Understanding Children's Work
An Inter-Agency Research Cooperation Project

Understanding Children's Work Project Working Paper Series, January 2004

Measuring the vulnerability of children in developing countries: an application to Guatemala

F. Mealli
S. Pudney
F. C. Rosati

January 2004

Measuring the vulnerability of children in developing countries: an application to Guatemala

F. Mealli[*]
S. Pudney[**]
F. C. Rosati[***]

Working Paper
January 2004

Understanding Children's Work (UCW) Project
University of Rome "Tor Vergata"
Faculty of Economics
V. Columbia 2
00133 Rome Tor Vergata

Tel: +39 06.7259.5618
Fax: +39 06.2020.687
Email: info@ucw-project.org

As part of broader efforts toward durable solutions to child labor, the International Labour Organization (ILO), the United Nations Children's Fund (UNICEF), and the World Bank initiated the interagency Understanding Children's Work (UCW) project in December 2000. The project is guided by the Oslo Agenda for Action, which laid out the priorities for the international community in the fight against child labor. Through a variety of data collection, research, and assessment activities, the UCW project is broadly directed toward improving understanding of child labor, its causes and effects, how it can be measured, and effective policies for addressing it. For further information, see the project website at www.ucw-project.org.

This paper is part of the research carried out within UCW (Understanding Children's Work), a joint ILO, World Bank and UNICEF project. The views expressed here are those of the authors' and should not be attributed to the ILO, the World Bank, UNICEF or any of these agencies' member countries.

[*] University of Florence
[**] University of Leicester
[***] UCW project and University of Rome Tor Vergata

Measuring the vulnerability of children in developing countries: an application to Guatemala

Working Paper
January 2004

ABSTRACT

Anti-poverty policy in developing countries has focused mainly on the measurement and location of poverty and the targeting of policy towards those who are currently poor. Recently, the research effort has been extended to cover those judged to be not poor at present but vulnerable to poverty in the future. We concentrate on two aspects: inadequate education and child labor, which are closely associated with chronic poverty. We develop and apply new methods for the measurement and empirical analysis of vulnerability to future premature school leaving and/or onset of child labor. Guatemalan survey data are used for the illustrative application.

Measuring the vulnerability of children in developing countries: an application to Guatemala

Working Paper
January 2004

CONTENTS

1. Introduction 1
2. The concept of vulnerability 2
3. A model of the risks of school drop-out and child labor 6
   - 3.1 Hazard rates 6
   - 3.2 Data requirements 8
   - 3.3 Estimation of the model 8
   - 3.4 Derivation of $f(S, W \mid X)$ 9
   - 3.5 The likelihood 9
   - 3.6 Functional forms 10
4. Defining indicators of vulnerability 11
   - 4.1 Individual-level indicators 11
   - 4.2 Risk measures 11
   - 4.3 Welfare weights 12
   - 4.4 Measures of high vulnerability 13
   - 4.5 Identifying the vulnerable 13
5. The data set 16
   - 5.1 The Guatemalan Living Standards Measurement Survey 2000 16
   - 5.2 Children's work in Guatemala 17
6. Estimates of the transition model for Guatemala 19
   - 6.1 The time profile of hazard rates 19
   - 6.2 The impact of individual and household characteristics and of external shocks 20
7. Vulnerability in Guatemala 22
   - 7.1 The distribution of vulnerability 22
   - 7.2 Aggregate vulnerability 24
   - 7.3 The role of credit constraints and shocks 29
8. Policy targeting 31

9. Conclusions 38
References 39
Appendix 1: Data definitions and description 40
Appendix 2: Parameter estimates .... **Errore. Il segnalibro non è definito.**

1. INTRODUCTION

1. Anti-poverty policy in low income countries has, for obvious reasons, focused mainly on the measurement of the extent of poverty, the identification of the poor and the design of policies to alleviate their plight. Recently, there has been an effort to extend the research effort to identify those who are judged to be not poor at present, but vulnerable to poverty in the future (Morduch, 1994; Jalan and Ravallion, 1999; Pritchett *et. al.*, 2000; Chaudhuri *et. al.*, 2001; Tesliuc and Lindert, 2002). If this can be done, then it is possible to design policies intended to protect these vulnerable people from future adverse outcomes. There is a helpful parallel with health care. Conventional treatment programmes are designed to provide care for those who are currently suffering from disease. In contrast, preventative care is forward-looking and is designed to protect those who are at high risk of contracting a disease in the future. Effective health care policy requires both treatment and preventative components. Increasing attention has been devoted to policies aimed at reducing the vulnerability of households and to promote risk reduction and risk coping strategies. In particular the World Bank has developed its Social Risk Management strategy (Holzmann and Jorgensen, 2002) to promote this approach to policy design.

2. An important distinction here is that between vulnerability to transient poverty and vulnerability to long-term, sustained disadvantage (Jalan and Ravallion, 2001). There is a strong case for focusing the analysis of vulnerability and the consequent policy design on factors that are associated with the transmission of chronic poverty between generations. Inadequate education and child labor are closely associated with chronic poverty and have received a great deal of attention from researchers and policy-makers (for a recent review see Cigno *et.al.*, 2002).[2] In this paper, we concentrate on these two aspects and, as a complement to 'treatment'-style studies of the extent and distribution of inadequate education and child labor, we develop and apply new methods for the measurement and empirical analysis of vulnerability to future premature school leaving and/or onset of child labor.

3. In the next section we discuss alternative concepts of vulnerability. Sections 3 and 4 develop the statistical transition model that underlies our approach and discuss the construction of empirical indicators of vulnerability. Sections 5-8 describe the Guatemalan survey data used for the illustrative application and set out the results. Section 9 concludes.

[2] See also the UCW web site www.ucw-project.org.

2. THE CONCEPT OF VULNERABILITY

4. The concept of vulnerability expresses the potential realisation of an adverse outcome. In terms of welfare analysis this concept can be expressed in terms of the possibility that a household or individual will experience a reduction in well being. Vulnerability is also associated with exposure to risks that might lead to the realisation of such an adverse outcome.

5. The analysis of vulnerability has been mainly focused on poverty, even if the need to extend the focus also to other dimension of welfare has been recognized and some tentative analysis has been carried out. The survey of Dercon (2000) contains a review of most of the approaches that have been followed to analyse and measure vulnerability to poverty. While it has been widely recognized that vulnerability is an intrinsically dynamic problem, limitations in data availability and statistical methodology have forced researchers to adopt largely static approaches. The main focus has been to identify some source of risk, typically relating to income or employment, and to assess the effect of such factors on the probability of becoming poor. Few of these studies are able to exploit real panel data and they have largely had to rely on cross section information to generate indicators of possible risks.

6. In this study we extend the previous literature on the subject of vulnerability in two directions: vulnerability to child labor and to the premature cessation of education leading to a lack of accumulation of children's human capital. We propose a methodology that relies on retrospective information, available in many cross sectional data sets to give a time profile of individual behaviour. This allows us to measure the hazard that a given adverse event occurs, without the requirement for panel data. Nevertheless, the availability of a full longitudinal or panel dataset would allow the analysis to be greatly enriched.

7. Children can spend their time at school, at work, doing both or doing nothing. We assume that lack of education is harmful to the child in terms of different aspects of future welfare, mainly in terms of income and of health (as education is known to be a very important influence on subsequent health status). Working can also directly affect future health because of the risks and physical stresses of the activity carried out. At this stage we also assume that any short run benefit associated with child labor (higher income, better nutrition etc.) is outweighed by the long term negative effects due to the lack of capital accumulation and to the long-term physical consequences of work. Vulnerability to child labor and to underinvestment in human capital accumulation can be interpreted also as a proxy for vulnerability to poverty of future (adult) generations. Lower health status, lower returns to work, employment in less protected sectors *etc*. have all been shown to be associated with school drop out and/or early entry in the labor market.

8. It is important, in discussing the slightly nebulous concept of vulnerability, to be very clear about the use of terminology. One of the principal aims of this paper is to set the concept of vulnerability on a firmer foundation by defining it clearly in relation to the concept of a hazard rate, which has a precise technical meaning, outlined in the next section. We use the terms risk and hazard interchangeably to refer to the current probability of the onset of a particular adverse event (in our case, school leaving or entry into the labour force). Measures of vulnerability can then be constructed in various ways from knowledge of the hazard rates for any given class of specific adverse outcomes.

9. The Oxford English Dictionary defines the word "vulnerable" as "liable to be damaged" or "not protected against attack". These are two distinct but related ideas.

First, note that both relate to possible events in an uncertain future. Thus unpredictability and the passage of time are inherent in both. Secondly, note that the second definition is wider than the first. One can be unprotected against an attack (for example, a lightning strike) but still have little liability to damage (since lightning strikes are rare). Thirdly, note that, since there is unpredictability, observed shocks and outcomes may differ randomly from their expectations during any given observation period: in other words, *liability* to damage is not the same as *occurrence* of damage. Thus the situation is as depicted in Figure 1, where solid arrows denote the underlying structure generating outcomes and dashed arrows denote the random generation of observed outcomes.

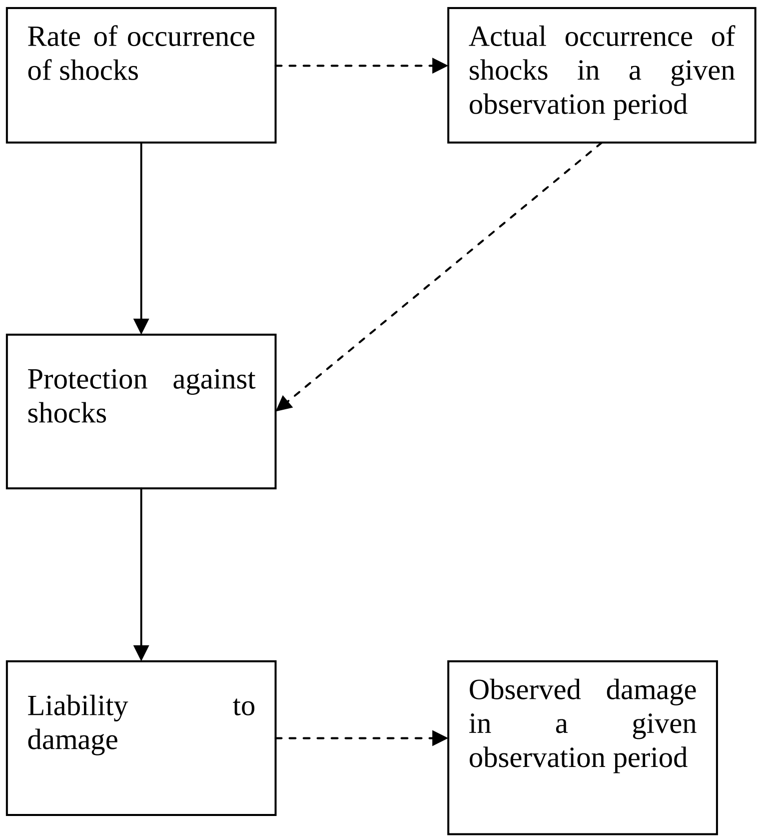


Figure 1. - The logical structure of vulnerability and actual outcomes

10. There are several decisions to be made before the notion of vulnerability becomes operational.

11. *Structural or reduced-form analysis?*. Figure 1 decomposes the vulnerability "process" into two stages: the rate of occurrence of shocks and the factors giving protection from shocks. One approach is to analyse these two stages separately and then combine the results to give an overall picture of vulnerability in the sense of liability to damage. This has some advantages: it helps policy-makers by giving a detailed view of the scope for policies (such as flood defences) that might reduce the number and severity of shocks, compared with policies (such as famine relief) that mitigate the effects of shocks. Arguably, it may also give a basis for more robust simulations of the effect of changing circumstances on the pattern of vulnerability. However, there are also some disadvantages of the structural approach. Firstly, it may be difficult to identify the occurrence of shocks and isolate their consequences those of from other events: Jalan and Ravallion (1999) and Tesliuc and Lindert (2002) discuss the issues here. A second difficulty is that there is no natural definition of a "shock". Indeed, future poverty is often associated with persistent disadvantage, manifested in low educational attainment, limited skills, poor physical state and low non-human capital. For a particular disadvantaged family, there may be nothing exceptional about this state - it need not be the result of natural disaster, poor harvest, etc, but simply the normal cycle of transmission of poverty through the generations. In terms of their future prospects, such people are highly vulnerable in the sense of liability to damage, but without any set of specific "shocks" being involved.

12. *The outcome measure*. Damage can only be judged in relation to some specific outcome measure. This requires the definition of a welfare indicator - do we use income, consumption or some other observable entity to capture the impact of adverse events? Consumption expenditure is the preferred indicator in much of the literature on vulnerability (Chaudhuri *et. al.*, 2001; Tesliuc and Lindert, 2002; Ligon and Schechter, 2003) but income, wealth and other variables are also possibilities.

13. *The time horizon*. In practice, we must choose the horizon over which the outcome measure is defined. Temporary dips in income and consumption may be judged unimportant in comparison to the long-term factors that underpin the inter-generational transmission of poverty (see Jalan and Ravallion, 2000). There is generally a trade-off between the choice of a welfare measure and the time horizon: even if we are concerned with long-term welfare defined in consumption terms, the non-availability of suitable longitudinal data may lead us to use non-consumption variables that are better indicators of long-term welfare than available consumption data. This is the approach used in this study. Our view is that much of the literature has a weakness here. Although the concept of vulnerability is inherently intertemporal, several influential analyses have been forced to use methods that amount to statistical modelling of consumption conditional on other contemporaneously-observed covariates. Examples of this include Chaudhuri *et. al.* (2001) whose analysis amounts to a parametric heteroskedastic probit model of the probability of being below a consumption threshold conditional on current observed characteristics; and Ligon and Schechter (2003), who make a decomposition of the the difference between a threshold 'social utility' level and the expected value of a concave transformation of consumption, using the law of iterated expectations in relation to a set of contemporaneously-observed covariates. We would argue that both of these analyses are essentially conditional descriptions of the distribution of consumption at a point in time and do not capture the intertemporal nature of the concept of vulnerability, since there is no direct analysis of the probability of a future

adverse event conditional on current circumstances. To do this would require an analysis of the dynamics of consumption, not its cross-section distribution.[3]

14. Measures of vulnerability based on the contemporaneous association between demographic characteristics and a welfare indicator rest on the implicit assumption that the probability of a movement into poverty is directly related to current distance from poverty. Although this is a reasonable assumption for many families, it is not hard to think of counter-examples involving high mean income, high risk households contrasted with households having lower mean income but lower risk. The difficulty with this type of measure implemented on cross-section data is that long-term temporal variations in income or consumption cannot be observed, so the relationship between the probability of furture movement into poverty and current circumstances cannot be established reliably. Our analysis based on education and child labor is special, since the retrospective information in our cross-section survey does permit the modelling of actual transitions conditional on past circumstances.

[3]Despite the fact that Ligon and Schechter (2003) use monthly panel data, their model of consumption (equation 4 in their paper) is static. Moreover, the unexplained variance of consumption that they use represents the month-to-month volatility in consumption, which may reflect individual exposure to predictable seasonality as much as the occurrence of events with persistent adverse consequences.

3. A MODEL OF THE RISKS OF SCHOOL DROP-OUT AND CHILD LABOR

3.1 Hazard rates

We are working with vulnerability interpreted as the current risk of the adverse events of school drop-out and child labor. The key concept underlying this approach is the hazard function (see Lancaster, 1990). In the analysis of mortality data, the hazard function h(t) is usually defined in continuous time and measures the instantaneous risk of dying, in that h(t)dt is the probability of dying in the next small time interval dt given survival to time t. Here we work with events, such as school dropout, that are usually observed in discrete time; for example we know the numbers of years of school attendance, but are not usually able to know the exact date when school attendance stopped. So, in general terms, the hazard rate for a given time period (the year) is defined as the probability of an event occurring within that period, conditional it not having occurred previously. We measure time as the age of the child in years, so that the hazard function is the relationship of the hazard rate with age. Thus, for example, at age 11, the hazard rate for school drop-out is the probability of drop-out at t = 11 (i.e. during the 12th year of life) for those who are still enrolled in school on their 11th birthday. To be more precise we consider that a child can be in any of the following mutually exclusive states at any time: studying only, working only, working and studying, doing neither. Given this initial state, we work with the hazards of leaving school and/or entering the labor force. The variables involved in the analysis, the hazard functions and the distribution of outcomes are defined in Table 1, which sets out our notation. Consider, for example, a child studying only: she faces the hazard P1(A | X) of leaving school without starting work (i.e. becoming idle), the risk P2(A | X) of starting work without leaving school, and the hazard P3(A | X) of leaving school to start working. Analogous hazards, Q1(A | X) and Q2(A | X), can be defined for the other possible initial states. Hazard functions can also be defined in continuous time but the available data is rarely sufficiently precise in terms of dating for this to be a fruitful generalisation. Note also that we treat children not entering school at all, as if they drop out at age 7; conversely, late school entrants are treated as entrants at age 7. As the number of children entering school after the age of 7 is small, the results are not sensitive to this convention.[4, 5]

[4] Note that $Q_1(A|X)$ and $Q_2(A|X)$ are not defined for age 7, because they are hazards conditioned on having left school or having started to work before age 7.

[5] We are assuming here that the covariates *X* are not time-varying. This assumption is justified for factors such as parents' education but is not tenable for variables (such as income, household structure and the occurrence of "shocks") which are directly observed only for the year in which the survey is conducted. In this paper we treat all covariates as time-invariant; work is in progress on a variant of the model that allows for time-variation and partial observability.

Table 1. - Notation

X	Vector of measured characteristics describing the child and his/her family, social and economic background
S	Age of leaving school
W	Age of starting working
T	Age at which basic schooling finishes and below which work is defined as child labor
$f(S, W \mid X)$	Discrete distribution of S, W conditional on X
$P_1(A \mid X)$	School dropout hazard function for a non-working schoolchild of age A = $\Pr(S{=}A, W{>}A \mid S{\geq}A, W{\geq}A, X)$
$P_2(A \mid X)$	Work hazard function for a non-working schoolchild of age A = $\Pr(S{>}A, W{=}A \mid S{\geq}A, W{\geq}A, X)$
$P_3(A \mid X)$	Simultaneous school dropout and work hazard function for a non-working schoolchild of age A = $\Pr(S{=}A, W{=}A \mid S{\geq}A, W{\geq}A, X)$
$Q_1(A \mid X)$	School dropout hazard at age A for a working schoolchild =$\Pr(S{=}A \mid S{\geq}A, W{<}A, X)$
$Q_2(A \mid X)$	Work hazard at age A for a child who is not at school =$\Pr(W{=}A \mid S{<}A, W{\geq}A, X)$

For the Guatemalan survey analysed below, the two following questions are used to establish the ages of school-leaving.

Q1 *"Did <child> drop out or is s/he not attending the school where s/he was registered?"* (Allowable responses: is attending / definitely dropped out)

Q2 *"What is or was the highest education level achieved?"* (Allowable responses: none / knows how to read and write / preparatory / complete primary / incomplete primary / secondary / higher education / don't know)

For children still at school, we know that school drop-out did not occur prior to the current recorded age. For those who are not attending school, age of drop-out is taken to lie between the age at which the attained grade is normally reached and the child's current age.

The age of onset of child labor is taken directly from the response to the following question:

Q3 *"How old were you when you had your first paying job or your first job helping without pay on the family farm or business?"*

15. Our model does not allow for reverse transitions and repeat spells. This can be rationalized in either of two ways. One is to assume that school drop-out and commencement of child labor are once-and-for-all events, so that a school-leaver never returns to education and a child laborer never subsequently quits the labor force. Alternatively (and preferably) we can regard the model as dealing with entry into the labor force rather than employment *per se* and with permanent drop-out from school rather than temporary absence (see the allowable responses to *Q1*). Thus, for example, the model would ignore the educational impact of a brief absence from school to help with seasonal farm work, rather than treating it as equivalent to leaving school permanently. The interpretation of work as child labor rests on its interpretation by respondent and interviewer as substantial enough to merit the term "job" (see Q3). This seems reasonable in the context of a poor rural family.

16. With sufficiently detailed longitudinal data it would be possible in principle to estimate transition models that allow for the possibility of spells of schooling interrupted by occasional periods of absence from school and/or child labor.

However, this degree of detail is absent from most available surveys, so we do not allow for more complex patterns of schooling and working. Table 2 synthesizes the transition probabilities implied by our model.

Table 2. - Transition probabilities for movements between states

		Destination state			
		School only	hool & work	Work only	Neither
Origin state	School only	1 - $P_1(A \mid X)$ - $P_2(A \mid X)$ - $P_3(A \mid X)$	$P_2(A \mid X)$	$P_3(A \mid X)$	$P_1(A \mid X)$
	School and work	-		$Q_1(A \mid X)$	-
	Work only	-	-	-	-
	Neither	-	-	$Q_2(A \mid X)$	

If we are able to estimate the five hazard functions $P_1(A \mid X)$, $P_2(A \mid X)$, $P_3(A \mid X)$, $Q_1(A \mid X)$ and $Q_2(A \mid X)$, they can be used as the basis of vulnerability indicators reflecting the risk for this child of dropping out of school and/or commencing work at age A. By summing the relevant probabilities over all children who are at school and not working, we can estimate, for example, the number of children that will start working next year. Similarly, for a child of age A-1 at school and already working, the estimated $Q_1(A \mid X)$ reflects the risk of school drop-out. For a child of age A-1 already out of schooling and not yet in work, the estimated hazard $Q_2(A \mid X)$ reflects the risk of starting work. The risk of entering the "idle" status is $P_1(A \mid X)$ for a non-working schoolchild.

3.2 Data requirements

17. Estimation of hazard rates require adequate observation of the processes by which children leave school and become workers. The minimum requirement for reliable estimation is that we can observe the following:

(i) Whether or not the child is still at school at the time of interview
(ii) The stage of schooling completed, for those not at school or the date of leaving school.
(iii) Whether or not the child has already begun working at the time of interview

18. Direct observation of the age at which child labor began is also a great advantage and is available in the Guatemalan survey data (survey question Q3, above).

3.3 Estimation of the model

19. Survey data tell us about the outcomes of the transition processes rather than the processes themselves. We can observe any given child in one of the following mutually exclusive states: (*i*) studying only, (*ii*) working only, (*iii*) working and studying, (*iv*) doing neither. For children in states (*ii*) and (*iv*) we will also observe the age at which schooling finished[6] and for children in states (*ii*) and (*iii*) we observe the age at which child labor began. So, using the notation set out in Table 2, the estimation process fits the following probabilities to the observed outcomes:

[6] Or a range of possible ages at which schooling finished.

(i) Pr(studying only | A, X) = $f(S>A, W>A \mid A, X)$
(ii) Pr(left school at age S, work began at age W | A, X) = $f(S, W \mid A, X)$
(iii) Pr(studying, work began at age W | A, X) = $f(S>A, W \mid A, X)$
(iv) Pr(left school at age S, not working | A, X) = Pr(S, $W>A \mid A, X$)

20. These probabilities are derived from the hazard functions P_1 ... P_3 , Q_1 and Q_2, but it is the hazard funstions, rather than the directly observable probabilities (i)-(iv), which we are primarily interested in. This is because the hazard rates are the natural indicators of vulnerability at a point in time. To infer P_1 ... P_3 , Q_1 and Q_2 from the data we proceed in two stages: firstly derive the distribution $f(S, W \mid A, X)$ from the functions P_1 ... P_3 , Q_1 and Q_2; and secondly construct a likelihood function for the survey observations on school and work status, using $f(S, W \mid X)$, in the case of the minimal set of information. The likelihood can be easily specialized for cases of more informative data.

3.4 Derivation of $f(S, W \mid X)$

21. Suppressing the conditioning on X and A for notational simplicity, the joint distribution of outcomes (S, W) is as follows:

$$f(S,W)=\begin{cases} \left[\prod_{t=1}^{S-1} P_0(t)\right] P_3(S) & S=W \\ \left[\prod_{t=1}^{S-1} P_0(t)\right] P_1(S)\left[\prod_{t=S+1}^{W-1}(1-Q_2(t))\right] Q_2(W) & S<W \\ \left[\prod_{t=1}^{W-1} P_0(t)\right] P_2(W)\left[\prod_{t=W+1}^{S-1}(1-Q_1(t))\right] Q_1(S) & W<S \end{cases} \qquad (1)$$

where $P_0(t) = 1-P_1(t)-P_2(t)-P_3(t)$. Note that $f(S,W)$ is the probability of leaving school at age S and starting to work at age W: so to explain (1), if $S < W$ it means that for the first S-1 years the child attended school and did not start to work ($\prod_{t=1}^{S-1} P_0(t)$), when he turned S he dropped out of school ($P_1(S)$), between S+1 and W-1 he did not start to work ($\prod_{t=S+1}^{W-1}\left(1-Q_2(t)\right)$), and finally at age W he entered the labor market ($Q_2(W)$).

3.5 The likelihood

22. There are four types of observation that we may encounter: a child still at school and not working; a child still at school and working; a child who left school sometime in the interval $[S_1,S_2]$ and is not working; and a child who left school sometime in $[S_1,S_2]$ and is working. Denote the corresponding sets of observations T_1 ... T_4 and let a_i be the observed age of child i. Then the log likelihood is:

$$L=\sum_{i\in T_1}\ln\left[\sum_{s=a_i+1}^{T}\sum_{w=a_i+1}^{T}f(s,w\mid x_i)\right]+\sum_{i\in T_2}\ln\left[\sum_{s=a_i+1}^{T}\sum_{w=1}^{a_i}f(s,w\mid x_i)\right]$$
$$+\sum_{i\in T_3}\ln\left[\sum_{s=S_{1i}+1}^{S_{2i}}\sum_{w=a_i+1}^{T}f(s,w\mid x_i)\right]+\sum_{i\in T_4}\ln\left[\sum_{s=S_{1i}+1}^{S_{2i}}\sum_{w=1}^{a_i}f(s,w\mid x_i)\right] \quad (2)$$

23. Maximising this function numerically then gives ML estimates of the parameters embedded in the hazard functions P_1 ... P_3, Q_1 and Q_2.[7] To explain (2), consider for example a child belonging to the T_3 group, of age a_i, who left school sometime in the interval $[S_1,S_2]$ and is not working: for him we only know that S is between S_{1i} and S_{2i} and that W is greater than a_i. This explains the summations in (2) required to form the log likelihood.

3.6 Functional forms

For this approach to be made operational, we need to specify functional forms for the basic components of the model: the five hazard functions $P_1(A|X)$ ... $P_3(A|X)$, $Q_1(A|X)$, $Q_2(A|X)$. We specify the $P_j(a \mid X)$ to have multinomial logit form:

$$P_j(A\mid X)=\frac{\exp\{X\alpha_j+D(A)\beta_j\}}{1+\sum_{k=1}^{3}\exp\{X\alpha_k+D(A)\beta_k\}} \qquad j = 1...3 \qquad (3)$$

Q_1 and Q_2 are specified as binary logits:

$$Q_j(A\mid X)=\frac{\exp\{X\gamma_j+D(A)\delta_j\}}{1+\exp\{X\gamma_j+D(A)\delta_j\}} \qquad j = 1,2 \qquad (4)$$

The vector of constructed variables $D(A)$ represents the time profile of the hazard rates. We explore two alternative forms for the time profiles:

Quadratic: $D(A) = [1, A, A^2]$

Semi-parametric: $D(A) = [1(A=7), 1(A=8), ..., 1(A=T-1)]$

where $1(A=t)$ is the indicator function, equal to 1 if $A = t$ and 0 otherwise and where we have assumed that school begins at age 7. Consequently, the indicator function $1(A=7)$ is dropped in $Q_1(A \mid X)$ and in $Q_2(A \mid X)$. The probabilities $f(S, W \mid X)$ are constructed to sum to 1 over $S, W = 1 ... T$, by setting the terminal values: $P_1(T) = P_2(T) = 0$ and $P_3(T) = Q_1(T) = Q_2(T) = 1$.

---

7 A Gauss 5.0 program was developed to maximize the log-likelihood numerically.

4. DEFINING INDICATORS OF VULNERABILITY

24. The school-work transition model is of interest in itself but we concentrate mainly on a range of indicators of vulnerability that can be constructed from the estimated models. There are three important issues here. Firstly, what measures should be used to capture the level of vulnerability for any given individual? Secondly, how can these individual-level measures be aggregated to give a measure of the vulnerability of a specific group of individuals, or of the child population in one region or country compared to another? Thirdly, given the vulnerability measure for each individual within a population, how can we identify the personal characteristics most closely associated with vulnerability and thus target policy most effectively?

4.1 Individual-level indicators

25. The risks faced by a child depend on his or her current position. A child who is attending school and not in the labor force faces potentially adverse outcomes: school drop-out; onset of child labor; and both simultaneously. For a child aged A with characteristics X, these risks are measured by the relevant transition probabilities $P_1(A|X)$ ... $P_3(A|X)$. A child who is at school and already working faces the risk of school drop-out, measured by $Q_1(A \mid X)$. A child who is not at school and not working faces the single risk of onset of child labor, measured by $Q_2(A|X)$. Once the model has been estimated, these conditional risk measures can be evaluated for each sampled individual and the distributions across individuals summarized in various ways. However, it will be useful to combine or modify these risk measures to produce a more clearly focused measure of risk for each individual. Two developments will be useful: condensing the 3 risk measures in the case of non-working school children; and constructing modified measures to identify cases of high risk.

4.2 Risk measures

26. The estimated model involves five hazard functions, which we reduce to three vulnerability indicators. Define VS and VW respectively as the age-specific probabilities of exit from school and of entry into the labor force. These will depend on the child's current educational and labor force status as follows:

Non-working schoolchild: $V_S(A|X) = P_1(A|X) + P_3(A|X)$

$V_W(A|X) = P_2(A|X) + P_3(A|X)$

Working schoolchild: $V_S(A|X) = Q_1(A|X)$

$V_W(A|X) = 0$

Non-working, non-school child: $V_S(A|X) = 0$

$V_W(A|X) = Q_2(A|X)$

27. If we aggregate V_S and V_W, we get an unbiased estimate of the number of children who will exit from school or entry into the labor force during the current

year. Let Y_i be a binary variable assuming value 1 if individual *i* will drop out of school during the current year (i.e. at age a_i) and 0 otherwise; then the expected number of children exiting from school can be estimated as follows:

$$\hat{E}\left(\sum_{i=i}^{N} Y_i\right) = \sum_{i=1}^{n} V_S\left(a_i | x_i\right)\omega_i$$

where *n* is the sample size, *N* is the population size and ω_i is the survey weight for individual *i*. Note that the sum of the estimated individual specific school-leaving risks can be decomposed as follows:

$$\hat{E}\left(\sum_{i=i}^{N} Y_i\right) = \sum_{i=1}^{n} V_S\left(a_i | x_i\right)\omega_i = \sum_{i:V_S\left(a_i | x_i\right)>k} V_S\left(a_i | x_i\right)\omega_i + \sum_{i:V_S\left(a_i | x_i\right)\leq k} V_S\left(a_i | x_i\right)\omega_i$$

so that both the children with high vulnerability (i.e with V_s greater than *k*) and the children with low vulnerability contribute to form the group of children who will leave school. This opens a question whether it is better to target policies towards the highly vulnerable, or have an overall reduction of the vulnerability level. If the policy objective is to reduce $\sum_{i=i}^{N} Y_i$, it is not obvious a priori which of the two targets if preferable.

4.3 Welfare weights

28.As an alternative, to separate analyses for schooling and labor vulnerability, we can construct a single vulnerability indicator capturing the risks of both kinds of adverse transition. This requires a suitable weighting scheme that makes assumptions about the relative seriousness of school drop-out and child labor, i.e., attribute a welfare loss to the different states a child can find herself in. If, for example, we believe that the welfare loss associated with school drop-out at age *A* is $k_S(A)$ and that associated with entry into the labor force at age *A* is $k_W(A)$, then we can construct the following composite vulnerability indicator $V(A \mid X)$ for a child aged *A* with characteristics *X*:

$$V(A|X) = k_S(A)\, V_S(A|X) + k_W(A)\, V_W(A|X) \qquad (5)$$

29.Note that this approach can be generalized considerably. The welfare costs $k_S(A)$ and $k_W(A)$ can be made dependent on the initial status of the child and on other characteristics besides age. However, there is a virtue in simplicity and we suggest the following welfare weights:

$$k_S(A) = \theta(1+g)^{T-A} \qquad (6)$$
$$k_w(A) = (1-\theta)(1+g)^{T-A} \qquad (7)$$

30. This reflects two measurement conventions: that the welfare losses of school drop-out and child labor are in the proportions $\theta : (1-\theta)$ and that losses occurring at earlier ages increase geometrically at the rate g.[8] While the comparison of welfare losses arising from dropping out of school and from working do reflect the need for a synthetic indicator, but are open to questions, this is not the case for age. There is substantial evidence that individual welfare loss increases the earlier the child left school and/or started to work. Neglecting this aspect might have important implication in terms of identification of group at risk, as we will show later.

31. It is only necessary to assume a pair of values (θ , g) to make the measure operational. In practice, one would use sets of alternative values, to explore sensitivity.

4.4 Measures of high vulnerability

32. Assume we have chosen an individual-level measure of vulnerability $V(A \mid X)$, using either one of the approaches discussed in the previous section. We then identify a highly-vulnerable individual by specifying a threshold v and a corresponding binary indicator $d_v(A \mid X)$:

$$d_v = 1(V(A \mid X) > v) \qquad (8)$$

where 1(.) is the indicator function. The relationship between d_v and A, X can be explored by means of tabulation and other statistical techniques. This is analogous to the use of a poverty indicator to construct headcount measures of poverty. However, there is an important difference: measures like (8) measure vulnerability to a certain type of serious future welfare loss, not existing low welfare. A child who has already dropped out of school and become involved in child labor may have very low welfare, but is not vulnerable in the sense we are using the term – the worst outcome has already been realized for that child. Therefore, working non-school children make no contribution to the vulnerability index.

4.5 Identifying the vulnerable

33. Once individual-level indicators of vulnerability $V(A \mid X)$ or of high vulnerability d_v have been defined, they can be averaged to give measures of the mean level or incidence, or aggregated to give the total number of highly vulnerable people, within different population groups. This involves the standard problem of estimating population means, proportions and totals from survey data.

34. Preventative policy has the objective of moderating the risks of school drop-out and child labor for the relevant group of children. In practice, careful targeting of policy is necessary to avoid the excessive costs associated with untargeted policy

---

[8] Note that we are attaching no special welfare loss to the state of idleness (no schooling and no work). Thus a non-working schoolchild who left school without taking up work would be attributed a welfare loss of $k_S(A)$. Conversely, an idle child who starts working will be attributed a welfare loss of $k_W(A)$. Other more complex welfare weighting systems are possible. In the case of school drop-out, the rate of harm accumulation g can be interpreted as the rate of return to schooling and the weight $(1+g)^{T-A}$ as the ratio of potential earnings with full schooling to age T to potential earnings with schooling truncated at age A.

options. The targeting of policy requires that it be linked to a characteristic, captured by a categorical indicator *Z*, which is observable by policy-makers and can therefore be built into policy rules. Thus alternative values of *Z* identify different groups within the child population. Let $V = V(A \mid X)$ be the relevant individual vulnerability indicator. For policy design purposes, we need to know the conditional probability $\Pr(Z=z \mid V > v)$ for each group *z*, where *v* is any specified vulnerability threshold. Call this probability $f_v(z)$. By altering the value *v* we can trace out the population groups which are associated with high, medium and low vulnerability, thus giving a clear set of priorities for policy design.

35. The natural sample analog of $f_v(z)$ is the ratio:

$$\hat{f}_v(z) = \frac{\sum_{i=1}^{n} \omega_i 1(v_i > v, z_i = z)}{\sum_{i=1}^{n} \omega_i 1(v_i > v)} \qquad (9)$$

where v_i and z_i are the observed values of *V* and *Z*, ω_i is the survey weight for household *i* and 1(.) is the indicator function. In surveys which are non-self-weighting, the sums in (9) can be replaced by weighted sums to avoid bias.

36. The sample distribution (9) gives the proportion of the set of children with vulnerability higher than *v* who are members of the identifiable group defined by $Z = z$. It therefore tells us where the vulnerable tend to be located within the population of children. In policy terms, it also tells us the proportion of the vulnerability problem that could be eradicated by a perfectly efficient policy targeted at group *z*. However, it does not say anything about the cost of a policy directed at group *z*. Assume that there is no possibility of finer targeting of policy within groups (for example, because individual means-testing is impractical). The cost of the policy is then $c_z M(z) N$ where c_z is the cost per head of the policy within group *z*, *N* is the size of the school-age child population and *M*(*z*) is the weighted proportion of the population who are in

group *z*: $$M(z) = \frac{\sum_{i=1}^{n} \omega_i 1(z_i = z)}{\sum_{i=1}^{n} \omega_i} \qquad (10)$$

37. Note that any group *z* which has a substantially higher value for $\hat{f}_v(z)$ than for *M*(*z*) can be regarded as a group with a high concentration of vulnerability.

38. A policy analysis designed to locate socio-economic-geographic areas where preventative policy could protect vulnerable children in a cost-efficient way should focus attention on groups with high values for $\hat{f}_v(z)$ and low values for $c_z M(z)$ - the policy "treatment" would thus be cheap but efficiently targeted. A useful initial step in the process of policy design is to carry out the following steps:

- Partition the population (for example, by age, location, type of land holding, etc.) to define the groups *z*;
- Select a vulnerability threshold *v*;

- Calculate the vulnerability map $\hat{f}_v(z)$ and the relative group sizes $M(z)$;
- Rank the groups by their values of $\hat{f}_v(z)$ and $[c_z M(z)]^{-1}$ and select for particular policy attention those groups that rank high on both criteria.

39. This procedure can be repeated for different values of the vulnerability threshold v to explore the effect of using a more or less stringent definition of vulnerability.

5. THE DATA SET

5.1 The Guatemalan Living Standards Measurement Survey 2000

40. Information on poverty, household conditions and other variables was collected in Guatemala through the 2000 Living Standards Measurement Survey (ENCOVI, 2000). The survey followed a probabilistic design, covering 7,276 households (3,852 rural and 3,424 urban) . The survey is representative at the national and regional level as well as in urban and rural areas. ENCOVI included questions to elicit a unique level of detail on themes related to vulnerability. The survey included modules on risks and shocks; conflict, crime, and violence; social capital and migration. The data set for Guatemala is also unique in containing information on access to credit and insurance. As most of our attention will be devoted to such variables, we now discuss their exact definition[9] and present some summary statistics.

41. *Credit rationing.* The survey contains a set of questions related to access to credit. In particular, households are asked whether they have applied for credit and, in case of application, whether they were denied the credit. We define as "credit rationed" households that did not apply for credit for one of the following reasons: a) Institutions offering credit not available b) Does not know how to ask for credit c) Does not have the required characteristics d) Does not have collateral e) Interest rates too high f) Insufficient income g) Institutions do not give credit to household in that conditions. We also classify as credit rationed households that applied for, but were denied, credit.

42. *Shocks.* ENCOVI 2000 contains a set of questions pertaining to the occurrence of shocks. Shocks are divided into two broad categories: collective and individual. Collective shocks include events like earthquakes, floods, fires etc. Individual shocks include loss of employment, death, *etc.*[10] Households can report more then one shock for each group. We have, however, classified a household as being hit by a shock if it reported at least one shock. In the analysis we used separate dummies for collective and individual shocks. Other classifications were also tried, but did not change the main results. About 50% of households surveyed reported experiencing one or more shock in year 2000; of these, 12% reported natural or economic shocks affecting the community, 20% shocks directly affecting the family and 20% affecting both. Of the 7,276 households surveyed, 38% were affected by individual (idiosyncratic) shocks and about 30% by collective shocks. The most frequently reported collective shock is a general increase of prices. This could reflect a misperception of the economic environment or just a generic complaint about the cost of living. In any case, excluding this form of shock from the definition of the dummy variables does not change the results obtained.

43. *Risk reduction and risk coping mechanisms*. The questionnaire allows us to identify whether an individual has medical insurance (public or private). A dummy variable was created, taking value of 1 if at least one member of the household has medical insurance. Information was insufficient to identify whether households belonged to an informal social support network.

---

[9] For a detailed discussion and analysis of these variables see Guarcello, Mealli and Rosati (2002).

[10] For a further description and analysis see Tesliuc and Lindert (2002).

44. *Child and household characteristics.* We have employed a set of control variables (see Appendix 1) to take into consideration individual and household characteristics.[11] The control variables include: a gender dummy; a dummy variable taking the value 1 if the child belongs to an indigenous household; the number of the household members; the number of children aged 0-5 in the household and the number of school age children; an interaction dummy variable taking the value 1 if the child is a girl and there are children aged 0-5 in the household; and a series of dummy variables for the education of the mother and of the father.

5.2 Children's work in Guatemala

45. Child labor is very common in Guatemala. Some 506,000 children aged 7-14 years, one-fifth of total children in this age group, are engaged in work. Most are employed on the family farm or in petty business and are located in rural areas. Guatemala ranks third highest in child work prevalence of the 14 Latin America and Caribbean countries where data are available, behind only Bolivia and Ecuador. In terms of GDP per capita, the country ranks fifth lowest of the 14 countries. Guatemala's relative level of child work is therefore high compared to its relative level of income.

Table 3. - Children aged 7-14, by sex, type of activity and residence

Sex	Activity	Urban		Rural		Total	
		%	No.	%	No.	%	No.
Male	Work only	4.3	19,285	12.3	104,161	9.5	123,446
	Study only	73.9	334,299	53.9	455,964	60.9	790,263
	Work and study	10.1	45,587	19.7	166,924	16.4	212,511
	Total work*	14.4	64,872	32.0	271,085	25.9	335,957
	Total study**	78.2	379,886	73.6	622,888	67.3	1,002,774
	Neither	11.8	53,308	14.1	119,329	13.3	172,637
Female	Work only	4.1	17,820	6.8	54,249	5.9	72,509
	Study only	74.6	323,451	58.4	464,030	64.1	787,76
	Work and study	7.6	32,764	8.3	66,386	8.1	99,546
	Total work*	11.7	50,584	15.1	120,635	14.0	172,055
	Total study**	82.2	356,215	66.7	530,416	72.2	887,310
	Neither	13.8	59,770	26.5	210,491	22	270,371
Total	Work only	4.2	37,105	9.7	158,410	7.7	195,515
	Study only	74.2	657,750	56.1	919,994	62.4	1,577,744
	Work and study	8.8	78,351	14.2	233,310	12.3	311,661
	Total work*	13.0	115,456	23.9	391,720	20.0	507,176
	Total study**	83.0	736,101	70.3	1,153,304	74.7	1,889,405
	Neither	12.8	113,078	20.1	329,820	17.5	442,898

* 'Total work' refers to children that work only and children that work and study.

** 'Total study' refers to children that study only and children that work and study.

Source: *Encuesta de Condiciones de Vida* (ENCOVI) 2000. *Instituto Nacional de Estadisticas* (INE) Guatemala

Note that the figures here are not comparable with those reported below because of differences in the age group and because some children classified here as non-working may have worked previously and are therefore classified as members of the child labor force subsequently.

46. In Table 2 we use the age range 7 to 14 to define child labor. School starts at 7 in Guatemala and no significant amount of child labor is found below the age of 7. The basic cycle of education (*ciclo basico*) requires in most cases 9 years of study to be completed. However, note that current legislation allows children to work legally

[11] The rationale for the use of these variables is well known in the literature on child work, see Cigno et al (2001) and the literature cited therein.

from the age of 14. We decided to keep the age range coherent with the completion of the basic cycle of education, also to facilitate international comparison. Nothing of substance changes in the results if we define child work over the age range 7- 13.

47. Table 2 gives detailed information on children's activities in Guatemala. Note in particular that a large proportion, 17%, of children is 'idle': reportedly neither working nor attending school. This group includes children (mainly girls) performing full time household chores, 'hidden' workers and children for whom school attendance is too expensive or impossible due to lack of infrastructure but lacking opportunities to perform productive activities. These idle children, a group almost as large as that of working children, also constitute an important policy concern. They not only do not go to school, but are at risk of becoming part of the labor force. This group is the most sensitive to changes in policy and in exogenous variables. Table 2 shows that gender differences in child activity status are important: boys are more likely to work, but girls are more likely to be neither working nor attending school. It also shows that children of indigenous households have a lower school attendance rate and a higher work participation rate than the rest of the population.

6. ESTIMATES OF THE TRANSITION MODEL FOR GUATEMALA

6.1 The time profile of hazard rates

48. Two sets of estimated parameters from model (2) are given in Appendix Tables A2.1 and A2.2. These correspond to the quadratic and semi-parametric models respectively. Comparing the two log-likelihood values via the Akaike Information criterion, the semi-parametric specification (AIC = 3.026) is clearly superior to the quadratic model (AIC = 3.078). The estimated time profiles, which are based on semi-parametric specification, are shown in Figure 2 for the three hazards $P_1(A|X)$ ... $P_3(A|X)$ relating to non-working schoolchildren and in Figure 3 for the two hazards $Q_1(A|X)$ and $Q_2(A|X)$ relating respectively to working schoolchildren and 'idle' children. In plotting these hazard functions we have set the X-vector at its sample mean value. Similar graphs for the quadratic model are reported in Appendix 2. Unlike the parametric model which forces the time dependence to be quadratic, the semi-parametric specification shows an obvious peak at age 10 for working related hazards, $P_2(A|X)$, $P_3(A|X)$, $Q_2(A|X)$. This suggests that age 10 is commonly considered an appropriate time to start working, and thus children close to that age are more vulnerable with respect to child labor. School-leaving hazards are instead smoother, showing a rising pattern until age 13. It is interesting to note that work hazards, though rising, are never higher than 0.35 for the 'average' schoolchild, while the rising pattern for idle children is steeper, with a work hazard of 0.75 by age 13.

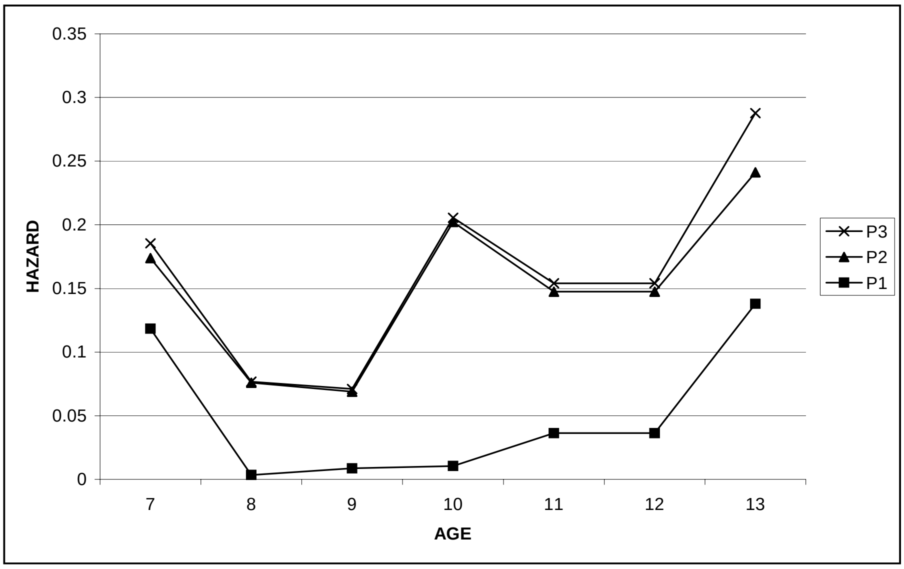


Figure 2. - Hazard functions for non-working schoolchildren (semi-parametric model)

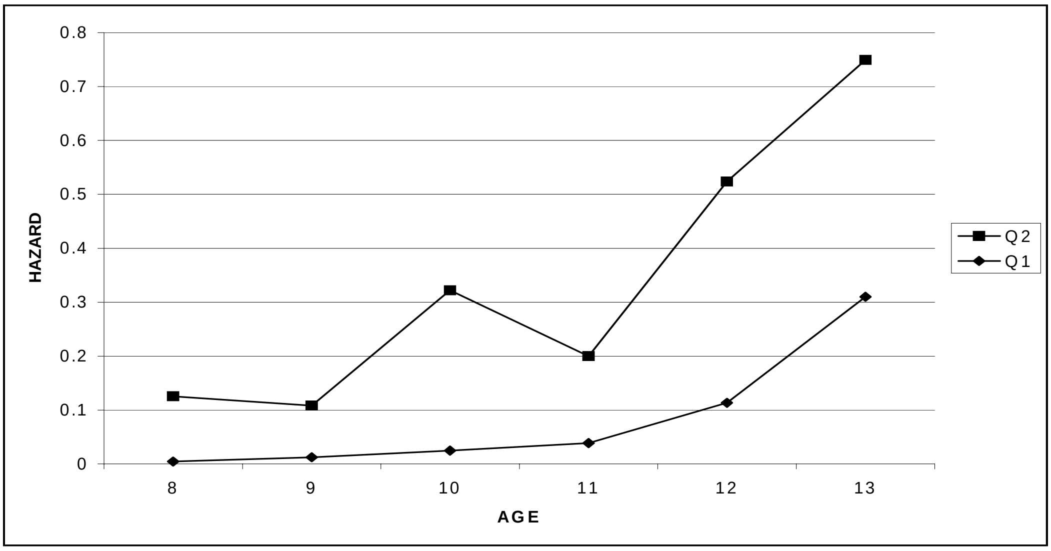


Figure 3. - Hazard functions for working schoolchildren and 'idle' children (semi-parametric model)

6.2 The impact of individual and household characteristics and of external shocks

49. The degree of complexity of the estimated model does not allow interpretation of the results by simply looking at the estimated coefficients. It is more useful to plot the three hazard functions $P_1(A|X)$ ... $P_3(A|X)$ relating to non-working schoolchildren and the two hazard functions $Q_1(A|X)$ and $Q_2(A|X)$ related respectively to working schoolchildren and 'idle' children for different levels of one explanatory variable, while holding all the others at their sample mean value. Figures 4a and 4b show the hazards for males and females: gender differences are more pronounced for $P_2(A|X)$ and $Q_2(A|X)$ (the hazards of entry into work), with males having a higher probability to start working than females. Females are at higher risk of dropout from school, whether or not they have started working. Note also that $P_1(A|X)$ at age 7 is higher for females than for males, and this is due to the likely event that girls never enter school.

50. In Appendix 2, similar figures show the effects of other individual and household characteristics and external shocks.

51. For example, indigenous have a higher risk of starting to work, irrespective of being or not still at school.

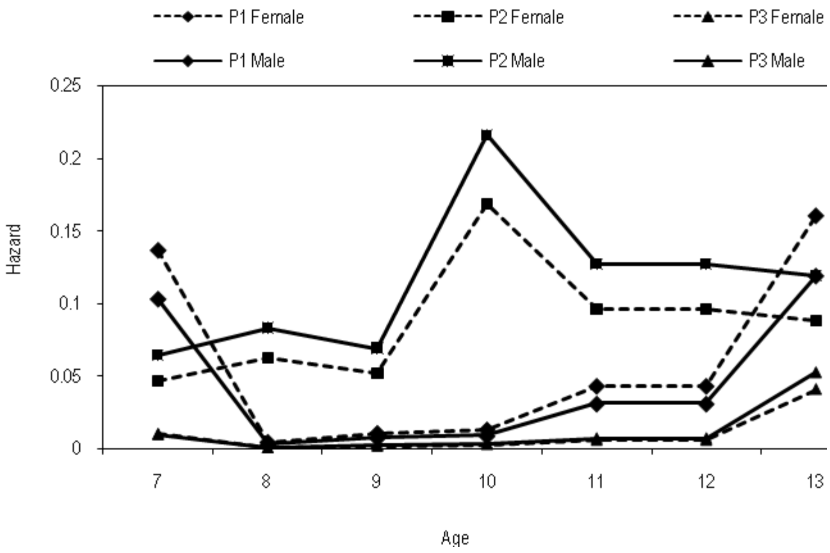


Figure 4.a. - Hazard functions for non-working schoolchildren by gender (semi-parametric model)

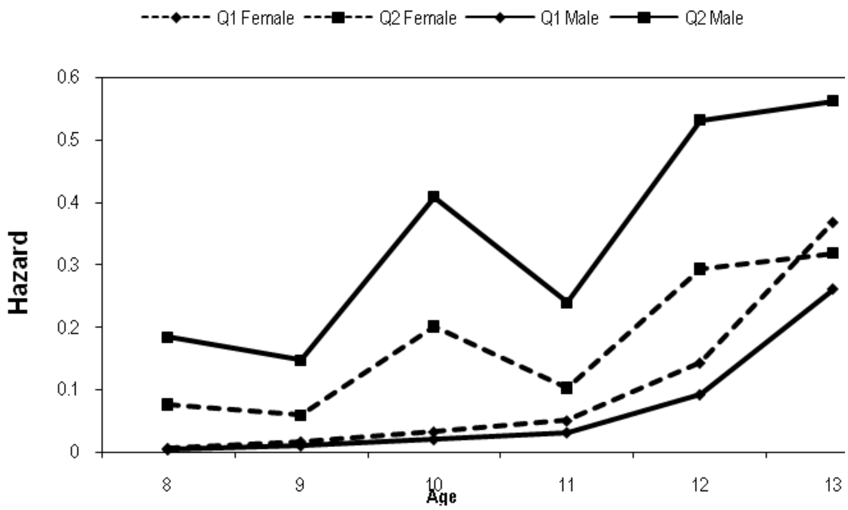


Figure 4.b. - Hazard functions for working schoolchildren and 'idle' children by gender (semi-parametric model)

7. VULNERABILITY IN GUATEMALA

52. The estimated transition model can be used to construct a range of indicators of vulnerability. We consider the size distribution of predicted vulnerabilities in section 7.1 and consider them in aggregated form by demographic group in section 7.2. Issues of policy targeting are addressed in section 8.

7.1 The distribution of vulnerability

Figures 5a and 5b depict the cross section distribution of vulnerability with respect to school drop-out (V_S, equivalent to $\theta = 1$) and child labor (V_W, equivalent to $\theta = 0$) for the survey year 2000. In calculating V_S and V_W, we have assumed a harm accumulation rate of g = 10% and made separate calculations for male and female children. Figure 5a shows the empirical density function and Figure 5b shows the cumulative distribution function, indicating the proportion of children with vulnerabilities below any given threshold. In general, the value 0.4 seems a reasonable choice as the threshold to define particularly high vulnerability and we adopt that value henceforth.

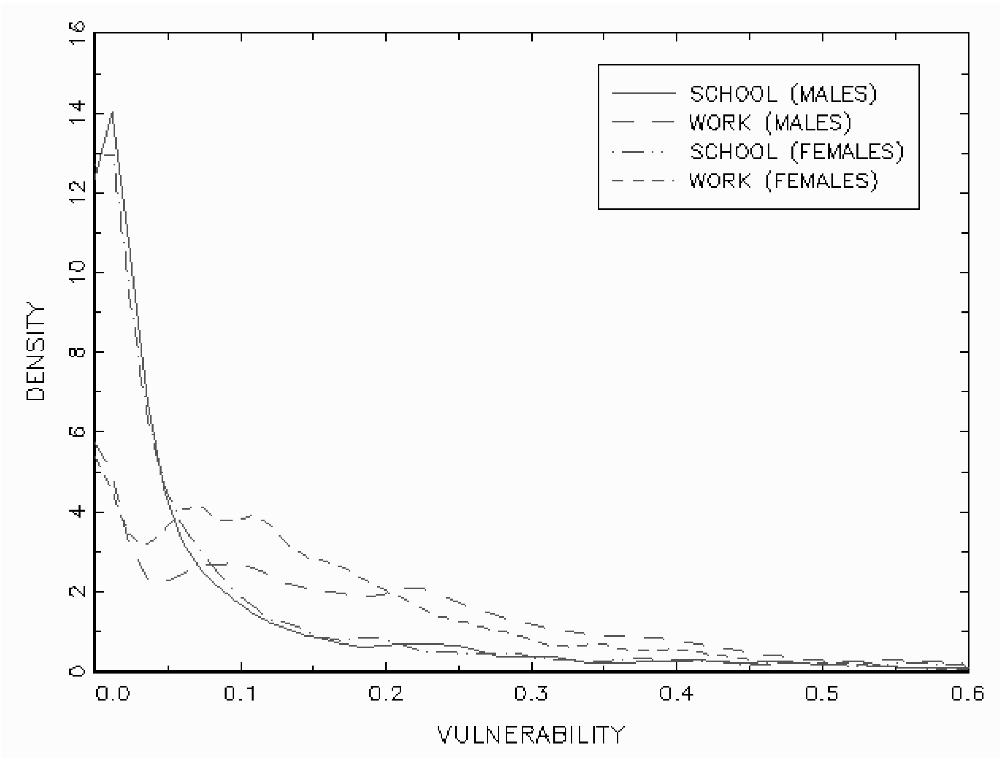


Figure 5. a. - The cross-section distribution of school and work vulnerabilities (weighted Gaussian kernel density estimates with adaptive bandwidth)

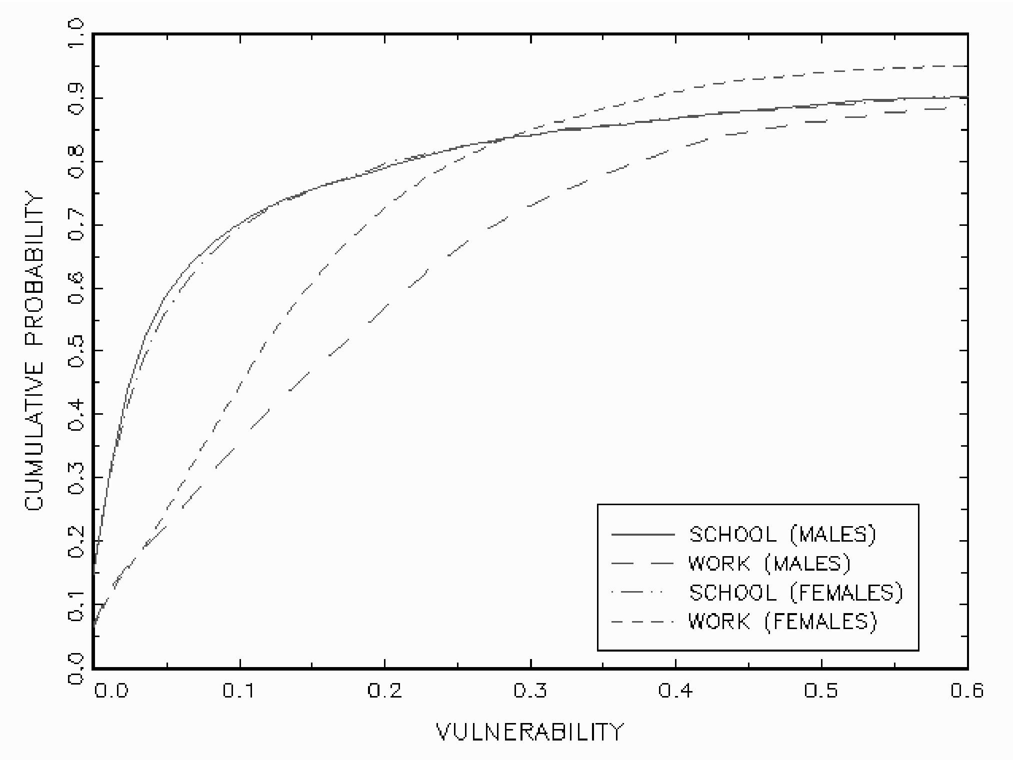


Figure 5.b. - The cross-section distribution of school and work vulnerabilities (weighted Gaussian kernel density estimates with adaptive bandwidth)

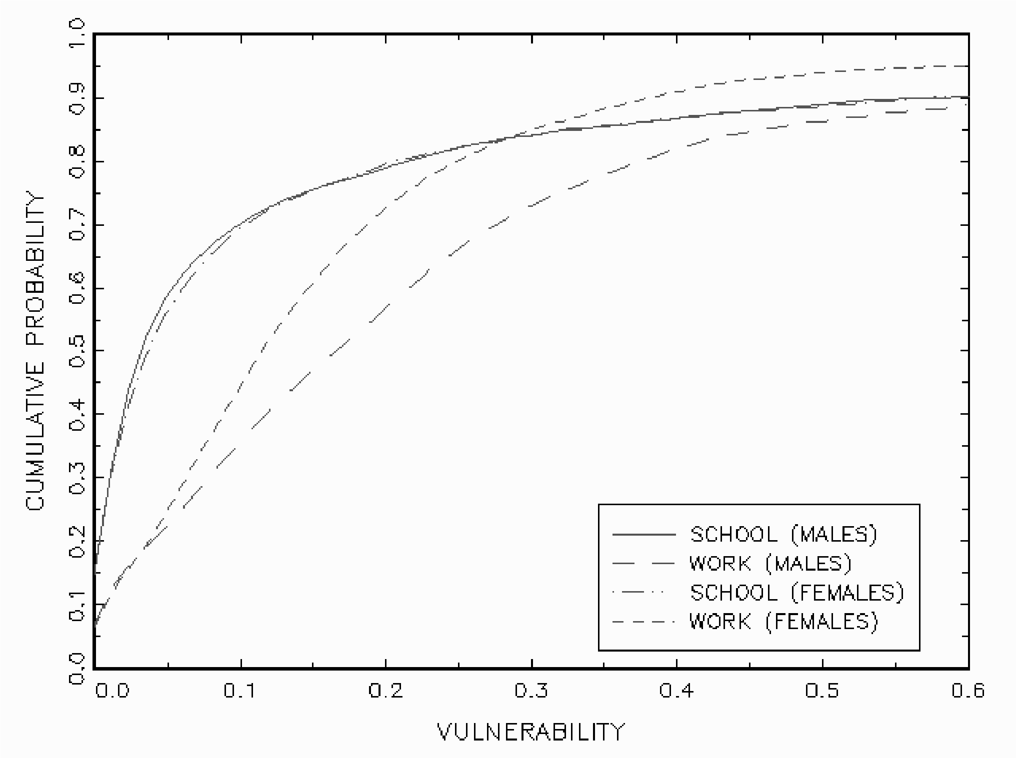


Figure 5.b. - The cross-section distribution of school and work vulnerabilities (weighted Gaussian kernel density estimates with adaptive bandwidth)

53. Vulnerability to school drop-out has a very similar distribution for males and females, with the vast majority of children having low vulnerability but a significant minority highly vulnerable to drop-out: for example, 13% of boys and girls have vulnerability above a threshold of 0.4. This has a very important important implication for policy: the existence of a small group of high estimated vulnerabilities implies that there is scope for targeting of policy towards children possessing the specific combination of characteristics associated with high risk of school drop-out.

54. The picture is rather different for child labor, where there is a large difference between the vulnerability of boys and girls. For boys, 18% have vulnerabilities to child labor in excess of 0.4, compared with 9% for girls.

7.2 Aggregate vulnerability

55. The estimates of the hazard functions $P_1(a_i|x_i)...P_3(a_i|x_i)$, $Q_1(a_i|x_i)$ and $Q_2(a_i|x_i)$, calculated at a_i = age of child i at time of the interview + 1, give individual-level measures of vulnerability to different risks conditional on current school-labor force status. Figure 6a shows the number of children who were school-going in 2000 when aged A-1 but who have a particularly high risk (above 0.4) of leaving school at age A, compared with the number of children who had already left school before age A. These are plotted against age A. Figure 6b is similar, but shows the predicted addition to the stock of children not attending school: in other words total vulnerability rather than high vulnerability. Figures 6c and 6d show analogous plots for child labor. The numerical values underlying Figures 6a-d are given in Table A2.3 of Appendix 2.

56. Figure 6a shows that high vulnerability to school drop-out is clearly associated primarily with the final year of compulsory schooling at age 13, although a significant number of boys are at high risk of drop-out when aged 12. However, as Figure 6b shows, there is more widely-spread general risk of school drop-out, even for those as

young as 7.[12] This finding has implications for the design of policy: very early school-leaving does not appear to be associated with an specific combination of the observable characteristics we have used in the econometric modelling. Thus, a policy designed to combat this problem needs to be of wide scope.

57. Figure 6c suggests that high vulnerability to the onset of child labor is significant rather earlier than vulnerability to school drop-out, particularly for boys, where age 10 appears to be a critical stage.[13] There is a considerably larger group of boys at high risk of child labor than of girls. The large number of predicted transitions into child labor apparent in Figure 6d indicate a widely-distributed low level of vulnerability to child labor in addition to the specific pools of highly vulnerable children.

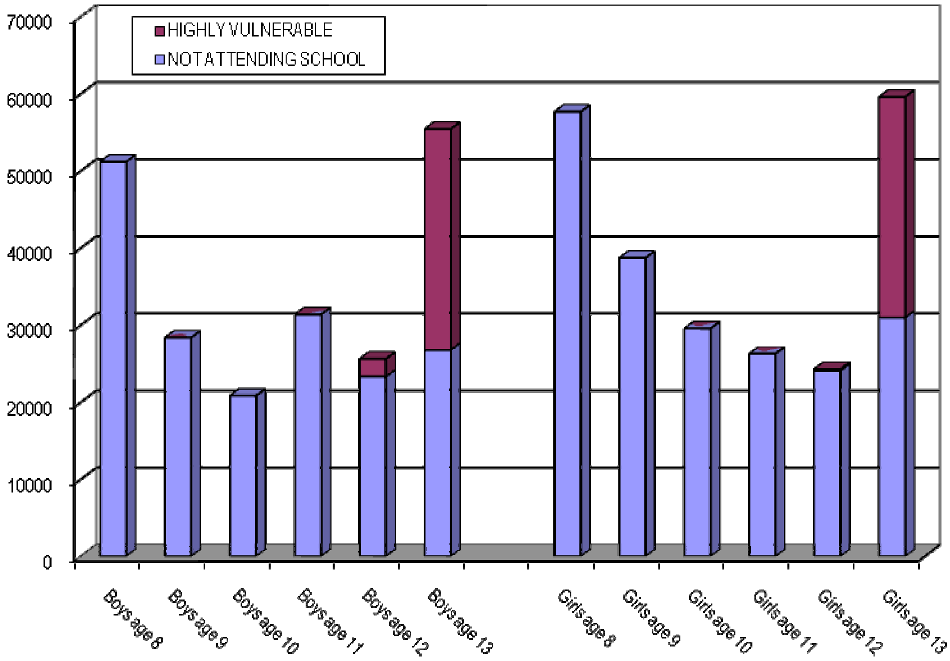


Figure 6.a. - Numbers of children not attending school and children highly vulnerable to drop-out in the coming year (v = 0.4)

[12] Recall that we treat children who never enroll in school as drop-out at age 7.

[13] Note that there is a possibility that the peak at age 10 is partly the result of rounding-off error in recall of the timing of entry into the labor force.

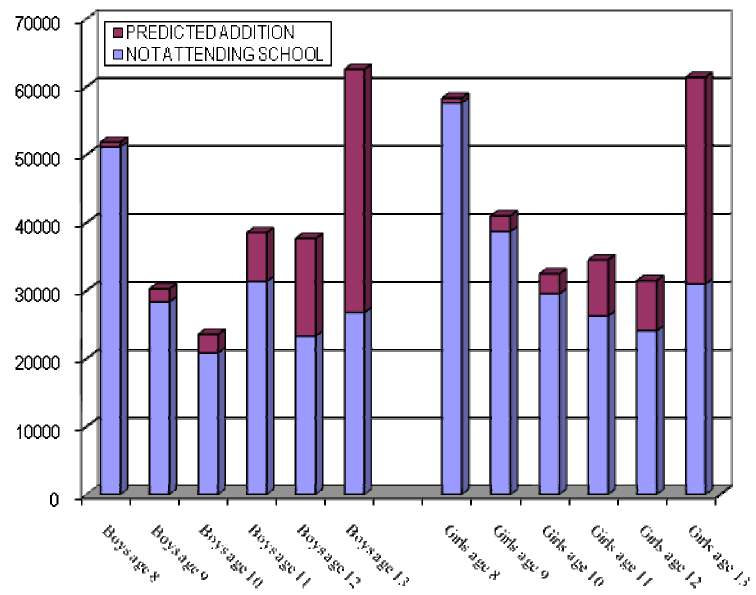


Figure 6.b. - Numbers of children not attending school and predicted flow of children dropping-out in the coming year

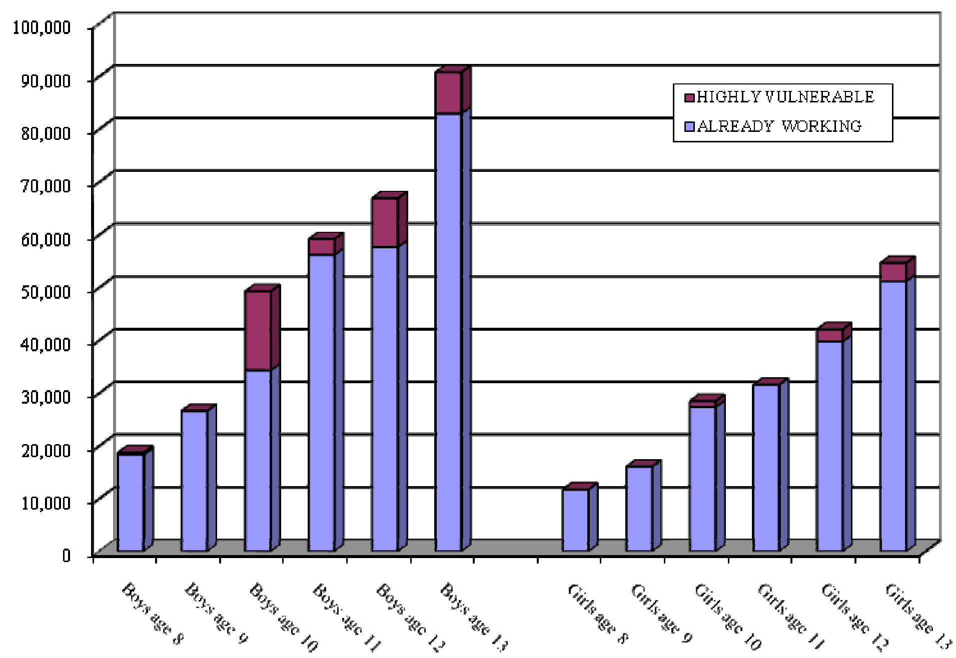


Figure 6.c. - Numbers of children in work and children highly vulnerable to onset of child labor in the coming year (v = 0.4)

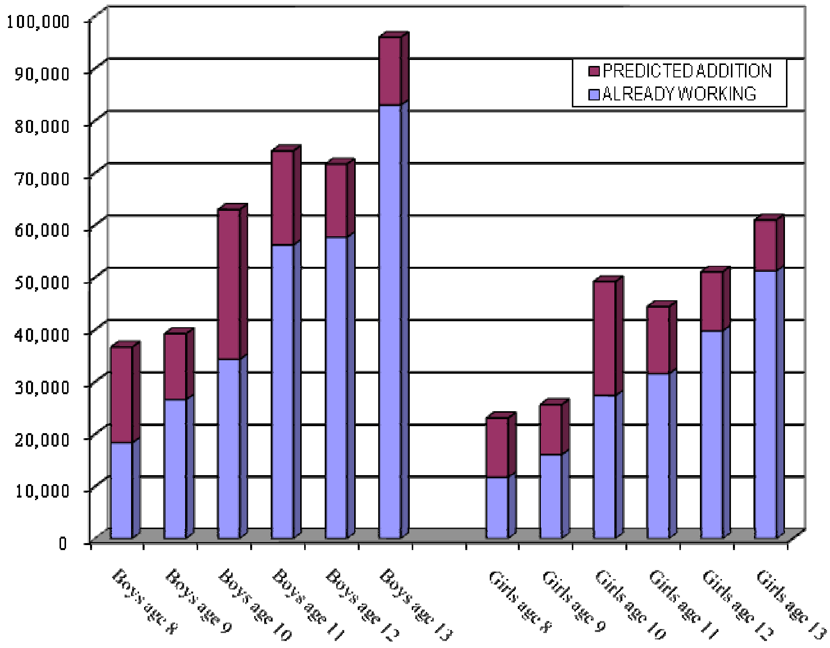


Figure 6.d. - Numbers of children in work and predicted flow of children commencing child labor in the coming year

58. To show how the vulnerability indicators can be combined, we now examine the distribution across gender-age groups of education-specific and work-specific vulnerabilities. Figure 7a shows the sample mean values of the indicators V_W and V_S defined above, by age-gender sub-groups. As for the previous tables, age refers to age at the time of the interview, but the indicators are predictive, relating to the following year. The pattern is revealing. The age profile of the risk of drop-out from school rises strongly as the legal school leaving age of 14 is approached. In contrast, the risk of onset of child labor is much flatter, with a clear peak among 9-year-olds (who will turn 10 in the current year), which is a consequence of the shape of the hazard functions P_2, P_3 and Q_2 (see Figures 2b and 3b). For most of the age range, boys have a higher risk of child labor than girls.

59. Figures 7b and 7c give a similar analysis for females and males separately for different specifications of the combined vulnerability indicator V. The four cases are: (i) $\theta = 0.5$ and $g = 5\%$; (ii) $\theta = 0.5$ and $g = 10\%$; (iii) $\theta = 0.75$ and $g = 5\%$; (iv) $\theta = 0.75$ and $g = 10\%$. These give equal or greater weight to school leaving than to child labor and they assume a moderate or high rate of harm accumulation. The peak at age 9 (*i.e.* relating to the 10th year) remains striking, as does the broad similarity between the patterns for males and females.

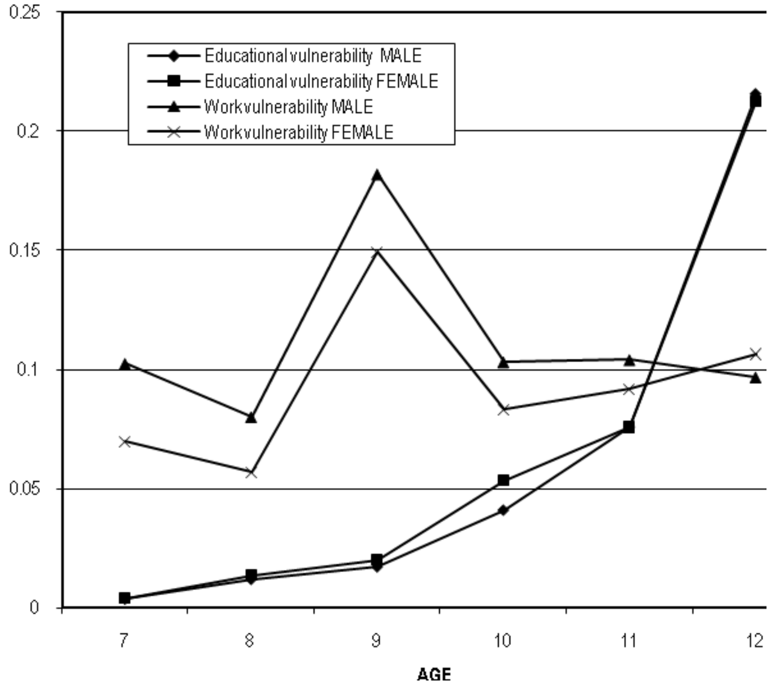


Figure 7.a. - Age-gender profile of work-specific and education-specific vulnerability indicators

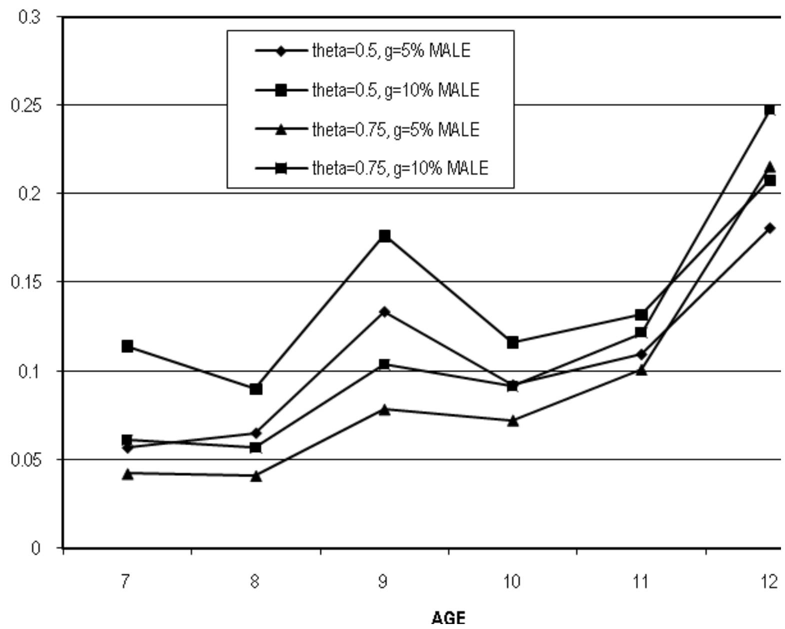


Figure 7.b. - Age profile for combined vulnerability indicators for males

7.3 The role of credit constraints and shocks

60. It is widely believed that poor access to credit is often the cause of premature school-leaving and child labor. Essentially, for families that are unable to use credit markets to smooth their consumption stream, children's education and work potential act as buffers against adverse events. This is explored in Table 4a, which shows the estimated causal impact of credit-rationing on school drop-out and child labor. The method here is as follows. We first estimate the hazards of school-drop out and child labor for each child, with the credit rationing covariate set to zero. We then repeat the simulation with the credit rationing variable set to unity for each child. The results are then summarized within three education-work status groups. Analogous simulations of the impact of access to insurance and of individual and collective shocks are described in Tables 4b-d.

61. The simulation results do not show any very large impacts. In the case of shocks this is at least in part due to the fact that we do not have data on shocks in years prior to 2000 and we would expect this to bias their predicted impact downwards. Access to credit and insurance are likely to be long-term characteristics, so one might expect their influence to be more reliably estimated. The estimated coefficients on which the simulated impacts in Table 4a are based are statistically insignificant except for that relating to school drop-out for non-working schoolchildren. For this group there is a moderately increased risk of school drop-out but not of child labor. Thus credit rationing tends to increase the flow of children into the "idle" state. Table 4b relates to the effect of access to insurance, which is mainly to reduce the risk of premature school-leaving. "Idle" children in particular are also afforded some protection against child labor. Individual shocks (Table 4c) raise the risk of child labor for non-working schoolchildren and of school drop-out for schoolchildren who are already working. Both these effects are consistent with the idea that children's work tends to increase to meet the urgent family needs that may follow a family-specific shock. In contrast, there is little evidence of a major impact of wider-scale collective shocks (Table 4d). One interpretation of this finding is that there may be little demand for additional child labor following collective shocks such as failed harvests or adverse market conditions.

Table 4.a. - The causal effect of credit rationing (semi-parametric model)

Current activity	% expected to leave school in the next year	% expected to enter labor force in the next year	% with hazard of leaving school > 0.4	% with hazard of entering labor force > 0.4
	Credit rationed			
Non-working schoolchildren	5.6	11.1	1.7	1.0
Working schoolchildren	15.7	-	12.3	-
Idle children	-	18.4	-	9.1
	Not credit rationed			
Non-working schoolchildren	4.5	12.7	0.9	1.7
Working schoolchildren	15.9	-	12.7	
Idle children	-	21.8		14.1

Table 4.b. - The causal effect of insurance (semi-parametric model)

Current activity	% expected to leave school in the next year	% expected to enter labor force in the next year	% with hazard of leaving school > 0.4	% with hazard of entering labor force > 0.4
		Insurance		
Non-working schoolchildren	5.0	10.5	1.3	0.7
Working schoolchildren	15.0	-	11.3	
Idle children	-	17.7		7.6
		No insurance		
Non-working schoolchildren	5.1	12.2	1.3	1.2
Working schoolchildren	15.9	-	12.5	
Idle children	-	20.1		11.3

Table 4.c. - The causal effect of individual shocks (semi-parametric model*)*

Current activity	% expected to leave school in the next year	% expected to enter labor force in the next year	% with hazard of leaving school > 0.4	% with hazard of entering labor force > 0.4
		Individual shock		
Non-working schoolchildren	5.2	13.6	1.4	2.1
Working schoolchildren	17.4	-	15.3	-
Idle children	-	19.3	-	9.8
		No individual shock		
Non-working schoolchildren	5.1	10.6	1.2	0.5
Working schoolchildren	14.2	-	9.4	
Idle children	-	20.0		10.6

Table 4.d. - The causal effect of collective shocks (semi-parametric model)

Current activity	% expected to leave school in the next year	% expected to enter labor force in the next year	% with hazard of leaving school > 0.4	% with hazard of entering labor force > 0.4
		Collective shocks		
Non-working schoolchildren	4.8	13.2	1.1	1.7
Working schoolchildren	14.2	-	10.2	
Idle children	-	23.0		15.7
		No collective shock		
Non-working schoolchildren	5.2	11.3	1.3	0.8
Working school children	16.9	-	14.3	
Idle children	-	18.4		8.5

8. POLICY TARGETING

62. We now consider the issue of policy targeting, discussed in section 4.2 above. To implement the approach proposed there, we need to partition the population into a set of identifiable subgroups who might be separately targeted by policy interventions. We use two alternative 32-group partitions of 8 regions by 2 ethnicity groups (indigenous/non-indigenous) by either a male/female classification or a landowner/landless classification. Tables 5a-c give the results for the former and Tables 6a-c for the latter. In these tables, we use alternative values for the relative weight attached to premature school-leaving of $\theta = 0$ (only child labor matters), 0.5 (equal weight for premature school-leaving and child labor), 0.75 (school drop-out and child labor weighted 3:1), and 1 (only school drop-out matters). We also use alternative harm accumulation rates of g = 0%, 2.5%, 5%, 10%, implying increasingly high relative weights given to early years of school drop-out and child labor. In defining high vulnerability, we use a threshold of $v = 0.4$ so, for example, in the case (g = 0%, $\theta = 1$), a highly vulnerable child is one who has at least a 40% chance of dropping out of school in the coming year.

63. Two indicators are given in these tables. M(z) (defined by equation (10) above) is the survey-based estimate of the proportion of the total population of 7-12 year-old children who are in each group z, while fv(z) is the proportion of the total population of highly vulnerable children who are in group z. Note that for $\theta = 0$ the vulnerability index captures only work vulnerability, whereas for $\theta = 1$ the vulnerability index captures only school-leaving vulnerability. Tables 5a and 6a contrast these two cases assuming a harm accumulation rate of 0% - equivalent to ignoring the age at which school drop-out or child labor occurs, while Tables 5b and 6b explore the effect on these two extreme cases of assuming different harm accumulation rates. Tables 5c and 6c indicate the effect of using different weights for school drop-out and child labor, for alternative harm accumulation rates.

64. The results suggest very strongly that it is among the indigenous population in the North, North-West, Peten and Central regions that vulnerability is most heavily concentrated. For example, consider the final column of Table 6c, corresponding to a heavy weight for school drop-out relative to child labor and for early relative to late events. These tell us that 74.3% of aggregate high vulnerability is located among the indigenous people of those four regions, who make up less than 36.7% of the total child population. The degree of high vulnerability among the male and female populations also differs substantially by region (see Table 5c).

65. Land tenure turns out to be important and the evidence here suggests that much of the problem could be reached by policy directed towards the landowning indigenous population. This might suggest that an educational/child labor policy linked in some way to an agricultural support policy directed towards indigenous peasant farmers in these four regions of Guatemala could be an efficient way of reaching the most vulnerable children. This is, of course, only a suggestive finding. The detailed design of such a policy would present challenging problems.

Table 5.a. - Indicators for policy targeting (v = 0.4)

Region	Gender	Ethnicity	*M*(z) (%)	The percentage of highly vulnerable children in group *z*: $\hat{f}_v(z)$	
				School drop-out only (θ = 1)	Child labor only (θ = 0)
				g = 0%	*g* = 0%
South	M	N	7.9	1.21	3.0
	F	N	7.2	0.8	2.6
	M	I	1.3	-	4.5
	F	I	1.3	1.76	1.1
North	M	N	0.9	1.5	0.6
	F	N	0.7	0.6	-
	M	I	3.6	12.4*	14.2*
	F	I	3.4	8.3*	0.7
North East	M	N	3.5	0.8	3.4*
	F	N	2.8	4.5*	1.1
	M	I	0.9	5.3	5.9
	F	I	0.6	-	-
South East	M	N	4.7	2.5*	2.4*
	F	N	4.2	3.3*	1.7
	M	I	0.1	-	0.4
	F	I	0.2	-	-
Central	M	N	3.0	3.8*	3.9*
	F	N	2.9	3.9*	-
	M	I	2.4	6.1*	6.9*
	F	I	2.1	3.5*	2.4*
South West	M	N	7.1	1.7	3.4*
	F	N	6.7	1.3	-
	M	I	7.6	3.0	12.3*
	F	I	7.2	2.1	-
North West	M	N	1.6	2.7*	-
	F	N	1.4	2.0*	-
	M	I	5.5	10.8*	18.4*
	F	I	5.4	11.1*	4.9*
Peten	M	N	1.6	1.2*	3.5*
	F	N	1.3	1.0*	-
	M	I	0.3	0.5	1.1*
	F	I	0.5	2.1*	1.7*
Total			100	100	100

Key Gender: M = male, F = female; Ethnicity: I = indigenous, N = non-indigenous; θ = welfare weight assigned to premature school exit relative to premature entry into labor force; *g* = accumulation rate; - denotes inadequate cell sample to estimate *f*; * denotes an estimate significantly different from 0 at the 5% level (2-sided)

Table 5.b. - Indicators for policy targeting (v = 0.4)

Region	Gender	Ethnicity	$M(z)$ (%)	The percentage of highly vulnerable children in group z: $\hat{f}_v(z)$					
				School drop-out only (θ = 1)			Child labor only (θ = 0)		
				g = 2.5%	g = 5%	g = 10%	g = 2.5%	g = 5%	g = 10%
South	M	N	9	1.07	0.0	0.7	2.8	2.0	3.7*
	F	N	7.2	0.7	1.8	1.8	1.9	1.3	0.9
	M	I	1.3	-	0.1	0.1	3.2	2.2	2.0*
	F	I	1.3	4.1	3.6	2.8	0.8	1.1	1.5
North	M	N	0.9	1.4	1.2	1.4*	1.0	1.0	0.6*
	F	N	0.7	1.1	1.0	0.9	-	-	0.1
	M	I	3.6	11.7*	11.0*	10.1*	12.4*	11.1*	9.4*
	F	I	3.4	7.6*	7.1*	6.8*	1.3	3.4*	3.8*
North East	M	N	3.5	1.3	1.3	1.6*	2.4*	1.8*	2.5*
	F	N	2.8	4.0*	3.5*	2.8*	1.4	2.2	1.5
	M	I	0.9	5.9	5.1	1.1	4.1	3.8	2.2*
	F	I	0.6	-	-	0.6	0.5	0.7	0.8
South East	M	N	4.7	2.5*	2.4*	3.2*	3.5*	4.6*	4.9*
	F	N	4.2	3.1*	2.7*	2.4*	1.5	1.0	1.4*
	M	I	0.1	-	-	-	0.3	0.2	0.1
	F	I	0.2	-	-	0.1	-	-	0.1
Central	M	N	3.0	3.4*	3.3*	3.7*	3.6*	3.4*	3.3*
	F	N	2.9	3.4*	3.0*	2.9*	0.4	0.5	0.7*
	M	I	2.4	5.5*	6.0*	5.4*	5.3*	5.3*	6.0*
	F	I	2.1	3.4*	3.6*	3.6*	3.4*	3.7*	3.4*
South West	M	N	7.1	1.5	2.3	2.5*	5.7*	4.8*	4.2*
	F	N	6.7	1.2	1.0	2.1	-	0.6	1.1*
	M	I	7.6	3.2	4.6*	4.5*	12.6*	11.3*	11.3*
	F	I	7.2	2.8	2.5	5.5*	1.6	3.8*	5
North West	M	N	1.6	2.9*	2.5*	2.0*	1.4	1.4*	1.8*
	F	N	1.4	1.8*	1.9*	1.7*	-	-	0.4
	M	I	5.5	11.4*	11.6*	11.9*	17.6*	16.2*	13.9*
	F	I	5.4	10.4*	10.7*	10.2*	6.5*	8.4*	8.8*
Peten	M	N	1.6	1.3*	1.9*	1.7*	2.7*	2.2*	1.7*
	F	N	1.3	1.0*	1.1*	1.1*	-	-	0.4*
	M	I	0.3	0.8*	0.7*	0.6*	1.0*	0.8*	0.9*
	F	I	0.5	1.9*	1.7*	1.3*	2.0*	1.6*	1.5*
Total			100	100	100	100	100	100	100

Key Gender: M = male, F = female; Ethnicity: I = indigenous, N = non-indigenous; θ = welfare weight assigned to premature school exit relative to premature entry into labor force; g = accumulation rate; - denotes inadequate cell sample to estimate f; * denotes an estimate significantly different from 0 at the 5% level (2-sided)

Table 5.c. - Indicators for policy targeting (v = 0.4)

				The percentage of highly vulnerable children in group z: $\hat{f}_v(z)$					
				Equal weights for school-leaving and child labor (θ = 0.5)			School-leaving and child labor weighted 3:1(θ = 0.75)		
Region	Gender	Ethnicity	$M(z)$ (%)	g = 2.5%	g = 5%	g = 10%	g = 2.5%	g = 5%	g = 10%
South	M	N	7.9	-	-	-	-	-	0.6
	F	N	7.2	-	-	-	-	0.9	0.8
	M	I	1.3	-	-	-	-	-	0.1
	F	I	1.3	-	-	-	-	2.8	3.7
North	M	N	0.9	4.1	2.7	1.4	-	1.0	2.1*
	F	N	0.7	4.5	3.0	1.5	1.0	0.8	0.6
	M	I	3.6	28.4*	24.7*	19.1*	18.3*	14.8*	11.9*
	F	I	3.4	13.7	10.2	6.3	8.3*	9.8*	8.2*
North East	M	N	3.5	-	-	-	-	-	-
	F	N	2.8	-	-	-	5.2	5.8	4.4
	M	I	0.9	-	-	1.0	-	-	2.8
	F	I	0.6	-	-	-	-	-	-
South East	M	N	4.7	-	4.2	5.7	3.9	3.1	2.4
	F	N	4.2	-	-	-	-	1.2	1.1*
	M	I	0.1	-	-	1.0	-	-	-
	F	I	0.2	-	-	-	-	-	-
Central	M	N	3.0	-	-	2.7	3.4*	3.3*	3.4*
	F	N	2.9	-	-	-	2.2*	4.5*	4.4*
	M	I	2.4	-	1.9	8.5*	7.6*	6.5*	5.8*
	F	I	2.1	2.8	1.9	4.6*	4.7*	4.2*	2.8*
South West	M	N	7.1	-	-	-	-	-	3.1
	F	N	6.7	-	-	-	-	2.1	1.4
	M	I	7.6	-	-	-	-	-	3.2
	F	I	7.2	-	-	-	1.6	1.3	0.9
North West	M	N	1.6	2.2	1.4	2.2	4.0*	3.1*	3.4*
	F	N	1.4	-	3.5	1.8	3.2	3.1*	2.3*
	M	I	5.5	14.7*	14.0*	18.2*	14.5*	13.3*	9.8*
	F	I	5.4	8.3	11.5*	12.1*	15.2*	11.9*	12.8*
Peten	M	N	1.6	1.7	6.6	4.9*	1.5	1.8*	3.1*
	F	N	1.3	-	-	0.5	0.4	0.5	0.7*
	M	I	0.3	2.6	1.7	2.1	0.7	0.9	0.6
	F	I	0.5	17.2*	12.8*	6.4*	4.3*	3.4*	2.3*
Total			100	100	100	100	100	100	100

Key Gender: M = male, F = female; Ethnicity: I = indigenous, N = non-indigenous; θ = welfare weight assigned to premature school exit relative to premature entry into labor force; g = accumulation rate.

- denotes inadequate cell sample to estimate f; * denotes an estimate significantly different from 0 at the 5% level (2-sided)

Table 6. a. - Indicators for policy targeting (v = 0.4)

Region	Land-ownership	Ethnicity	*M*(z) (%)	The percentage of highly vulnerable children in group *z*: $\hat{f}_v(z)$	
				School drop-out only (θ = 1)	Child labor only (θ = 0)
				g = 0%	*g* = 0%
South	NL	N	14.8	2.0	5.7*
	L	N	0.3	-	-
	NL	I	1.7	-	3.1
	L	I	1.0	1.8	2.6
North	NL	N	1.1	2.1*	-
	L	N	0.5	-	0.6
	NL	I	3.4	11.9*	5.4*
	L	I	3.7	8.8*	9.5*
North East	NL	N	4.7	2.7*	3.2*
	L	N	1.7	2.6	1.2
	NL	I	0.7	-	0.5
	L	I	0.7	5.3	5.4
South East	NL	N	4.9	2.5*	2.6*
	L	N	4.0	3.4*	1.5
	NL	I	0.1	-	-
	L	I	0.2	-	0.4
Central	NL	N	4.7	6.3*	3.3*
	L	N	1.3	1.4*	0.6
	NL	I	1.9	5.4*	3.1*
	L	I	2.6	4.2*	6.2*
South West	NL	N	9.6	2.3	3.4
	L	N	4.2	0.7	-
	NL	I	4.2	2.2	3.5
	L	I	10.6	2.9	8.9*
North West	NL	N	0.9	0.8	-
	L	N	2.1	3.9*	-
	NL	I	1.4	2.9*	1.6
	L	I	9.4	19.1*	21.7*
Peten	NL	N	1.7	1.1*	0.7
	L	N	1.2	1.1*	2.8*
	NL	I	0.5	1.8*	0.8
	L	I	0.3	0.9	2.0
Total			100	100	100

Key Land ownership: NL = landless, L = landowner; Ethnicity: I = indigenous, N = non-indigenous; θ = welfare weight assigned to premature school exit relative to premature entry into labor force; *g* = accumulation rate; - denotes inadequate cell sample to estimate *f*; * denotes an estimate significantly different from 0 at the 5% level (2-sided)

Table 6.b. - Indicators for policy targeting (v = 0.4)

Region	Land-ownership	Ethnicity	$M(z)$ (%)	The percentage of highly vulnerable children in group z: $\hat{f}_v(z)$					
				School drop-out only (θ = 1)			Child labor only (θ = 0)		
				g = 2.5%	g = 5%	g = 10%	g = 2.5%	g = 5%	g = 10%
South	NL	N	14.8	1.8	2.8	2.6*	4.7*	3.9*	4.6*
	L	N	0.3	-	-	-	-	-	-
	NL	I	1.7	1.0	0.9	0.7	2.1	1.5	2.5*
	L	I	1.0	3.1	2.7	2.2	1.8	1.8	0.9
North	NL	N	1.1	1.9*	1.7*	1.6*	0.1	0.4	0.4*
	L	N	0.5	0.6	0.5	0.7	0.9	0.6	0.3
	NL	I	3.4	11.3*	11.0*	10.8*	5.4*	5.7*	5.1*
	L	I	3.7	8.0*	7.0*	6.1*	8.3*	8.7*	8.0*
North East	NL	N	4.7	3.0*	2.7*	2.7*	2.2*	3.0*	2.2*
	L	N	1.7	2.3	2.0	1.6	1.5	1.1	1.7
	NL	I	0.7	1.1	1.0	1.4	0.8	0.9	0.7*
	L	I	0.7	4.7	4.1	3.3	3.8	3.5	2.2*
South East	NL	N	4.9	2.6*	2.5*	3.6*	2.5*	1.7*	2.1*
	L	N	4.0	3.0*	2.6*	2.1*	2.5*	3.9*	4.2*
	NL	I	0.1	-	-	0.1*	-	-	0.1
	L	I	0.2	-	-	-	0.3	0.2	0.1
Central	NL	N	4.7	5.5*	5.2*	5.6*	3.0*	2.7*	2.4*
	L	N	1.3	1.2*	1.1*	1.0*	0.4	1.1	1.6*
	NL	I	1.9	5.2*	5.5*	5.2*	3.0*	2.9*	2.8*
	L	I	2.6	3.7*	4.2*	3.7*	5.7*	6.0*	6.5*
South West	NL	N	9.6	2.0	2.8	3.5	2.4	1.7	1.5*
	L	N	4.2	0.6	0.5	1.1	3.3	3.7*3.7*	
	NL	I	4.2	2.9	3.8*	4.6*	2.4	2.4	2.6*
	L	I	10.6	3.1	3.2	5.4*	11.8*	12.7*	14.3*
North West	NL	N	0.9	1.2	1.1*	0.8	-	-	0.2
	L	N	2.1	3.5*	3.4*	2.9*	1.1	1.4*	2.0*
	NL	I	1.4	2.5*	2.9*	3.0*	1.6	1.7*	1.6*
	L	I	9.4	19.2*	19.4*	19.1*	22.5*	22.8*	21.0*
Peten	NL	N	1.7	1.2*	1.7*	1.4*	0.6*	0.4*	0.5*
	L	N	1.2	1.1*	1.3*	1.3*	2.1*	1.7*	1.5*
	NL	I	0.5	1.8*	1.6*	1.2*	0.8*	0.7*	1.1*
	L	I	0.3	1.0	0.8	0.6	2.2*	1.7*	1.2*
Total			100	100	100	100	100	100	100

Key Land ownership: NL = landless, L = landowner; Ethnicity: I = indigenous, N = non-indigenous; θ = welfare weight assigned to premature school exit relative to premature entry into labor force; g = accumulation rate.

- denotes inadequate cell sample to estimate f; * denotes an estimate significantly different from 0 at the 5% level (2-sided)

Table 6.c. - Indicators for policy targeting (v = 0.4)

Region	Land-ownership	Ethnicity	$M(z)$ (%)	The percentage of highly vulnerable children in group z: $\hat{f}_v(z)$					
				Equal weights for school-leaving and child labor (θ = 0.5)			School-leaving and child labor weighted 3:1 (θ = 0.75)		
				g = 2.5%	g = 5%	g = 10%	g = 2.5%	g = 5%	g = 10%
South	NL	N	14.8	-	-	-	-	0.9	1.4
	L	N	0.3	-	-	-	-	-	-
	NL	I	1.7	-	-	-	-	-	0.1
	L	I	1.0	-	-	-	-	2.8	3.7
North	NL	N	1.1	4.5	3.0	1.5	1.0	1.1	2.2*
	L	N	0.5	4.1	2.7	1.4	-	0.7	0.5
	NL	I	3.4	20.1*	16.5*	11.6*	14.8*	14.7*	11.9*
	L	I	3.7	22.0*	18.3*	13.8*	11.8*	9.9*	8.2*
North East	NL	N	4.7	-	-	-	-	1.7	1.7
	L	N	1.7	-	-	-	5.2	4.1	2.7
	NL	I	0.7	-	-	1.0	-	-	-
	L	I	0.7	-	-	-	-	-	2.8
South East	NL	N	4.9	-	4.2	3.2	0.7	0.9	2.4*
	L	N	4.0	-	-	2.5	3.2	3.3	2.2
	NL	I	0.1	-	-	-	-	-	-
	L	I	0.2	-	-	1.0	-	-	-
Central	NL	N	4.7	-	-	1.5	5.5*	7.6*	7.0*
	L	N	1.3	-	-	1.2	-	0.2	0.8
	NL	I	1.9	2.8	1.9	7.2*	8.8*	7.5*	6.1*
	L	I	2.6	-	1.9	6.0*	3.5*	3.2*	2.6*
South West	NL	N	9.6	-	-	-	-	2.1	3.8
	L	N	4.2	-	-	-	-	-	0.7
	NL	I	4.2	-	-	-	-	-	0.9
	L	I	10.6	-	-	-	1.6	1.3	3.1
North West	NL	N	0.9	-	3.5	1.8	1.6	1.2	1.4
	L	N	2.1	2.2	1.4	2.2	5.6*	5.0*	4.3*
	NL	I	1.4	7.3	4.9	4.2*	4.5*	3.5*	2.8*
	L	I	9.4	15.7*	20.6*	26.1*	25.3*	21.7*	19.8*
Peten	NL	N	1.7	1.7	1.1	1.0	0.8	1.2*	1.8*
	L	N	1.2	-	5.5	4.3	1.1	1.2	2.0*
	NL	I	0.5	10.8*	8.5*	5.5*	3.6*	3.2*	2.1*
	L	I	0.3	10.0	6.0	3.0	1.4	1.1	0.8
Total			100	100	100	100	100	100	100

Key Land ownership: NL = landless, L = landowner; Ethnicity: I = indigenous, N = non-indigenous; θ = welfare weight assigned to premature school exit relative to premature entry into labor force; g = accumulation rate; - denotes inadequate cell sample to estimate f; * denotes an estimate significantly different from 0 at the 5% level (2-sided)

9. CONCLUSIONS

66. Inadequate education and child labor are closely associated with chronic long-term child poverty. Policy intervention should target both the children out of school and/or working, but also those vulnerable to such negative outcome. While it has been since long recognised the need to extend the vulnerability analysis to other dimension of welfare, beside poverty as measured by consumption (or income), little has been done in this area. This paper, building on the risk management approach, proposes a methodology to assess the vulnerability of children to leave education and/or to become child labourers.

67. We make use of retrospective information about children's school attendance and age at which they begun to work to simultaneously estimate a set of hazards functions that defines the risk the children faces to drop out of school and/or begin to work. The methodology proposed does not need panel data, as it exploits retrospective information to build time profiles of the child. As panel data are seldom available, the approach proposed aims also to help producing reliable estimates of vulnerability based on relatively easily available information.

68. We apply this methodology to data from Guatemala and on the basis of the estimates we build different measure of vulnerability, both individual and aggregated and also some welfare indicators that tries to aggregate both the risk of leaving school and that of beginning to work.

69. The results indicate that few retrospective information are sufficient to build reasonable indicators of children's vulnerability.

70. In the case of Guatemala the estimates show that about 8 per cent of children attending school are at risk of dropping out. Of this group about half (4 per cent of the children attending school) has an high vulnerability. At the same time, about 12 per cent of children not working is vulnerable to begin work, with 8 per cent being highly vulnerable (i.e. with a probability higher than 0.4 to begin work). To put this figure in context consider that in Guatemala about 25 per cent of children aged 7 – 14 are out of school, if we include the vulnerable group this percentage would increase to 31 per cent (or to 28 per cent including only the highly vulnerable). Analogously, about 20 per cent of the children in the relevant age group work: including vulnerable children this figure would increase to 24 for per cent (or to 22 per cent including only the highly vulnerable).

71. While it goes beyond the scope of this paper to discuss in details the source of vulnerability, our estimates indicate that shocks, lack of risk copying mechanisms (insurance), ethnicity are among the major causes determining the risk of dropping out of school and/or beginning to work.

72. Finally, we have developed a set of simple indicators to identify characteristics that are associated with high vulnerability and that can be useful for policy targeting. It appears that a combination of gender, ethnicity, region of residence and land ownership are useful to identify group at risk. For example, 20 per cent of the children highly vulnerable to start working are indigenous boys located in the North West of the country. In the same region, a relatively large group of children is highly vulnerable to drop out of school and/or begin working if the household owns land.

73. The estimates about the vulnerability of children to work and/or school drop out, confirm the role of risk management policies have to play also in the area of child labor and schooling and how vulnerability analysis is essential to a correct design and targeting of intervention policies.

REFERENCES

Chaudhuri, S., Jalan, J. and Suryahadi, A. (2001). "Assessing household vulnerability to poverty: a methodology and estimates for Indonesia". University of Columbia: Working Paper.

Dercon, S. (2001). "Assessing vulnerability", Oxford: CSAE Working Paper.

Jalan, J. and Ravallion, M. (1999). "Are the poor less well insured? Evidence on vulnerability to income risk in rural China", *Journal of Development Economics* **58**, 61-81.

Jalan, J. and Ravallion, M. (2000). "Is transient poverty different? Evidence from rural China", *Journal of Development Studies* **88**, 82-99.

Lancaster, T. (1990). "The Econometric Analysis of Transitions Data". Cambridge: Cambridge University Press.

Holzmann, R. and S. Jorgensen (2002), "Social Risk Management: A New Conceptual Framework for Social Protection and Beyond". Washington: The World Bank.

Morduch, J. (1994). "Poverty and vulnerability", *American Economic Review* 84, 221-225.

Pritchett, L., Suryahadi, A. and Sumarto, S. (2000). "Quantifying vulnerability to poverty - a proposed measure applied to Indonesia". Washington: World Bank Policy Research Working Paper no. 2437.

Tesliuc, E. D. and Lindert, K. (2002). "Vulnerability: a quantitative and qualitative assessment", Washington: World Bank Working Paper, Guatemala Poverty Assessment Program.

APPENDIX 1: DATA DEFINITIONS AND DESCRIPTION

Table A1 - Variable definitions and sample means

Variable	Definition	Mean in sample for model estimation (n = 7,936)
Female	Dummy = 1 if individual is female	0.485
Indigenous	Dummy if individual is of indigenous origin	0.420
Log income	Natural logarithm of normal household income	6.316
Household size	Number of household members	7.124
No. of children under age 6	Number of children in the household aged 5 or under	1.142
No. of children over age 5	Number of children in the household aged 6 or over	3.015
Female × young children	= number of children aged under 6 if child is female; = 0 if child is male	0.725
Uneducated father	Dummy = 1 if father has no education	0.360
Father primary education	Dummy = 1 if father has primary education only	0.482
Uneducated mother	Dummy = 1 if mother has no education	0.514
Mother primary education	Dummy = 1 if mother has primary education only	0.364
Collective shock	Dummy = 1 if at least one collective shock is experienced	0.325
Individual shock	Dummy = 1 if at least one individual shock is experienced	0.403
Credit rationed	Dummy = 1 if household believes it is credit rationed	0.522
Medical insurance	Dummy = 1 if household has medical insurance	0.251
North	Dummy = 1 for residence in Northern region	0.114
North-east	Dummy = 1 for residence in North-eastern region	0.069
South-east	Dummy = 1 for residence in South-eastern region	0.110
Central	Dummy = 1 for residence in Central region	0.162
South-west	Dummy = 1 for residence in South-western region	0.174
North-west	Dummy = 1 for residence in North-western region	0.194
Peten	Dummy = 1 for residence in Peten region	0.091
Landowner	Dummy = 1 if the household owns its land	0.433
Age	Age of child in years	10.354

At school, not working	Dummy = 1 for enrolment in school and non-working status	0.546
At school, working	Dummy = 1 for enrolment in school and working status	0.213
Not at school, not working	Dummy = 1 for non-enrolment in school and non-working status	0.130
Not at school, working	Dummy = 1 for non-enrolment in school and working status	0.111

Table A2. - Parameter estimates: quadratic model

Covariate	$P_1(A\|X)$		$P_2(A\|X)$		$P_3(A\|X)$		$Q_1(A\|X)$		$Q_2(A\|X)$	
Age	-3.004	(0.112)	0.366	(0.065)	-2.474	(0..245)	-0.059	(0.480)	0.216	(0.233)
Age2	0.395	(0.016)	-0.021	(0.001)	0.352	(0..032)	0.099	(0.046)	0.020	(0.028)
Female	0.310	(0.103)	-0.308	(0.078)	-0.295	(0.260)	0.509	(0.225)	-0.947	(0.220)
Indigenous	0.275	(0.082)	0.523	(0.064)	0.208	(0.191)	-0.082	(0.179)	0.469	(0.172)
Log income	-0.707	(0.119)	-0.459	(0.094)	-1.182	(0.335)	-0.982	(0.282)	-0.159	(0.258)
Household size	-0.216	(0.039)	-0.221	(0.031)	-0.455	(0.106)	-0.407	(0.099)	0.015	(0.085)
No. of children under age 6	0.043	(0.051)	0.156	(0.038)	0.127	(0.124)	0.393	(0.102)	-0.042	(0.117)
No. of children over age 5	0.054	(0.039)	0.075	(0.033)	0.050	(0.111)	-0.007	(0.094)	-0.040	(0.081)
Female × no. young children	0.064	(0.048)	-0.064	(0.039)	0.053	(0.126)	-0.284	(0.108)	0.067	(0.103)
Uneducated father	1.066	(0.173)	0.349	(0.110)	0.814	(0.439)	1.451	(0.438)	0.021	(0.385)
Father primary education	0.551	(0.172)	0.270	(0.104)	0.534	(0.434)	1.194	(0.429)	0.275	(0.393)
Uneducated mother	1.290	(0.234)	0.400	(0.141)	4.668	(29.990)	0.789	(0.796)	0.781	(1.911)
Mother primary education	0.623	(0.235)	0.396	(0.136)	4.547	(29.989)	1.046	(0.789)	1.162	(1.917)
Collective shock	-0.161	(0.073)	0.181	(0.056)	0.153	(0.178)	-0.260	(0.165)	0.329	(0.151)
Individual shock	-0.038	(0.069)	0.276	(0.055)	0.422	(0.176)	0.292	(0.155)	-0.069	(0.146)
Credit rationed	0.351	(0.067)	-0.144	(0.051)	-0.110	(0.155)	-0.011	(0.144)	-0.258	(0.131)
Medical insurance	-0.050	(0.093)	-0.198	(0.074)	-0.092	(0.223)	-0.061	(0.219)	-0.220	(0.211)
North	-0.417	(0.183)	-0.197	(0.155)	0.323	(0.616)	0.033	(0.444)	-0.346	(0.367)
North-east	-0.314	(0.184)	-0.173	(0.155)	-0.991	(0.773)	0.487	(0.484)	0.191	(0.422)
South-east	-0.903	(0.191)	-0.245	(0.147)	-0.543	(0.675)	-0.231	(0.459)	0.194	(0.410)
Central	-0.668	(0.172)	0.244	(0.132)	0.260	(0.595)	0.380	(0.404)	0.039	(0.373)
South-west	-0.810	(0.174)	-0.291	(0.136)	-0.669	(0.620)	-0.559	(0.431)	-0.274	(0.360)
North-west	-0.493	(0.182)	-0.306	(0.146)	0.129	(0.613)	-0.091	(0.427)	-0.107	(0.350)
Peten	-0.645	(0.181)	-0.076	(0.144)	0.348	(0.611)	-0.001	(0.433)	0.195	(0.370)
Landowner	-0.322	(0.077)	0.512	(0.062)	-0.341	(0.188)	-0.164	(0.170)	0.197	(0.154)
Constant	5.147	(1.146)	0.139	(0.899)	3.755	(30.15)	1.537	(2.965)	-2.204	(2.992)

Sample: All individuals aged 7-14 (n = 7,936); ln L = -24,169.965; Akaike Information Criterion = 3.078

Table A2.2 - Parameter estimates: semi-parametric model

Covariate	$P_1(A\|X)$		$P_2(A\|X)$		$P_3(A\|X)$		$Q_1(A\|X)$		$Q_2(A\|X)$	
Female	0.292	(0.104)	-0.306	(0.079)	-0.258	(0.265)	0.506	(0.229)	-1.008	(0.221)
Indigenous	0.279	(0.083)	0.530	(0.065)	0.139	(0.191)	-0.062	(0.183)	0.502	(0.175)
Log income	-0.735	(0.120)	-0.448	(0.096)	-1.306	(0.331)	-0.982	(0.290)	-0.108	(0.261)
Household size	-0.223	(0.040)	-0.219	(0.032)	-0.470	(0.105)	-0.414	(0.102)	0.040	(0.086)
No. of children under age 6	0.037	(0.052)	0.163	(0.039)	0.116	(0.123)	0.398	(0.107)	-0.072	(0.117)
No. of children over age 5	0.055	(0.039)	0.075	(0.033)	0.007	(0.110)	-0.003	(0.097)	-0.042	(0.082)
Female × no. young children	0.067	(0.049)	-0.068	(0.039)	0.023	(0.126)	-0.278	(0.110)	0.090	(0.105)
Uneducated father	1.053	(0.174)	0.345	(0.112)	0.877	(0.469)	1.400	(0.434)	-0.016	(0.387)
Father primary education	0.537	(0.173)	0.268	(0.106)	0.627	(0.465)	1.132	(0.425)	0.237	(0.396)
Uneducated mother	1.265	(0.236)	0.389	(0.142)	4.953	(45.168)	0.744	(0.815)	0.733	(1.935)
Mother primary education	0.591	(0.236)	0.391	(0.138)	4.884	(45.168)	1.018	(0.808)	1.097	(1.941)
Collective shock	-0.163	(0.074)	0.187	(0.057)	0.163	(0.177)	-0.261	(0.168)	0.322	(0.153)
Individual shock	-0.036	(0.070)	0.281	(0.056)	0.417	(0.176)	0.302	(0.159)	-0.055	(0.148)
Credit rationed	0.348	(0.067)	-0.147	(0.052)	-0.110	(0.155)	-0.024	(0.147)	-0.243	(0.133)
Medical insurance	-0.038	(0.093)	-0.199	(0.075)	-0.097	(0.225)	-0.090	(0.227)	-0.179	(0.214)
North	-0.449	(0.184)	-0.200	(0.158)	0.279	(0.639)	-0.041	(0.451)	-0.304	(0.373)
North-east	-0.314	(0.184)	-0.177	(0.157)	-1.153	(0.808)	0.383	(0.491)	0.305	(0.430)
South-east	-0.909	(0.192)	-0.247	(0.149)	-0.635	(0.701)	-0.322	(0.462)	0.361	(0.415)
Central	-0.690	(0.173)	0.242	(0.134)	0.233	(0.621)	0.294	(0.406)	0.099	(0.380)
South-west	-0.810	(0.174)	-0.291	(0.137)	-0.704	(0.646)	-0.681	(0.437)	-0.190	(0.365)
North-west	-0.513	(0.183)	-0.306	(0.148)	0.114	(0.638)	-0.181	(0.432)	-0.037	(0.357)
Peten	-0.664	(0.183)	-0.073	(0.146)	0.344	(0.632)	-0.093	(0.437)	0.189	(0.380)
Landowner	-0.317	(0.078)	0.520	(0.063)	-0.356	(0.186)	-0.140	(0.176)	0.215	(0.155)

Constant	3.180	(1.170)	1.240	(0.935)	3.948	(45.276)	6.145	(2.784)	-0.315	(3.019)
Age = 7	-0.287	(0.185)	-0.758	(0.209)	-1.517	(0.285)				
Age = 8	-3.941	(0.310)	-0.613	(0.210)	-4.320	(0.540)	-4.628	(0.748)	-1.737	(0.307)
Age = 9	-3.035	(0.269)	-0.805	(0.214)	-3.344	(0.420)	-3.570	(0.385)	-2.002	(0.323)
Age = 10	-2.684	(0.305)	0.510	(0.209)	-2.769	(0.470)	-2.868	(0.311)	-0.617	(0.306)
Age = 11	-1.508	(0.232)	-0.097	(0.215)	-2.122	(0.381)	-2.408	(0.267)	-1.406	(0.351)
Age = 12	-1.508	(0.232)	-0.097	(0.215)	-2.122	(0.381)	-1.254	(0.242)	-0.120	(0.327)

Sample: All individuals aged 7-14 (n = 7,936); ln L = -23,725.649; Akaike Information Criterion = 3.026

Table A2.2 - Parameter estimates: semi-parametric model

Covariate	$P_1(A\|X)$		$P_2(A\|X)$		$P_3(A\|X)$		$Q_1(A\|X)$		$Q_2(A\|X)$	
Female	0.292	(0.104)	-0.306	(0.079)	-0.258	(0.265)	0.506	(0.229)	-1.008	(0.221)
Indigenous	0.279	(0.083)	0.530	(0.065)	0.139	(0.191)	-0.062	(0.183)	0.502	(0.175)
Log income	-0.735	(0.120)	-0.448	(0.096)	-1.306	(0.331)	-0.982	(0.290)	-0.108	(0.261)
Household size	-0.223	(0.040)	-0.219	(0.032)	-0.470	(0.105)	-0.414	(0.102)	0.040	(0.086)
No. of children under age 6	0.037	(0.052)	0.163	(0.039)	0.116	(0.123)	0.398	(0.107)	-0.072	(0.117)
No. of children over age 5	0.055	(0.039)	0.075	(0.033)	0.007	(0.110)	-0.003	(0.097)	-0.042	(0.082)
Female × no. young children	0.067	(0.049)	-0.068	(0.039)	0.023	(0.126)	-0.278	(0.110)	0.090	(0.105)
Uneducated father	1.053	(0.174)	0.345	(0.112)	0.877	(0.469)	1.400	(0.434)	-0.016	(0.387)
Father primary education	0.537	(0.173)	0.268	(0.106)	0.627	(0.465)	1.132	(0.425)	0.237	(0.396)
Uneducated mother	1.265	(0.236)	0.389	(0.142)	4.953	(45.168)	0.744	(0.815)	0.733	(1.935)
Mother primary education	0.591	(0.236)	0.391	(0.138)	4.884	(45.168)	1.018	(0.808)	1.097	(1.941)
Collective shock	-0.163	(0.074)	0.187	(0.057)	0.163	(0.177)	-0.261	(0.168)	0.322	(0.153)
Individual shock	-0.036	(0.070)	0.281	(0.056)	0.417	(0.176)	0.302	(0.159)	-0.055	(0.148)
Credit rationed	0.348	(0.067)	-0.147	(0.052)	-0.110	(0.155)	-0.024	(0.147)	-0.243	(0.133)
Medical insurance	-0.038	(0.093)	-0.199	(0.075)	-0.097	(0.225)	-0.090	(0.227)	-0.179	(0.214)
North	-0.449	(0.184)	-0.200	(0.158)	0.279	(0.639)	-0.041	(0.451)	-0.304	(0.373)
North-east	-0.314	(0.184)	-0.177	(0.157)	-1.153	(0.808)	0.383	(0.491)	0.305	(0.430)
South-east	-0.909	(0.192)	-0.247	(0.149)	-0.635	(0.701)	-0.322	(0.462)	0.361	(0.415)
Central	-0.690	(0.173)	0.242	(0.134)	0.233	(0.621)	0.294	(0.406)	0.099	(0.380)
South-west	-0.810	(0.174)	-0.291	(0.137)	-0.704	(0.646)	-0.681	(0.437)	-0.190	(0.365)
North-west	-0.513	(0.183)	-0.306	(0.148)	0.114	(0.638)	-0.181	(0.432)	-0.037	(0.357)
Peten	-0.664	(0.183)	-0.073	(0.146)	0.344	(0.632)	-0.093	(0.437)	0.189	(0.380)

Landowner	-0.317	(0.078)	0.520	(0.063)	-0.356	(0.186)	-0.140	(0.176)	0.215	(0.155)
Constant	3.180	(1.170)	1.240	(0.935)	3.948	(45.276)	6.145	(2.784)	-0.315	(3.019)
Age = 7	-0.287	(0.185)	-0.758	(0.209)	-1.517	(0.285)				
Age = 8	-3.941	(0.310)	-0.613	(0.210)	-4.320	(0.540)	-4.628	(0.748)	-1.737	(0.307)
Age = 9	-3.035	(0.269)	-0.805	(0.214)	-3.344	(0.420)	-3.570	(0.385)	-2.002	(0.323)
Age = 10	-2.684	(0.305)	0.510	(0.209)	-2.769	(0.470)	-2.868	(0.311)	-0.617	(0.306)
Age = 11	-1.508	(0.232)	-0.097	(0.215)	-2.122	(0.381)	-2.408	(0.267)	-1.406	(0.351)
Age = 12	-1.508	(0.232)	-0.097	(0.215)	-2.122	(0.381)	-1.254	(0.242)	-0.120	(0.327)

Sample: All individuals aged 7-14 (*n* = 7,936); ln *L* = -23,725.649; Akaike Information Criterion = 3.026

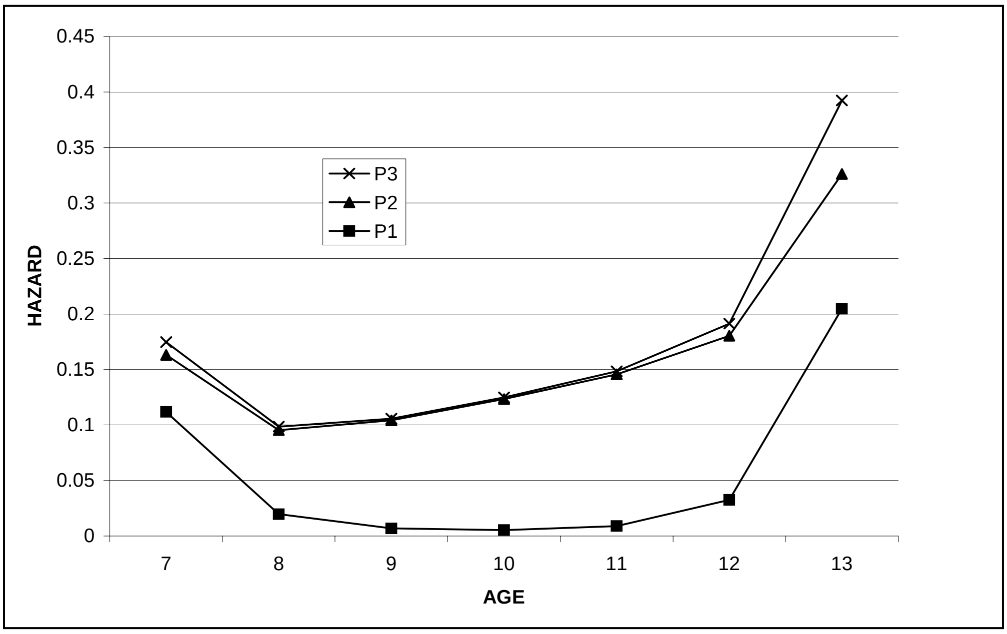


Figure A.1a Hazard functions for non-working schoolchildren (quadratic model)

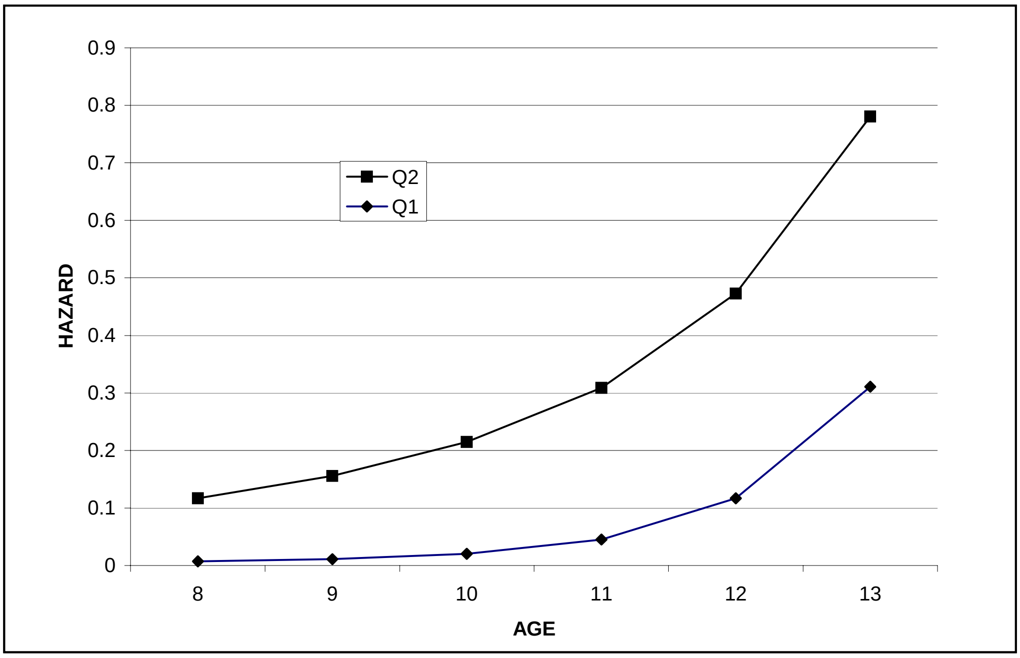


Figure A.2a Hazard functions for working schoolchildren and 'idle' children (quadratic model)

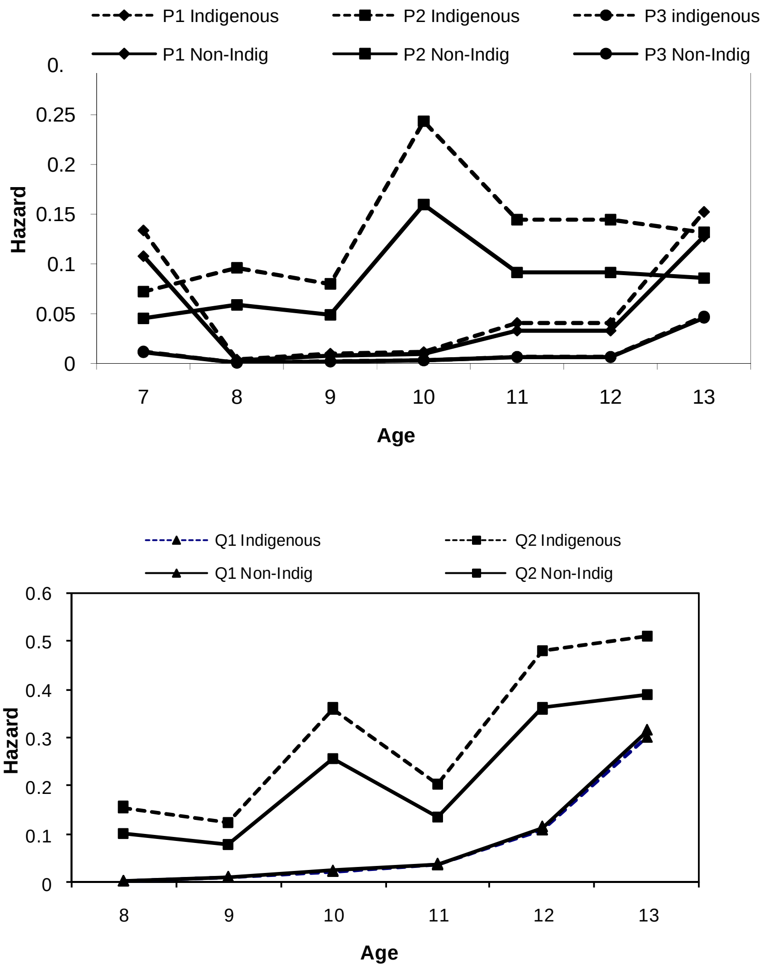


Figure A2.1 - Hazards by ethnicity

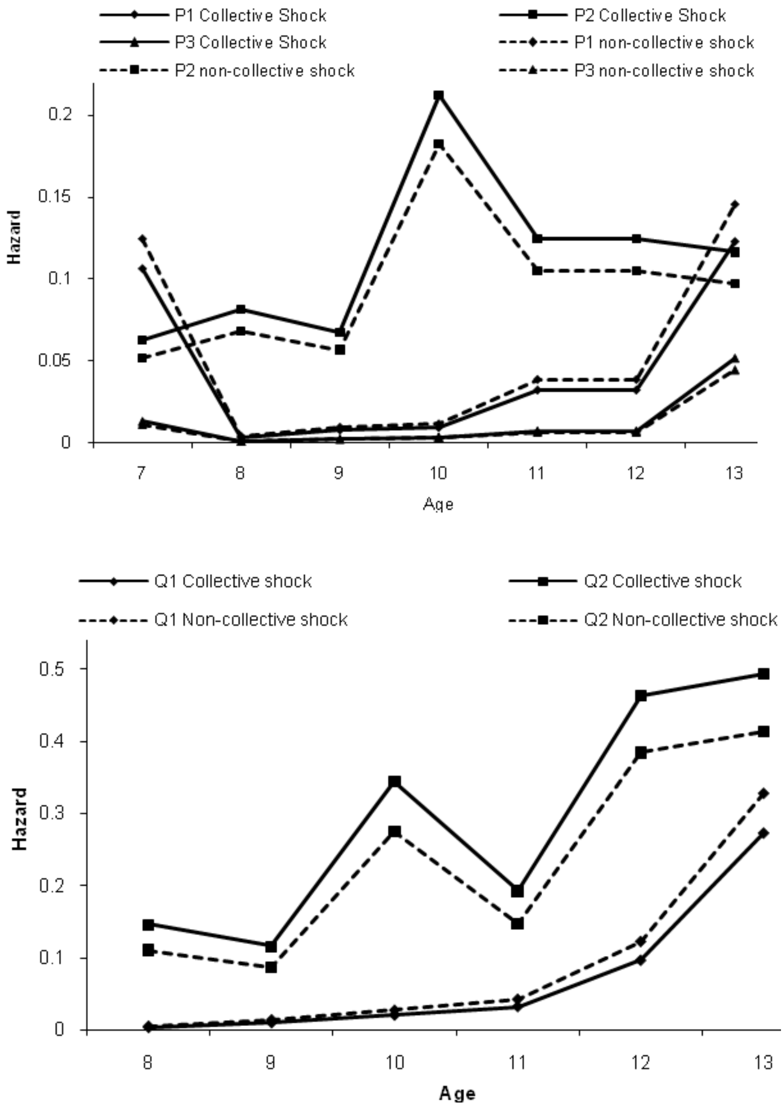


Figure A2.2 - Hazards by occurrence of collective shocks

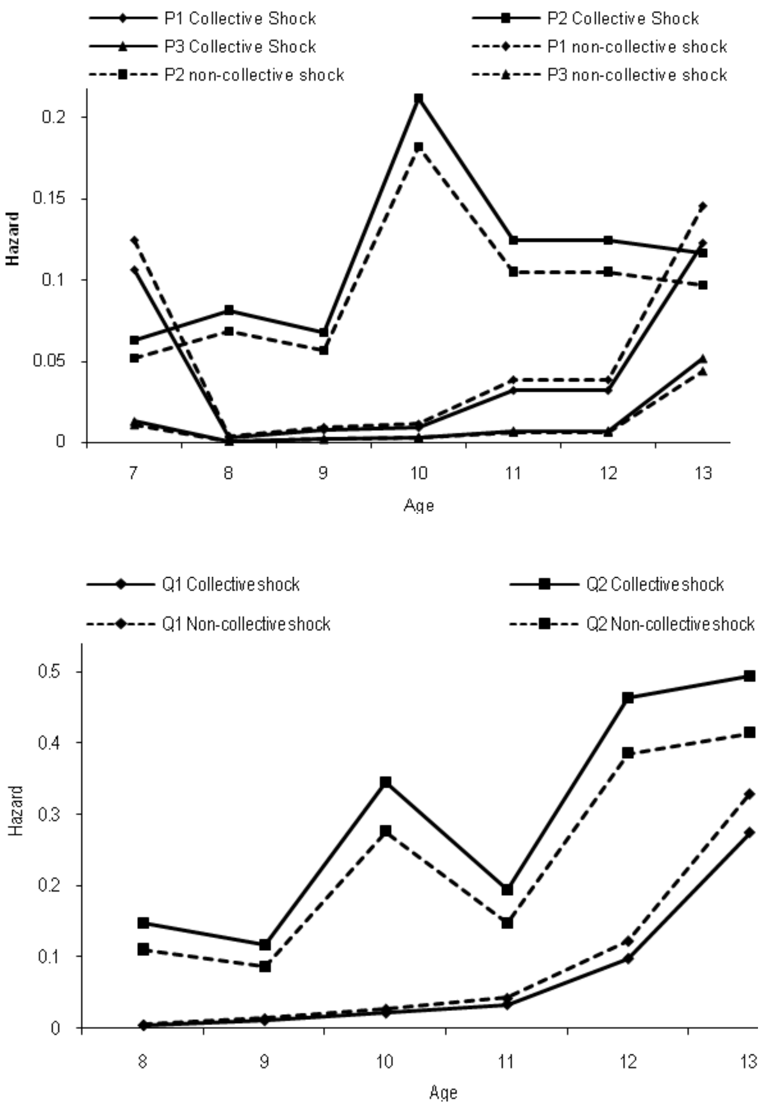
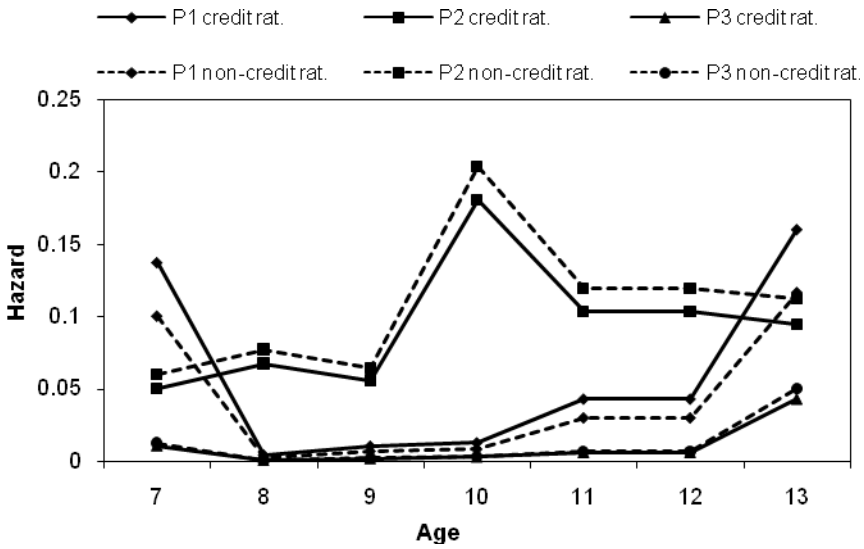


Figure A2.3 - Hazards by occurrence of individual shocks



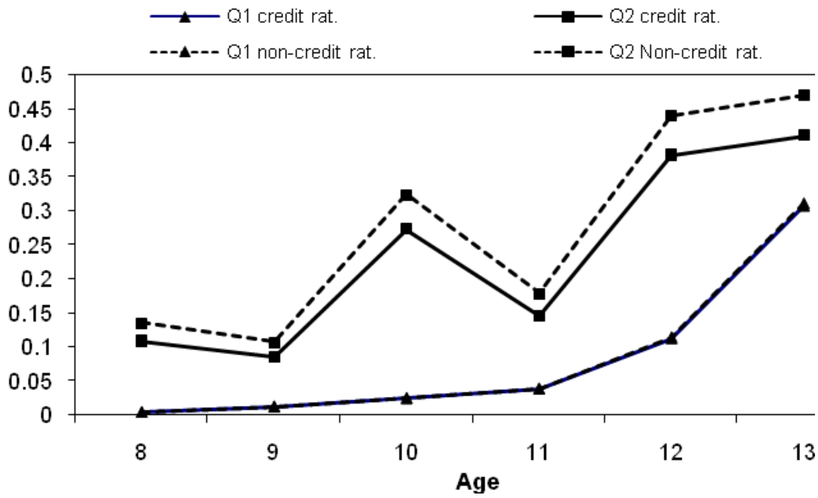


Figure A2.4 Hazards by experience of credit rationing

Figure A2.5 - Hazards by levels of father's education

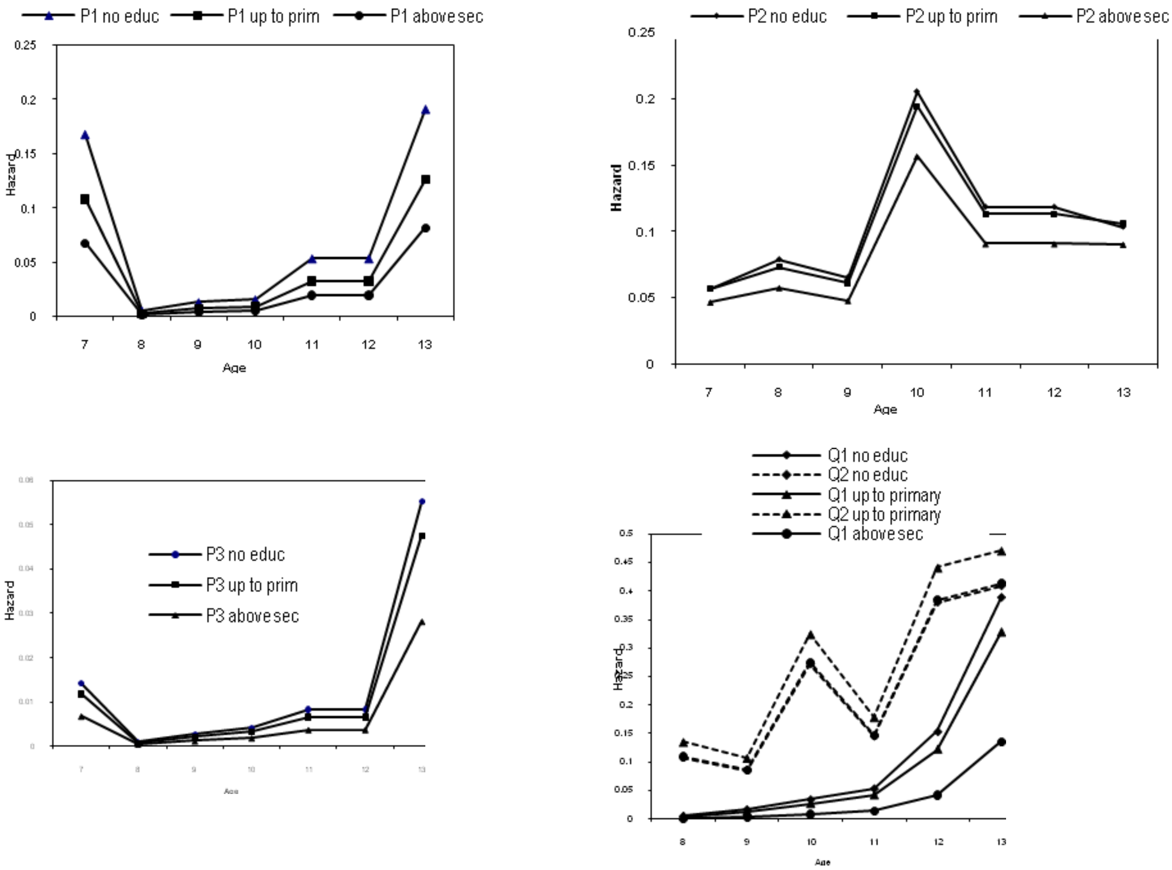


Table A2.3. - Numbers of children predicted to leave school or commence child labor, and numbers with high (*v* = 0.4) vulnerability, compared to school/work status in previous year (see Figures 6a-d in text).

	still at school	not attending school	predicted addition in coming year	Highly vulnerable to school drop-out	not working	already working	predicted addition in coming year	highly vulnerable to child labor
Boys age 8	127367	51156	686	0	160179	18,344	18297	400
Boys age 9	128991	28347	1909	0	130728	26610	12604	0
Boys age 10	136670	20823	2737	0	123186	34307	28649	14964
Boys age 11	143059	31348	7175	0	118215	56192	17991	2982
Boys age 12	110878	23333	14349	2286	76512	57699	13964	9227
Boys age 13	139087	26784	35763	28627	82910	82960	12962	7835
Girls age 8	106030	57654	657	0	151972	11712	11361	
Girls age 9	129674	38700	2293	0	152318	16056	9553	0
Girls age 10	116524	29538	2931	0	118689	27373	21795	1091
Girls age 11	127900	26271	8236	0	122605	31566	12812	0
Girls age 12	89844	24096	7369	179	83173	39767	11265	2295
Girls age 13	111981	30977	30389	28604	91756	51202	9751	3447

Table A2.4. - Numbers of children predicted to leave school or commence child labor, and numbers with high (*v* = 0.2) vulnerability, compared to school/work status in previous year.

	still at school	not attending school	predicted addition in coming year	highly vulnerable to school drop-out	not working	already working	predicted addition in coming year	highly vulnerable to child labor
Boys age 8	127367	51156	686	0	160179	18,344	18297	24209
Boys age 9	128991	28347	1909	0	130728	26610	12604	7375
Boys age 10	136670	20823	2737	0	123186	34307	28649	63282
Boy age 11	143059	31348	7175	795	118215	56192	17991	35396
Boys age 12	110878	23333	14349	13157	76512	57699	13964	23400
Boys age 13	139087	26784	35763	80745	82910	82960	12962	33528
Girls age 8	106030	57654	657	0	151972	11712	11361	2124
Girls age 9	129674	38700	2293	0	152318	16056	9553	679
Girls age 10	116524	29538	2931	0	118689	27373	21795	52068
Girls age 11	127900	26271	8236	3016	122605	31566	12812	7090
Girls age 12	89844	24096	7369	10994	83173	39767	11265	17040
Girls age 13	111981	30977	30389	67669	91756	51202	9751	28658