000 individuals in 321 enumeration areas. Between August and December 1997 the second full sample wave (IFLS2) successfully managed to reinterview over 94% of the IFLS1 households. Other two follow ups surveys were conducted in 2000 (IFLS3) and 2007 (IFLS4). Among the IFLS1 households 90.3% were either interviewed in all the four waves or died, and 87.6% were actually interviewed in all four waves (Frankenberg and Thomas 2000).

There are interesting features in the IFLS which make this data particularly suited to my research needs. First, these high recontact rates contribute significantly to data quality by lowering the bias due to non random attrition. Second, in addition to basic demographic and socio-economic characteristics of the respondents, the IFLS collected detailed information on various educational aspects (e.g., current schooling grade; age at which the child first enrolled at school;

number of correct answers given in a cognitive test) as well as on the anthropometric measures which are necessary to derive child nutritional status variables.

3.2 Description of key variables

I consider the panel of individuals surveyed in IFLS2, IFLS3 and IFLS4 and I shrink the initial IFLS2's size by keeping only ten cohorts of individuals born between 1990 and 1997[4]. These children are then tracked after three years (i.e. in IFLS3) and/or after ten years (i.e. in IFLS4) in order to get information on the current educational achievement. Data show that in 2000 and/or in 2007, 936 observations were traced from an initial sample of 2181 children[5] for which there was complete information available on basic demographic and socio-economic characteristics and on anthropometric measures such as weight and height that were used to construct sex and age-standardized z-scores for height and weight based on the standards provided in 2006 by the World Health Organization (WHO) in the Multicentre Growth Reference Study (MGRS)[6].

In order to consider different facets of learning achievements, education outcomes are measured by three distinct variables[7]: a) completed years of schooling; b) score obtained in a cognitive test (to consider the development of cognitive and learning skills), and c) age at which primary school was started (to proxy for readiness to enter school). The first variable is observed in 2007 (IFLS4) and is measured by summing the number of grades completed at each level of school[8]. The score on cognitive test is measured either in 2000 or in 2007, depending on child age. This test was indeed administered to children aged 8 to 14 years and the variable on its outcome is constructed as the ratio of correct answers on total questions[9]. Lastly, the information on the age at which the child started school is directly taken from the answer provided by the mothers either in IFLS3 or in IFLS4.

When observing the information available on nutritional status, it can be noted that on average the children sampled have poor height and weight for age relative to the MGRS sample of

---

[4] Despite the availability of one more wave of data from the IFLS administered in 1993, this was not used as the baseline survey, given that in my identification strategy I just needed observations whose child nutritional status was measured at one point in time during early childhood and I needed to include also children born after 1993 since the instrument used in this analysis identifies children who are aged 12-36 months in September 1997 (see Section 4.1 for further details).

[5] More details about tracking issues are discussed in Section 4.3.

[6] The MGRS was based on a sample of 8500 children from widely different ethnic backgrounds and cultural settings (Brazil, Ghana, India, Norway, Oman and USA). These children were breastfed during infancy, appropriately fed later on in life and raised in optimal conditions (WHO, 2006).

[7] See Table A1 in the Appendix.

[8] The Indonesian school system consists of six years of primary education, three years of junior secondary and other three years of senior general or vocational education. Primary school starts by law at the age six or seven.

[9] The test consisted of a set of 17 questions, of which 12 were cognitive and 5 were based on simple mathematics.

adequately nourished children. As suggested by the figures reported in Table A1 in the Appendix, given age and sex, the child height (weight) is -1.74 (-1.44) standard deviations below the median child in that age group.

Also, a relatively large percentage of children suffer from mild and moderate stunting and underweight conditions. Table 2 reports that –on average- while only 15.25% (7.15%) of children are severely stunted (underweight), and this figure grows of about 25-29 percentage points when considering moderate stunting and underweight conditions.

Lastly, as illustrated in Table 3, children who are suffering from moderate to severe stunting conditions are more likely to experience worst educational outcomes in later stages of their life if compared to their healthier peers.

Table 2 Nutritional status of children surveyed in IFLS2, by gender and place of residence

	Gender		Residence		Total
	Female	Male	Urban	Rural	
Stunting					
Mild (Height-for-Age z-score <-1 SD)	73.70%	75.97%	70.48%	79.36%	74.87%
Moderate (Height-for-Age z-score <-2SD)	41.30%	46.64%	40.12%	48.09%	44.06%
Severe (Height-for-Age z-score <-3 SD)	15.22%	15.27%	13.10%	17.45%	15.25%
Underweight					
Mild (Weight-for-Age z-score <-1 SD)	68.70%	70.47%	65.98%	73.35%	69.61%
Moderate (Weight-for-Age z-score <-2SD)	29.78%	35.03%	29.46%	35.61%	32.49%
Severe (Weight-for-Age z-score <-3 SD)	7.17%	7.13%	7.47%	6.82%	7.15%

Source: Author's elaboration from IFLS2 and WHO-MGRS.

Table 3 Mean educational achievements of children above and below moderate stunting and underweight thresholds

	Children with moderate to severe stunting	Non-malnourished	Children with moderate to severe underweight	Non-underweight
Completed years of schooling	6.55	6.87	6.84	6.67
Age start school	6.34	6.20	6.41	6.19
Cognitive Test Score	0.73	0.76	0.73	0.76

Source: Author's elaboration from IFLS 2-3-4

4 Findings

4.1 Estimation approach and instrumental validity

The empirical approach pursued in this paper is based on the estimation of the afore-specified schooling equation where three alternatives measures for E_{it} are used: completed years of schooling, the score obtained in the cognitive test and the age at which primary school was started. The effect of H_{it-1} (measured as height-for-age z-scores) is estimated by mainly relying on a mother-fixed effects-instrumental variable (MFE-IV) model, which-as argued before- addresses endogeneity in the relationship of interest.

It is important to note that the MFE-IV estimation procedure entails the choice of an instrument that should significantly affect a child nutritional status, be adequately variable across children born from the same mother and sufficiently transitory in order to not exert any direct effect on E_{it}.

The instrument that I use is the shock resulting from individual exposure to the drought associated to the late 1997's Indonesian forest fires.

From September to November 1997 large traits of tropical forest and arable land belonging to the islands of Sumatra and Kalimantan were severely hit by the worst fires ever recorded in Indonesia. The damage inflicted by these wildfires and the resulting haze was massive: the lives of the majority of the population living and working in rural areas were adversely affected by the destruction of farms and plantations, the interruption in transports and the bad respiratory problems which resulted from months of breathing heavy smoke-haze (Frankenberg et al. 2005; Jayachandran, 2009).

In Sumatra and Kalimantan, small and controlled burns have been traditionally used by small-scale farmers to clear land for planting new crops[10]. But these burns went quickly out of control from early September 1997 because of the extraordinary dry weather conditions that were brought by the El Niño Southern Oscillation (ENSO; Jim, 1999) phenomenon. Therefore, the drought associated with El Niño which exacerbated the intensity of the fires became particularly severe and prolonged. Only in mid-to-late November, when fires were quenched by the first rains, land and environmental conditions in these two islands began to recover (see Figure A1 reported in the Appendix).

[10] This type of "slash and burn" techniques which play an important ecological role for the local ecosystem have, nevertheless, been increasingly used in more extended areas during the Nineties because of the expansion of timber and palm oil industries in Indonesia.

The instrumental variable is therefore constructed as a dichotomous variable which equals one if the child was living in Sumatra or Kalimantan and was aged 12 to 36 months when the forest fires began (i.e. at September, 5th, 1997). Hence, variation in the exposure to such shock mainly comes from two sources: place of residence and age.

The choice of a specific age range is motivated by the large number of contributions provided by nutritionists, physiologists and social scientists which have posited the existence of a "critical period" in human life where brain development is most sensitive to poor nutrition (Dobbing, 1976; Waber et al., 1981; Stein et al., 1976; Villar et al., 1984; Glewwe and King, 2001). There have been mixed findings, however, concerning the exact age range where this critical period can be identified, although the bulk of the literature agrees that the impact of shocks on children older than 36 months is zero (Hoddinott and Kinsey, 2001; Glewwe and King, 2001; Shrimpton et al., 2001).

I've therefore conducted a preliminary analysis where I tested for different age range as well as for foetal exposure[11] and found that the shock experienced during the second and third year of life had the largest negative impact on child health status.

Table 4 reports the mother fixed effects estimates of the effect of the exposure to the forest fires on child nutritional status measured as height-for-age z-scores (Columns 1 and 2) and as weight for age z-scores (Columns 3 and 4)[12].

Results indicate that across all the specifications, there is a negative and significant effect at the 1% level on the endogenous variables and the magnitude of the effect is relatively larger for height-for-age z-score. Moreover, as suggested by the F test statistic, the instrument's validity (at least with respect to the strong correlation with the endogenous variable) is particularly high[13], and well above the thresholds recommended in Staiger and Stock (1997) and Bound et al. (1995).

A last important point to be discussed concerns the second condition for instrumental validity, i.e. the exclusion restriction assumption which implies that the only way though which the instrument affects the dependent variable is via its impact on the endogenous variable.

---

[11] Fetus exposure is proxied by health conditions experienced by mothers during pregnancy. All these tests, not shown in this paper, are available upon request.

[12] Note that weight and height were measured shortly after the beginning of the forest fires (i.e. from September 1997 to March 1998). The few observations (about 2.7% of the sample) which were measured before September 1997 were dropped.

[13] Additional tests, like the Anderson canonical correlation test for under-identification and the Cragg-Donald F statistic for weak identification have been implemented and their outcome further supports the validity and exogeneity of the instrument (see Table 6).

Table 4 Effect of exposure to forest fires on child nutritional status. First stage estimates

	Height for Age z-scores		*Weight for Age z-scores*	
	(1) Without controls	(2) With controls	(3) Without controls	(4) With controls
Exposure to Forest Fires	-1.103***	-1.181***	-0.744***	-0.971***
	(0.220)	(0.232)	(0.225)	(0.232)
Boy		-0.0887		-0.00746
		(0.121)		(0.111)
Age		-0.0374		-0.0965***
		(0.0277)		(0.0285)
Constant	-1.668***	-1.480***	-1.395***	-1.026***
	(0.0134)	(0.124)	(0.0137)	(0.123)
F-statistics on significance of the fires' shock	25.12***	25.95***	10.96***	17.50***
Observations	936	936	936	936
R-squared	0.092	0.101	0.048	0.111
Number of m.id	711	711	711	711

Note. Estimations are based on mother fixed effects.
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Given that there isn't any statistical test that can be performed to check whether this assumption is violated (at least when using one instrument only), the instrument validity can't never be known with complete certainty and can only by checked indirectly or falsified by the data. Therefore, I will investigate here some auxiliary hypothesis or implications which can add plausibility to the exclusion restriction.

One can think, indeed, of different situations that can violate the exclusion restriction in this context. The forest fires, for example, may have exerted their effect on children educational achievements through two other possible channels: 1) on the supply-side, they may have destroyed books, schools and harmed teachers and 2) on the demand-side, they also could have negatively affected household income and thereby probably depleted parental resources devoted to education.

With respect to the first point, it can be argued that this channel actually doesn't seem to have mattered: the damages that were reported by the press and by the literature basically consisted in the burning of millions of hectares of wild forest and in the spread of smoke and haze. The state of emergency declared by the Government of Indonesia, that implied the temporary closing of school, government offices, business, airport and port, lasted only ten days (Dauvergne, 1998). Since in this paper, I am identifying children who were hit by the fires in their earliest months of life and who therefore went to school several years later, this supply-side channel is conceivably not relevant.

Similarly, it can be argued that whether the forest fires may have hit household income as well, as long as these economic shortages were quite temporary, parents' investments in education were likely to not be affected by these income losses.

In order to see if this argument is actually confirmed by the data, I have tested (see Table A2 in the Appendix) whether there are significant differences in the effect of exposure to forest fires on household per capita expenditure along the time period considered in this paper and whether exposure to forest fires significantly affected changes in the share of education expenditure.

Results contribute to relief further concerns on instrumental validity as they clearly indicate that the instrument neither significantly affected household income in any of the years considered nor it had any impact on education expenditure.

4.2 Empirical Estimates

Before discussing the main findings based on the preferred MFE-IV approach, I will start by showing the estimates from three alternative and less precise econometric approaches, which despite being affected by some bias, are still interesting to the extent that they can provide a first picture on the strength of the nutrition-learning nexus, and-when compared to our core model-they inform about the magnitude and the direction of the endogeneity bias in the relationship under investigation. These approaches are: (1) the Province-fixed effects (PFE) that essentially control for unobserved heterogeneity within each province; (2) the Province fixed effect combined with instrumental variable (PFE-IV) that address only the correlation between child height and child specific characteristics; and (3) the Mother fixed effects (MFE) which only sweep out the bias due to aspects which are common across sibling. Results are presented in Table 5.

The estimates from the two 'naïve' PFE and PFE-IV regressions suggest that there are large positive associations between child height and educational achievements: thereby, children with higher nutritional status tend to complete more grades, perform better in the cognitive test and enter school at younger ages. The relatively large size of these coefficients is, however, much reduced once one controls for the endogeneity bias related to the correlation between child height for age and the sibling-invariant error term component (i.e. supposing a certain degree of substitutability between nutrition and school inputs, parents can engage in compensatory actions in order to equalize learning outcomes among their kids).

Yet the simple within-sibling estimator does not account for the remaining correlation between the endogenous variable and the specific error term component. The preferred mother-fixed effects-instrumental variable estimator, however, contributes to reduce attenuation bias from measurement errors, and-relying on the assumption that the shock is independent of child

Table 5 Child height for age and subsequent educational achievements. "Naïve" approaches

Outcome:	(A) *Years of schooling*			(B) *Cognitive test scores*			(C) *Age starting school*		
Model:	(1) PFE	(2) PFE-IV	(3) MFE	(4) PFE	(5) PFE-IV	(6) MFE	(7) PFE	(8) PFE-IV	(9) MFE
Height for Age (z scores)	0.104***	0.637*	-0.0171	0.00579	0.0490	0.0211*	-0.0410***	-0.309	-0.0647**
	(0.0323)	(0.350)	(0.0683)	(0.00343)	(0.0631)	(0.0120)	(0.00901)	(0.199)	(0.0324)
Boy	-0.287***	-0.219**	-0.203	0.000617	0.00448	0.00137	0.162***	0.129**	0.109
	(0.0927)	(0.111)	(0.162)	(0.0155)	(0.0142)	(0.0234)	(0.0285)	(0.0558)	(0.0777)
Age	0.839***	0.871***	0.884***	0.000810	0.00241	0.00600	0.0400***	0.0199	0.00768
	(0.0280)	(0.0332)	(0.0307)	(0.00274)	(0.00365)	(0.00550)	(0.0124)	(0.0194)	(0.0161)
Mother Education	0.139***	0.107***		0.00738***	0.00477		-0.0428***	-0.0265*	
	(0.0134)	(0.0250)		(0.00198)	(0.00429)		(0.0101)	(0.0147)	
Mother Age	-0.00454	-0.00869		0.00395***	0.00382***		-0.00199	0.000126	
	(0.00697)	(0.0117)		(0.000932)	(0.00135)		(0.00519)	(0.00518)	
Observations	928	928	928	783	783	783	834	834	834
Number of prov.id	13	13		13	13		13	13	
Number of moth.id			706			609			652
R^2 (within)	0.610	0.512	0.776	0.034	-0.077	0.024	0.098	-0.199	0.034
Anderson canon.corr.LR.stat		11.33			6.96			10.08	
		(0.001)			(0.008)			(0.001)	
Cragg-Donald F statistic		11.34			6.94			10.08	

Note: The sample consists of children with height for age z score in the range -6 to 6 and aged 0-8 in 1997. These children were subsequently aged 5-11 when started school, 8-14 when they took the cognitive test (either in 2000 or 2007) and 9-17 when completed years of schooling are observed (i.e. in 2007). PFE= Province Fixed Effects; PFE-IV= Province Fixed Effects combined with Instrumental Variable; MFE= Mother Fixed Effect. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

specific characteristics- allow us to identify the children with greater educational achievements than average (Card, 2001; Alderman et al., 2006).

According to our initial expectations, main findings from the MFE-IV regressions reported in Table 6, show that greater child height significantly contributes to improved educational performance, both in terms of completed years of schooling and of cognitive skills.

Results indicate that if the children sampled in this analysis had the nutritional status of the well nourished population of reference, they would have gained on average 0.8 additional grades of school. Interestingly this estimated impact is pretty close to the one found by Alderman et al. (2006).

It can be assumed, however, that part of the effect that child nutritional status has on future schooling achievement is transmitted by the age at which the child entered school. In order to detect this transmission channel, a dummy indicating whether the child entered the school on time (i.e. by the age of 7) was added to the baseline MFE-IV model (see Col. 2; 3; 5; 6). Results show that the inclusion of this variable indeed mitigates the magnitude of the impact exerted by height-for-age on the dependent variable, but leaves unaltered the statistical significance of the coefficient.

Lastly, since most of the children have not completed their schooling in 2007, a further specification (Column 3; 6 and 8) replaces the age regressor with age dummies that can better standardize the years of schooling (see Yamauchi, 2008). Here again the positive and significant results are confirmed.

When we look at the estimates on the score achieved on the cognitive test (Col. 4, 5 and 6), results indicate that increased height-for-age is associated to better performance on the test: a child would have obtained 18% more in her score if she had experienced the same nutritional status of her well nourished peers.

Consistently with the findings on years of schooling, the inclusion of the instrument variable estimator increases the magnitude of the impact, suggesting the presence of a downward bias in the within sibling estimator. Moreover, improvements in cognitive skills seem to be not significantly associated with readiness to enter school (see Col. 5), although they are positively associated with the age at which the test was taken, possibly implying some positive returns of years of schooling on the development of cognitive skills. The statistical significance of the estimated coefficient on height-for-age z-score is, nevertheless, weaker than the one found in the years of schooling models.

Table 6 The effect of child height for age on subsequent educational achievements. Main Findings, based on the MFE-IV estimator

	(A) *Years of schooling*			(B) *Cognitive test scores*			(C) *Age starting school*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Height for Age	0.480**	0.382**	0.455**	0.0790	0.107*	0.103*	-0.167	-0.271
(z scores)	(0.236)	(0.184)	(0.227)	(0.0589)	(0.0612)	(0.0586)	(0.118)	(0.172)
Boy	-0.134	-0.111	-0.141	0.00482	0.00764	0.00434	0.0940	0.0624
	(0.174)	(0.160)	(0.177)	(0.0259)	(0.0251)	(0.0234)	(0.0787)	(0.0802)
Age	0.885***			0.00792			0.00586	
	(0.0330)			(0.00610)			(0.0169)	
Entered school by the age of 7		0.0597	0.285		-0.0306	-0.0268		
		(0.252)	(0.257)		(0.0438)	(0.0404)		
Observations	928	885	885	783	769	769	834	834
Number of moth.id	706	682	682	609	602	602	652	652
Age Fixed Effects			YES			YES		YES
R^2 (within)	0.736	0.801	0.810	-0.112	-0.288	-0.140	-0.010	0.041
Anderson canon.	23.08	25.90	17.64	17.31	17.52	15.07	23.64	12.79
corr.LR.stat	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cragg-Donald F statistic	23.99	27.08	17.52	17.88	18.03	14.92	24.83	12.16

Note: The sample consists of children with height for age z score in the range -6 to 6 and aged 0-8 in 1997. These children were subsequently aged 5-11 when started school, 8-14 when they took the cognitive test (either in 2000 or 2007) and 9-17 when completed years of schooling are observed (i.e. in 2007). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Lastly, as argued before, part of the positive effect that child nutrition has on schooling achievement could be transmitted by the optimal timing of entering primary school and indeed, a negative direct relationship is found between height-for-age z-score on age starting school (i.e. better nourished children start school at younger ages). Nevertheless, the hypothesis that child nutrition improves readiness to enter primary school only finds weak statistical support in our regressions (see Columns 7-8).

4.3 Robustness checks

This section presents several checks which were conducted in order to test for the validity of the findings of the present analysis. Hence, I will deal with issues related to selection bias in the data and address other concerns related to the robustness of the main results.

Potential attrition bias

One of the major concerns arising from the use of longitudinal datasets is related to the presence of selection bias which can be caused by deaths, missing data in fundamental variables or by the screening out of multiple observations recorded in the same wave and of observations with discrepant information provided across the survey's waves .

Since the main analysis of this paper is based on variations among siblings, any attrition which stems from maternal, household and community characteristics is removed by the inclusion of mother fixed effects (Ziliak and Kniesner, 1998). Yet, there remain some attrition at the individual level which is needed to be tested for.

Table 7 reports the determinants of attrition from the 1997-2007 waves. This test which follows the methods set out in Fitzgerald et al. (1998a, 1998b) and Alderman et al. (2001b), is based on a linear probability model where the dependent variable equals 1 if any of the educational outcome were observed in the second period and 0 otherwise. As explanatory variables, I use the same main variables included in the schooling equation: height-for-age z-scores; weight for age z-scores; child sex; child age and mother's years of schooling. If these are not significantly correlated with attrition, I can assume that there will be no bias in my estimates stemming from attrition on the observables.

Baseline and province fixed effects estimates reported in Columns (1) and (2) of Table 7 indicate that it is more likely to observe individuals with poorer initial height-for-age, of male sex, higher initial age (due to the higher rates of mortality during infancy) and born from more educated mothers (probably because of their better accuracy in answering the questionnaire).

Table 7 Determinants of attrition

	(1)	(2)	(3)
Height for Age z score	-0.0149*	-0.0164	-0.0136
	(0.00817)	(0.0106)	(0.00857)
Weight for Age z score	0.00501	0.00586	-0.00818
	(0.00985)	(0.0102)	(0.0107)
Boy	0.0228	0.0218	-0.0111
	(0.0200)	(0.0195)	(0.0175)
Age in 1997	0.0107**	0.00947**	-0.00273
	(0.00502)	(0.00431)	(0.00450)
Mother education	0.0139***	0.0136**	
	(0.00278)	(0.00501)	
Constant	0.165***	0.170***	0.306***
	(0.0313)	(0.0392)	(0.0216)
Province Fixed Effects	NO	YES	NO
Mother Fixed Effects	NO	NO	YES
Observations	2,181	2,181	2,181
R^2	0.014	0.013	0.015
Number of prov.id		13	
Number of moth.id			1,706

Note: Dependent variable=1 if educational outcome was observed, 0 otherwise. Initial sample is 2181 children born from 1990 to 1997 that had their height for age z scores or weight for age z scores or both observed in 1997 and laying in the range -6 and +6. From this sample, 936 children have their educational outcome observed in subsequent waves. Estimation method: linear probability model. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The mother fixed effects estimates reported in Column 3 suggest, however, that controlling for mother's fixed effects, none of these variables has a significant effect on the probability of being observed, thereby attenuating any concern about selection bias due to attrition.

Other checks

I address here three main concerns that could cast some doubt on the validity of the main analysis: robustness of the estimates to the inclusion of additional covariates in the MFE-IV regressions; adequacy of the indicator used for child nutritional status; and sensitivity of the results to slight changes in the timing of exposure to the shock.

Tables A3-A5 in the Appendix report the results on the three different measures of educational attainments used in this analysis.

In the first specification, the presence of non-linearities in age has been tested for: while, as expected, there are non linear effects in completed years of schooling, the positive sign on age squared in the test scores regression, suggest that cognitive skills improves at older ages, probably due to positive complementarities with the numbers of years spent at school.

It can also be observed that boys tend to have worse educational attainments than girls, especially in school completion and readiness to school. This result is consistent with the evidence provided in many World Bank and Asian Development Bank reports (see, for example, Asian Development Bank, 2006) and may be partly explained by the higher returns to schooling (at later stages of education) for women than for men (Deolikar, 1993).

Column 3 in Tables A3 and A4 include a dummy variable for children who were already at school at the time that their height was measured. It can be indeed argued that parents may easily alter the amount of nutrition and education inputs for their kids, once they observe their school performance. Doing this check, slightly decreases the magnitude of the height-for-age coefficient but leaves unaltered its statistical significance. Next, I include birth order and interaction terms between this and sex. There might be, indeed, a concern that there is a competition over resources among siblings or that there is a gender bias in parental preferences which is mediated by birth order (Das Gupta, 1987). Results suggest that higher order children tend to perform relatively worse and that –for completed years of school- the effect is stronger among boys.

A second set of robustness checks is related to the adequacy of the measure of nutritional status considered in the main analysis, height-for-age. The measurement of children height was undertaken few months after the shock took place. Given this short time interval, it can be argued therefore that the use an indicator of chronic nutritional deprivation, such as height-for-age z- scores, may not be adequate to correctly identify the effect of the shock on nutritional status. The first stage regressions shown in Table 4, already confirm the validity of the instrument on height-for-age z-scores and exhibit a smaller impact (both in terms of magnitude and statistical validity) of the shock on the indicator for previous and current nutritional deficiencies given by the weight-for-age z-scores.

Table A6 in the Appendix summarizes the results for the MFE-IV regressions that use weight-for-age z-scores as the endogenous variable. The estimated coefficients indicate that the magnitude of the effect is relatively larger than the one resulting in Tables 6 although the statistical significance is –on average- lower.

A last concern is related to the robustness of the age range 12-36 months used to identify the exposure to the shock. As argued in Section 4.1, the majority of the empirical studies have excluded that ages above 36 months matter but some of them find stronger impact at the age range 12-24. In the preliminary analysis that I conducted, I found –in accordance to Glewwe and King, 2001- zero impact below the second year of life and above the third year of life, but similar results come out when using both age range 12-36 months and 12-24 months, although in the latter case, the power of the tests for instrumental validity and exogeneity is relatively lower.

Table A7 in the Appendix summarizes the results for the MFE-IV regressions using as instrument exposure to forest fires during the second year of life only. In the years of schooling and cognitive test scores regressions the magnitude and statistical significance of the coefficients is slightly higher than in the main analysis whether the weak effect on readiness to school is further confirmed.

5. Conclusions

This paper investigated empirically the relationship between child nutrition and subsequent educational attainments using longitudinal data from Indonesia. By applying a sibling difference model combined with instrumental variable estimation, the estimates obtained address endogeneity biases. There are three relevant remarks which emerge from this study. First, results suggest that good nutritional status in childhood significantly contributes to improved, school attainments. Well nourished children complete 0.8 grades more than their less fortunate peers and perform better in cognitive tests (+18%). Poor nutritional status also tends to delay enrolment, although the strength of this relationship has weak statistical support. The main findings, which are confirmed by robustness checks, imply that from a policy perspective school and nutrition objectives should not be seen as competing goals but are closely interlinked. Many countries-including Indonesia- have been struggling with the full realization of the Millennium Development Goals. Although remarkable progress has been made, there is still consistent space to be improved and there is need to continue on stable and durable paths of development. Financial resources devoted to child nutrition policies do not necessarily compete with those for the education agenda; instead, as implied in this study, they can be regarded as a more cost effective way to raise present and future socio-economic development.

Second, in line with the growing body of scientific studies on the long-run impact of childhood shocks (see, inter alia, Akresh et al. 2012; Almond and Currie, 2011a, 2011b; Bhalotra, 2010), exposure to environmental disasters may have long lasting effects on individuals, despite any compensatory actions which they or their caregivers may undertake to alleviate the impact of the shock.

Third, consistent with Glewwe and King (2001), I do not find any support for the hypothesis that prenatal and first six months of life's nutritional conditions have long term effects on cognitive and education outcomes. Instead, the critical period where health shocks may have permanent consequences on human capital development is found in the second and third years of life.

An important implication of these two last remarks is that nutrition interventions or broader food security programs that target children in this critical period of their life not only substantially contribute to protect vulnerable people against uninsured shocks but they can also deliver positive long term outcomes in educational achievements and thereby contribute to future socio-economic development.

Bibliography

Akresh, R., Bhalotra, S., Leone,M., & Osili, U.O. (2012). War and stature: growing up during the Nigerian Civil War. *The American Economic Review,* 102 (3): 273-277.

Alderman, H., Behrman, J.R., Lavy, V., & Menon, R. (2001a). Child health and school enrolment: A longitudinal analysis. *Journal of Human Resources,* 36: 185-205.

Alderman, H., Behrman, J., Kohler, H-P., Maluccio, J., & Watkins, S. (2001b). Attrition in longitudinal household survey data: some tests for three developing country samples. *Demographic Research*, 5: 78–124.

Alderman, H., Hoddinott, J., & Kinsey, B. (2006). Long term consequences of early child malnutrition. *Oxford Economic Papers*, 58: 450-74.

Almond, D., & Currie, J. (2011a). Human capital development before age five, In *Handbook of Labor Economics*, ed. David Card & Orley Ashenfelter. Amsterdam Elsevier.

———. (2011b). Killing me softly: the fetal origins hypothesis. *Journal of Economic Perspectives* 25 (3): 153–72.

Asian Development Bank. (2006). *Indonesia: country gender assessment*. Manila: Asian Development Bank.

Bhalotra, S. (2010). Fatal fluctuations? Cyclicality in infant mortality in India. *Journal of Development Economics*, 93 (1): 7-19.

Behrman, J. (1996). The impact of health and nutrition and education. *World Bank Research Observer*, 11, 23-77.

Bound, J., Jaeger, D., & Baker, R. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *Journal of the American Statistical Association*, 90: 443–50.

Card, D. (2001). Estimating the return to schooling: progress on some persistent econometric problems. *Econometrica*, 69: 1127–60.

Cunha, F., Heckman, J., Lochner, L., & Masterov. D. (2006). Interpreting the evidence on life cycle skill formation. In *Handbook of the Economics of Education*, ed. Eric Hanushek and Finis Welch. Amsterdam: Elsevier.

Das Gupta, M. (1987). Selective discrimination against female children in rural Punjab, India. *Population and Development Review*, 13 (1): 77-100.

Dauvergne, P. (1998). The political economy of Indonesia's 1997 forest fires. *Australian Journal of International Affairs*, 52 (1): 13-17.

Deolalikar, A. B. (1993). Gender differences in the returns to schooling and in school enrolment rates in Indonesia. *The Journal of Human Resources* 28 (4): 899-932.

Dobbing, J. (1976). *Vulnerable periods in brain growth and somatic growth*. In D. F. Roberts and A. M. Thomson, eds. The Biology of Human Fetal Growth. London: Taylor and Francis.

Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: evidence from an unusual policy experiment. *American Economic Review*, 91 (4): 795–813.

Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998a). An analysis of sample attrition in panel data: the Michigan Panel Study of Income Dynamics. *Journal of Human Resources,* 33: 251–99.

———. (1998b). The impact of attrition in the panel study of income dynamics on intergenerational analysis. *Journal of Human Resources,* 33: 300–344.

Frankenberg, E., McKee, D. & Thomas, D. (2005). Health consequences of forest fires in Indonesia. *Demography*, 42 (1): 109–129.

Glewwe, P., & Jacoby, H. (1995). An economic analysis of delayed primary school enrollment and childhood malnutrition in a low income country. *Review of Economics and Statistics*, 77 (1): 156-69.

Glewwe, P., Jacoby, H. & King, E. (2001). Early childhood nutrition and academic achievement: a longitudinal analysis. *Journal of Public Economics*, 81: 345-68.

Glewwe, P. & King, E. (2001). The impact of early childhood nutritional status on cognitive development: does the timing of malnutrition matter? *World Bank Economic Review*, 15: 81–114.

Glewwe, P., & Miguel, E. A. (2008). The impact of child health and nutrition on education in less developed countries. In *Handbook of Development Economics*, ed. T.P. Schultz and J. Strauss. Amsterdam: Elsevier.

Hoddinott, J. & Kinsey, B. (2001). Child growth in the time of drought. *Oxford Bulletin of Economics and Statistics*, 63: 409–36.

Jayachandran S. (2009). Air quality and early-life mortality during Indonesia's massive wildfires in 1997. *Journal of Human Resources*, 44 (4): 916–954.

Jim, C.Y. (1999). The forest fires in Indonesia 1997-1998: possible causes and pervasive consequences. *Geography,* 84 (364): 251-60.

Keller, M.C., Garver-Apgar, C.E., Wright, M.J., Martin, N.G., Corley, R.P., et al. (2013). The Genetic Correlation between Height and IQ: Shared Genes or Assortative Mating? *PLoS Genet* 9 (4).

Marioni, R.E., Batty, G.D., Hayward, C., Kerr, S.M., Campbell, A., Hocking, L.J., Porteous, D.J., Visscher, P.M., Deary, I.J. (2014). Common genetic variants explain the majority of the correlation between height and intelligence: the Generation Scotland Study. *Behavior Genetics,* 44: 91-96.

Morduch, J. (1995). Income smoothing and consumption smoothing. *Journal of Economic Perspectives*, 9 (3): 103-114.

Posthuma, D., de Geus, E.J., Neale, M.C., Hulsho Pol, H.E., Baare, W.E.C., Kahn, R.S., Boomsma, D. (2000). Multivariate genetic analysis of brain structure in an extended twin design. *Behavior Genetics*, 30: 311–319.

Preece, M.A. (1996). The genetic contribution to stature. *Hormone Research*, 45: 56-58.

Shrimpton, R., Victora, C., de Onis M., Costa Lima, R., Blössner, M., & Clugston, G. (2001). Worldwide timing of growth faltering: implications for nutritional interventions, *Pediatrics*, 107: 75–81.

Silventoinen, K, Posthuma, D, van Beijsterveldt, T, Bartels, M, Boomsma, D.I. (2006). Genetic contributions to the association between height and intelligence: evidence from Dutch twin data from childhood to middle age. *Genes, Brain and Behavior*, 5: 585–595.

Staiger, D., Stock J.H. (1997). Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.

Stein, Z., Susser, M., Saenger, G., & F. Marolla, F. (1975). *Famine and human development: the Dutch hunger winter of 1944-45*. New York: Oxford University Press.

Sundet, J.M., Tambs, K., Harris, J.R., Magnus, P., Torjussen, T.M. (2005). Resolving the genetic and environmental sources of the correlation between height and intelligence: a study of nearly 2600 Norwegian male twin pairs. *Twin Research and Human Genetics,* 8 (4), 307–311.

Townsend, R. (1995). Consumption insurance: an evaluation of risk-bearing systems in low-income countries. *Journal of Economic Perspectives*, 9, 83–102.

van Dam, P.S., de Winter, C.F., de Vries, R., van der Grond, J., Drent, M.L., Lijffijt, M., Kenemans, L.J., Aleman, A., de Haan, E.H.F., Koppeschaar, H.P.F. (2005) Childhood-onset growth hormone deficiency cognitive function and brain N-acetylaspartate. *Psychoneuroendocrinology* 30: 357–363.

Villar, J., Smeriglio, V., Martorell, R., Brown, C.H. & Klein, R.E. (1984). Heterogeneous growth and mental development of intrauterine growth-retarded infants during the first 3 Years of Life. *Pediatrics* 74 (5): 783-91.

Waber, D., Vuori-Christiansen, L., Ortiz, N., Clement, J., Christiansen, N., Mora, J., Reed, R., & Herrera, M.G. (1981). Nutritional supplementation, maternal education, and cognitive development of infants at risk of malnutrition. *American Journal of Clinical Nutrition* 34: 807-13.

Weedon, N.M. et al. (2008). Genome-wide association analysis identifies 20 loci that influence adult height . *Nature Genetics,* 39: 1245–1250.

World Bank. (2011). *Gender Equality. Kesetaraan Gender*. Policy Brief No. 73031. World Bank, Jakarta.

World Health Organization. (2006). *WHO Multicentre Growth Reference Study Group: WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development.* Geneva, World Health Organization, 2006.

Yamauchi, F. (2008). Early childhood nutrition, schooling, and sibling inequality in a dynamic context: evidence from South Africa. *Economic Development and Cultural Change*, 56 (3): 657–682.

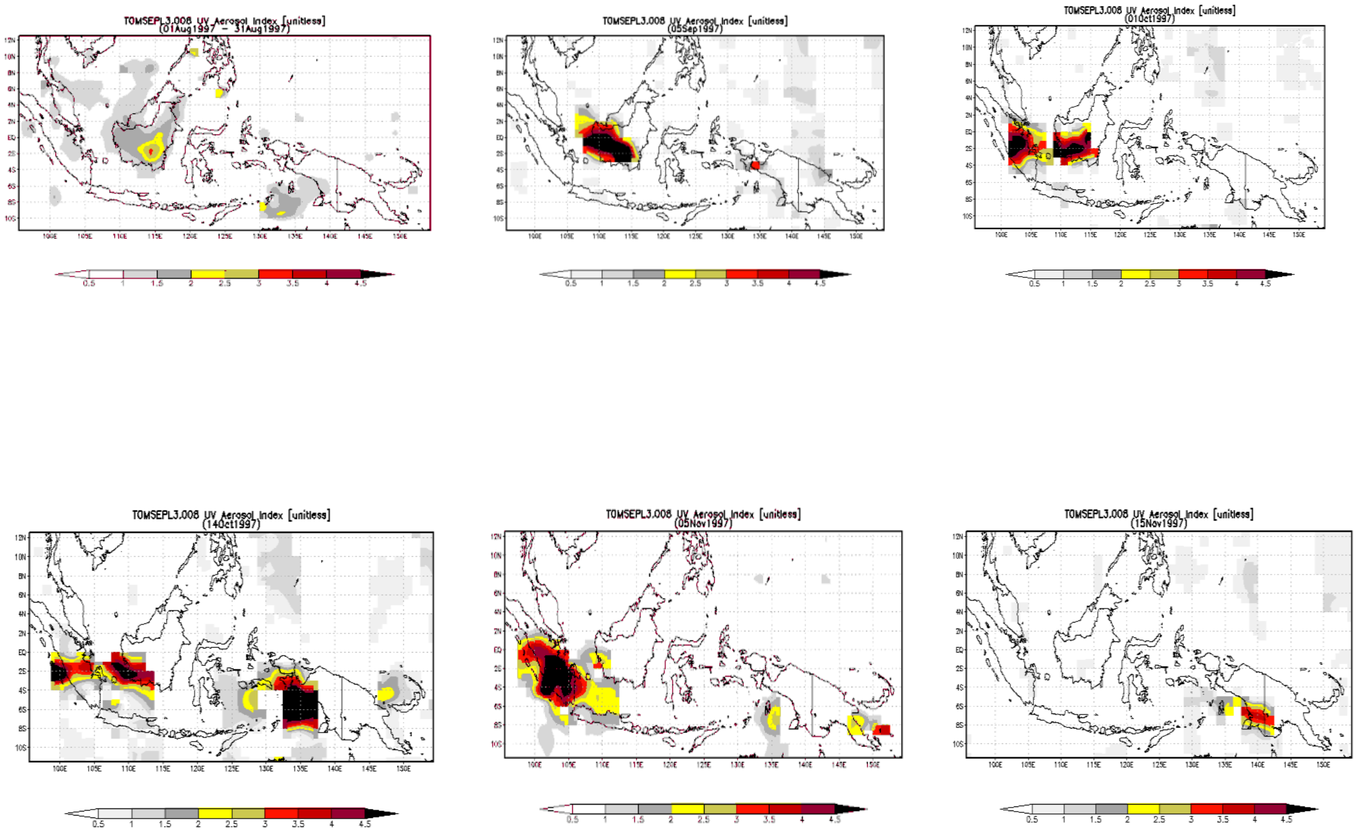
Ziliak, J. & Kneiser, T. (1998). The importance of sample attrition in life cycle labor supply estimation, *Journal of Human Resources*, 33: 507–30.

Appendix

Table A 1 Description of the variables used and descriptive statistics, IFLS (2-3-4)

Variable	*Definition*	*Obs*	*Mean*	*Std.Dev.*	*Min*	*Max*
Gender (male)	1= Child is a boy; 0= Child is a girl	936	.52	.50	0	1
Age (first period)	Child's age (in years)	936	3.67	2.06	0	8
Height-for-Age z-score	Height-for-Age(first period) z-score statistics	936	-1.74	1.34	-5.8	4.2
Weight-for-Age z-score	Weight-for-Age(first period) z-score statistics	936	-1.44	1.19	-5.7	4.1
Age (second period)	Child's age (in years)	936	13.5	2.0	9	17
Age start school	Age (years) at which the child entered school	834	6.3	.65	5	11
Cognitive Test Score	Score obtained for the cognitive test (range from 0 to 1)	783	.76	.18	0	1
Age C.T. score	Estimated age (years) at which the child took the cognitive test	783	10.93	2.07	8	14
Years of schooling	Completed years of education realized in the second period	928	6.7	2.22	0	12
Mother Education	Completed years of education realized by the mother	936	7.08	3.5	0	12
Mother age	Mother's age in years	936	30.13	5.3	15	50
Rural	1= household located in rural areas; 0= household located in urban areas	936	.49	.50	0	1
Fires Shock	1= Child was living in Sumatra or Kalimantan and was aged 12-36 months at the date 05.09.1997	936	.061	.24	0	1

Figure A 1 Location and timing of the Indonesian wildfires



Note: Figures display levels of haze and smoke in August 1997 (monthly average), September, 5th 1997; October, 1st 1997; October, 14th 1997; November, 5th 1997 and November, 15th 1997.Haze is measured using the UV aerosol index. Data Source: NASA Total Ozone Mapping Spectrometer (TOMS)

Table A 2 The effect of forest fires on income and education expenditure. A test for the assumption of exclusion restriction.

	Dep. Var.: Log of Household Per Capita Expenditure..			Dep. Var.: Change in the share of education expenditure..	
	(1) ..in 1997	(2) ..in 2000	(3) ..in 2007	(4) ..between 1997 and 2000	(5) ..between 1997 and 2007
Exposure to Forest Fires	-0.0788	0.116	-0.0772	-0.802	0.829
	(0.103)	(0.0978)	(0.0983)	(0.849)	(1.419)
Constant	11.28***	11.90***	12.85***	2.456***	7.556***
	(0.0226)	(0.0216)	(0.0216)	(0.208)	(0.344)
Observations	934	932	919	930	917
R-squared	0.001	0.002	0.001	0.001	0.000

Note: Household Per capita expenditure is measured in nominal terms. Provincial dummies are included in regressions (1), (2), and (3) in order to control for differences in price levels among provinces. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A 3 Robustness check 1: MFE-IV estimates of height-for-age on years of schooling. Additional covariates added

	(1)	(2)	(3)	(4)	(5)	(6)
Height for Age (z scores)	0.445**	0.643**	0.377**	0.225	0.219	0.226
	(0.197)	(0.309)	(0.177)	(0.162)	(0.162)	(0.161)
Boy	-0.103	-1.123**	-0.117	-0.145	-0.350	-0.00938
	(0.162)	(0.465)	(0.157)	(0.152)	(0.368)	(0.260)
Age	1.169***	0.927***	0.964***	0.760***	0.756***	0.780***
	(0.137)	(0.0311)	(0.0446)	(0.0690)	(0.0681)	(0.0831)
Entered school by the age of 7	0.00363	0.0216	0.131	0.160	0.154	0.151
	(0.246)	(0.264)	(0.263)	(0.250)	(0.253)	(0.259)
Age^2	-0.0355*					
	(0.0199)					
ZHFA*Boy		-0.537**				
		(0.246)				
At school in 1997			-0.357			
			(0.315)			
Birth Order				-0.556**	-0.624**	-0.556**
				(0.232)	(0.243)	(0.231)
Birth Order* Boy					0.107	
					(0.171)	
Age*Boy						-0.0375
						(0.0722)
Observations	885	885	885	819	819	819
Number of moth.id	682	682	682	626	626	626
R^2 (within)	0.798	0.792	0.804	0.823	0.824	0.823
Anderson canon.corr.LR.stat	22.89	18.68	27.72	21.97	21.93	22.03
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cragg-Donald F statistic	23.64	19.08	28.97	22.66	22.50	22.61

Note: The sample consists of children aged 0-8 in 1997 and 9-17 in 2007, with height for age z score in the range -6 to 6 in 1997 and aged 5 to 11 years when started school. ZHFA= Height for age z scores. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A 4 Robustness check 1: MFE-IV estimates of height-for-age on cognitive test score. Additional covariates added

	(1)	(2)	(3)	(4)	(5)	(6)
Height for Age (z scores)	0.0803	0.147	0.105*	0.0809*	0.0865*	0.0802*
	(0.0500)	(0.0943)	(0.0612)	(0.0487)	(0.0483)	(0.0484)
Boy	0.00593	-0.158	0.00674	0.00286	0.128**	-0.0155
	(0.0219)	(0.134)	(0.0249)	(0.0228)	(0.0602)	(0.138)
Age	-0.195**	0.00546	0.00609	0.00925	0.00869	0.00830
	(0.0767)	(0.00632)	(0.00669)	(0.00609)	(0.00597)	(0.0100)
Entered school by the age of 7	-0.0238	-0.0383	-0.0348	-0.0214	-0.0184	-0.0213
	(0.0391)	(0.0463)	(0.0452)	(0.0395)	(0.0386)	(0.0395)
Age^2	0.00902***					
	(0.00349)					
ZHFA*Boy		-0.0969				
		(0.0797)				
At school in 1997			0.0188			
			(0.0349)			
Birth Order				-0.0410**	-0.00276	-0.0408**
				(0.0192)	(0.0213)	(0.0191)
Birth Order* Boy					-0.0710**	
					(0.0297)	
Age*Boy						0.00168
						(0.0124)
Observations	769	769	769	769	769	769
Number of moth.id	602	602	602	602	602	602
R^2 (within)	-0.045	-0.367	-0.270	-0.065	-0.065	-0.062
Anderson canon.corr.LR.stat	20.94	14.49	17.48	22.45	22.24	22.85
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cragg-Donald F statistic	21.64	14.69	17.88	23.31	22.94	23.61

Note: The sample consists of children aged 0-8 in 1997 and 8-14 when they took the cognitive test (either in 2000 or in 2007), with height for age z score in the range -6 to 6 in 1997 and aged 5 to 11 years when started school. ZHFA= Height for age z scores. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A 5 Robustness check 1: MFE-IV estimates of height-for-age on age starting school. Additional covariates added

	(1)	(2)	(4)	(5)	(6)
Height for Age (z scores)	-0.198	-0.311*	-0.212*	-0.207	-0.212*
	(0.138)	(0.180)	(0.128)	(0.128)	(0.128)
Boy	0.0856	0.602**	0.0912	0.250	0.0626
	(0.0796)	(0.272)	(0.0827)	(0.204)	(0.139)
Age	-0.118	0.00287	-0.0418	-0.0368	-0.0460
	(0.0917)	(0.0164)	(0.0460)	(0.0452)	(0.0512)
Age^2	0.0188				
	(0.0135)				
ZHFA*Boy		0.275*			
		(0.148)			
Birth Order			-0.139	-0.0810	-0.138
			(0.133)	(0.139)	(0.132)
Birth Order*Boy				-0.0808	
				(0.0901)	
Age*Boy					0.00857
					(0.0352)
Observations	834	834	772	772	772
Number of moth.id	652	652	600	600	600
R^2 (within)	-0.019	0.023	-0.041	-0.031	-0.041
Anderson canon.corr.LR.stat	19.94	15.86	20.25	20.08	20.33
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cragg-Donald F statistic	20.61	16.21	20.99	20.68	20.95

Note: The sample consists of children that in 1997 were aged 0-8 and had height for age z score in the range -6 to 6 in 1997 and were subsequently aged 5 to 11 years when started school. The difference between age started school and age at which height was measured is larger or equal than zero. ZHFA= Height for age z scores.
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A 6 Robustness check 2: MFE-IV estimates of weight-for-age in baseline and alternative specifications

		Educational Outcome:					
		Years of schooling		Cognitive Test Score		Age starting school	
(0)	Baseline	0.525*	(0.279)	0.065	(0.060)	-0.189	(0.138)
(1)	Entered school by the age of 7 included	0.452**	(0.223)	0.103*	(0.058)	-	-
(2)	Age fixed effects included	0.266	(0.224)	0.097*	(0.056)	-0.384	(0.254)
(3)	Age squared included	0.509**	(0.232)	0.073	(0.047)	-0.224	(0.156)
(4)	ZHFA* boy included	0.967**	(0.485)	0.171*	(0.102)	-0.418*	(0.224)
(5)	At school in 1997 included	0.448**	(0.220)	0.101*	(0.058)	-	-
(6)	Birth order included	0.281	(0.220)	0.069	(0.044)	-0.257	(0.161)
(7)	Birth order*Boy included	0.269	(0.213)	0.0703	(0.043)	-0.247	(0.155)
(8)	Birth order and age* Boy included	0.284	(0.219)	0.0665	(0.044)	-0.257	(0.160)

Note: ZHFA= Height for age z scores. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A 7 Robustness check 3: MFE-IV estimates of height-for-age instrumented with exposure to forest fires during the 12-24 first months of life in baseline and alternative specifications

		Educational Outcome:					
		Years of schooling		Cognitive Test Score		Age starting school	
(0)	Baseline	0.661**	(0.289)	0.152	(0.105)	-0.053	(0.152)
(1)	Entered school by the age of 7 included	0.583***	(0.222)	0.192*	(0.103)	-	-
(2)	Age fixed effects included	0.462*	(0.265)	0.160*	(0.084)	-0.176	(0.171)
(3)	Age squared included	0.551***	(0.202)	0.147*	(0.082)	-0.046	(0.141)
(4)	ZHFA* boy included	1.179**	(0.496)	0.364*	(0.221)	-0.190	(0.318)
(5)	At school in 1997 included	0.534***	(0.200)	0.190*	(0.103)	-	-
(6)	Birth order included	0.368*	(0.205)	0.141*	(0.079)	-0.130	(0.169)
(7)	Birth order*Boy included	0.376*	(0.203)	0.130*	(0.072)	-0.137	(0.164)
(8)	Birth order and age* Boy included	0.369*	(0.204)	0.141*	(0.078)	-0.130	(0.169)

Note: ZHFA= Height for age z scores. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1