International COCOA Initiative

Using community-level data to understand child labour risk in cocoa-growing areas in Ghana

Research Report

May 2019

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

Lindt COCOA FOUNDATION

CHOCOLONELY foundation

ACKNOWLEDGEMENTS

This research study was prepared by Lorenzo Guarcello and Furio Rosati of the Understanding Children's Work (UCW) Programme, a joint research initiative by the International Labour Organization (ILO), the United Nations Children's Fund (UNICEF), and the World Bank.

Data collection in Ghana was collected by the *Institute of Statistical Social and Economic Research (ISSER)* and was funded jointly by the Chocoloney Foundation, the Lindt Foundation, and several ICI Board Members.

This is one of two studies commissioned by the International Cocoa Initiative to develop country-specific child labour risk models for Côte d'Ivoire and Ghana, based on community-level data. An executive summary of the findings of both studies is available on the ICI Knowledge Hub.

CONTENTS

EXECUTIVE SUMMARY

Monitoring the child labour situation at community level and identifying the communities more exposed to the risk of child labour is a complex task. On the one hand disaggregated and timely information at community level are essential for effective targeting of intervention. On the other hand, obtaining reliable estimates of child labour at community level would require to sample a relatively large number of individuals in each community. This would be complex in terms of sample design, costly and time consuming. Information on community characteristics, however, can be collected relatively easily and quickly and utilized to build a child labour "risk" indicator that can be computed and updated with relative ease.

The study presents a "risk" indicator for the presence of child labour at community level in Ghana. In particular, it aims at identifying classes of risk of child labour based on community characteristics. Different techniques are available to support the identification of community level classes of risk of the presence of child labour. The study utilized several tests and attempts to adopt techniques that do not establish a priori the number of classes, i.e. Latent Class or Finite Mixture models. However, the heterogeneity in terms of community-level child labour incidence is not substantial enough to apply these methods. Therefore, a different approach was followed. The sample was divided in three terciles on the basis of child labour incidence, constituting a priori three child labour risk classes. A multinomial logit model was then estimated to predict the probability of belonging to any of the three risk classes.

The study uses data on children's activities collected through an ad-hoc survey (Ghana Child Labour Prevalence Survey, 2017) and links them to a set of indicators observed at community level (Protective Cocoa Community Framework PCCF, 2017). In this way it is possible to build a risk indicator based on community characteristics only, when direct observation of the involvement of children in child labour is not feasible.

Results show that an additional primary school in the community increases the probability of belonging to the low risk class by 3 percentage points. Moreover, the probability of belonging to the low risk class is on average 27 percentage points higher for communities with connection to mobile network than for communities without connection. As the distance to the nearest senior secondary high school increases by one unit, the probability to be in the low risk class decreases by about 2 percentage points. However, communities with junior secondary high school are less likely to be in this class by 37 percentage points with respect to communities without the school.

The probability of belonging to the medium risk class increases by 3 percentage points for each additional primary school and by about 30 percentage points for communities with junior secondary high school. Communities with an accessible road throughout the year have a probability to be in the medium risk class lower by 37 percentage points than communities without an accessible road.

Turning to the high risk class, an additional primary school in the community decreases the probability to be in this class by 7 percentage points, while as the distance to Senior Secondary High School increases by one unit, the probability to be in this class increases by about 2 percentage points. Communities are less likely to be in the high risk class, by 42 percentage points, if there is a connection to the mobile network. Finally, the availability of casual adult labour increases the probability of belonging to the high risk class by 23 percentage points, indicating possible complementarities between child and adult work and/or high level of dependency on agriculture.

1 INTRODUCTION

Monitoring the child labour situation at community level and identifying the communities more exposed to the risk of child labour is a complex task. On the one hand disaggregated and timely information at community level are essential for effective targeting of intervention. On the other hand, obtaining reliable estimates of child labour at community level would require to sample a relatively large number of individuals in each community. This would be complex in terms of sample design, costly and time consuming. Information on community (potentially) relevant characteristics, however, can be collected relatively easily and quickly and utilized to build a child labour "risk" indicator that can be applied and updated with relative ease.

Moreover, as we shall discuss in more detail in the text, it is often more relevant and efficient to classify the communities in different risk groups rather than try to "predict" the specific incidence rate for each community. In the case considered, in fact, the problem lies in the development of a methodology allowing the inference of child labour risk at the community level in Ghana from a set of information collected at community level, *without* the direct observation of involvement of children in child labour.

In this study, we use data on children's activities collected through an ad-hoc survey and link them to a set of indicators observed at the community level. In this way we are able to build a risk indicator that can then be utilized on the basis of community level data, without the direct observation of the involvement of children in child labour.

Different techniques are available to support the identification of community level classes of risk of the presence of child labour. We carried out several tests and attempts to adopt techniques that do not establish a priori the number of classes, i.e. Latent Class or Finite Mixture models. However, the heterogeneity in terms of community-level risk was not substantial enough to apply these methods. Therefore, we decided to divide the sample in three terciles on the basis of child labour incidence, constituting a priori three risk classes: low risk, medium risk and high risk of child labour. We then estimate a multinomial logit model, to predict the probability of belonging to any of the three risk classes.

The rest of the paper is organized as follows. Section 2 illustrates the data used and presents the descriptive statistics. Section 3 discusses the risk class approach and introduces the econometric methodology. The results are discussed in section 4. In section 5, determinants of child

employment and child labour are analysed at the individual level. Conclusions are discussed in section 6.

2 DATA AND DESCRIPTIVE STATISTICS

The present study makes use of two primary data sources: (i) the Child Labour Prevalence Survey 2017; and (ii) the Protective Cocoa Community Framework 2017 (PCCF 2017).

The Child Labour Prevalence Survey was carried out in 2017 and was specifically designed to measure incidence and the characteristics of child labour in the cocoa growing areas of Ghana. It collected information at the individual and household levels.

The data collection process included: (i) the development of a sampling strategy; (ii) the development of the data collection instruments (questionnaires); and (iii) and the data collection phase, conducted by the Institute of Statistical Social and Economic Research (ISSER).

The sample was drawn following a two-stage stratified sample design approach. The sampling frame covered the 6 regions where cocoa is cultivated: Ashanti, Western, Central, Volta, Eastern and Brong-Ahafo. In the first-stage, a sample of enumeration areas (EA) was extracted, leading to a selection of 128 cocoa-growing communities (EAs).

In the second-stage, a fixed number of households were randomly selected within each sampled community, leading to a final sample of about 3,000 households, belonging to 128 communities (EAs) and 6,300 children were interviewed.

A survey based on two separate questionnaires was administered to the selected households. The household questionnaire, aimed at collecting background information on the household economic activities (agricultural and non-agricultural), on the occurrence of shocks and on the dwelling conditions. The individual level questionnaire aimed at collecting information on education, employment, unemployment and decisional power within the household.

At the same time, the ICI Protective Cocoa Community Framework Questionnaire (PCCF) was fielded. The PCCF is a community assessment tool, incorporating key indicators and proxies related to community development, community empowerment, education, child protection, gender and livelihoods in cocoa-growing communities in Ghana and Côte d'Ivoire. The PCCF survey in Ghana was conducted in 2017 in the same communities selected for the Child Labour Prevalence Survey. It was administered to key actors in the community depending on the specific sections of the questionnaire:

community leaders, community child protection committee members, community women's group, children, farmers and other organizations, school teachers and directors.

The data from the household survey, averaged at community level, were matched with the PCCF data to identify the community characteristics most relevant to the construction of the risk indicator.

On the basis of the individual data, we computed the average incidence of children's employment and child labour at the community level. The definitions of child employment and child labour are discussed in the next section, where we present also the community level characteristics used for the empirical analysis.

Child employment and child labour in the study communities

In this section we present the main characteristics of child employment and child labour in the 128 communities on the basis of the individual questionnaire.

The definition of child employment and child labour we use is conditioned by the structure of the questionnaire developed by the ISSER. The following table, reporting the relevant questions, helps to illustrate the information that can be obtained from the survey and how it is used to define child employment and child labour.

Table 1. **Questions used for child employment**

Question 1	Question 2	Question 3	Question 4	Question 5
Did you engage in any work for at least one hour during the last 7 days? (as employee, self-employed, employer or unpaid family work)	During the last 7 days, did you do any of the following activities, even for only one hour?	Even though you did not do any of these activities in the past week, do you have a job, business, or other economic or farming activity that you will definitely return to?	In what sector is this main activity?	Work in agriculture involves work on your own or the household's plot or any other farm, food garden, or help in growing farm produce or in looking after animals for the household. Did you work in agriculture during the last 7 days?

Each individual was asked whether she/he has been working for at least one hours in the past seven days (Question 1 and Question 2) and, if not, whether she/he has a job to return to (Question 3). If the individual responded "yes" to at least one of the first three questions, it was asked the sector of employment (Question 4). Then, a further probe was added specifying more in details the activities that can be considered as agricultural activities (Question 5). Questions concerning the hazards faced and the hours worked were asked only to the individuals who answered "yes" to Question 5[1] and this substantially limits the possibility of identifying child labourers, as we shall discuss below.

Given the structure of the questionnaire, we have defined a child to be in employment if she/he answered "yes" to at least one of the first three questions or, having answered no to all the three questions, if subsequently she/he answered "yes" to Question 5. Note that, following an internationally agreed convention in child labour statistics,[2] we did not consider fetching water as an economic activity.

The definition is more complex for child labour, as the questions on hazardous occupations, experienced injuries, dangerous tools used at work and hours worked (required to define child labour) were asked only to children who answered "yes" to Question 5. Therefore, child labour could be defined only for this subset of children, and can only be expressed as percentage of children working in agriculture and not as a percentage of all children in the relevant age group. Following the international standard definition, children are classified as in child labour on the basis of the following criteria: children aged 5-11 years in agriculture, children aged 12-14 years working in hazardous occupations or working from 14 to 42 hours per week and children aged 15-17 years working in hazardous occupations or working more than 42 hours per week. Hazardous occupations in turn include: children working in hazardous activities, children that experienced injuries at work and children that used dangerous tools.[3]

[1] About 1,410 children replied "yes" to Question 5.
[2] See 18th International Conference of Labour Statisticians, Resolution concerning statistics of child labour, and the appendix for a detailed description of the definition of Child Labour.
[3] Additional detail on the survey questions used to define child employment and child labour are reported in the Appendix.

Child employment

The main characteristics of child employment, disaggregated by gender and by age groups, are reported in Table 2.

As Table 2 shows, 31.91% of children aged 5-17 are in employment. Children's involvement in employment is markedly higher in rural areas (34.41%) than in urban areas (22.85%) and among older children aged 14-17 (45.66%) compared to younger children aged 5-13 (27.48%). Child employment differs by gender: 33.32% of boys, aged 5-17, are in employment as compared to 30.35% of girls in the same age range. The gender gap increases with age and is larger in rural areas.

As reported in Table 3, the majority of employed children are involved in unpaid family work (77%). The percentage of children in wage employment is small (9%).

Table 2. **Children in employment by region, age and sex**

	Residence		Age group		
	Urban	Rural	5-13	14-17	5-17
All	132,011	718,754	554,437	296,328	850,765
Boys	70,436	394,572	287,233	177,775	465,008
Girls	61,575	324,182	267,204	118,553	385,757
% of child population					
All	22.85	34.41	27.48	45.66	31.91
Boys	23.71	35.92	27.59	50.14	33.32
Girls	21.93	32.74	27.36	40.26	30.35

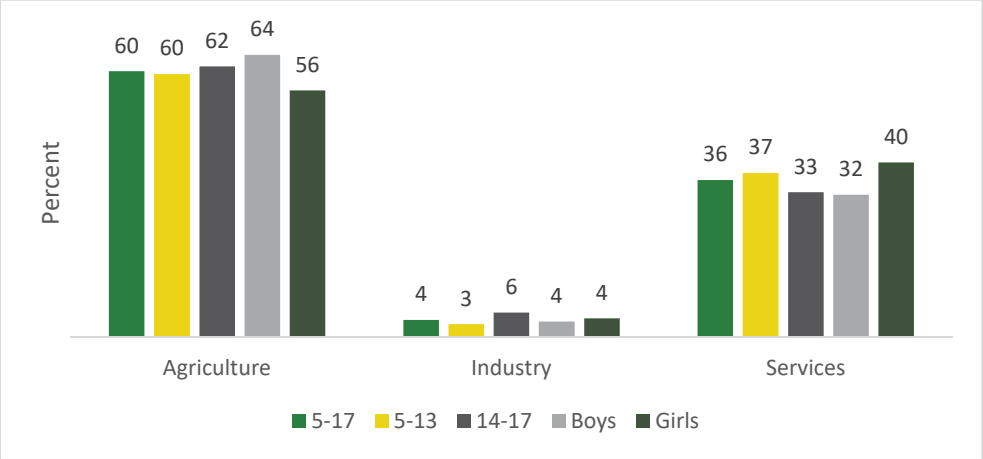
Table 3. **Child employment status, by age and sex (% of child population)**

	Wage employee	Self-employed	Contributing family worker	Other
5-17	9.18	0.56	76.86	13.40
5-13	7.99	0.30	75.53	16.18
14-17	11.40	1.06	79.34	8.19
Boys	11.95	0.65	72.09	15.30
Girls	5.84	0.46	82.60	11.10

Figure 1 reports the distribution of working children by sector of employment, age and sex. About 60% of children aged 5-17 work in the agricultural sector and 36% are in the service sector. The

number of children working in the industry sector is negligible (4%). There are no significant differences in terms of sector of employment between younger and older children. However, there are some differences between boys and girls: boys are more likely to work in the agricultural sector (64%) than girls (56%), while girls are more likely to work in the service sector (40%) compared to boys (32%).

Figure 1. **Sector of employment by age and sex (% share of total employment)**



As far as the time intensity of work is concerned, the average number of weekly hours worked is provided only for children who answered "yes" to Question 5 and reported in the following table. Overall, working children aged 5-17 work on average almost 8 hours per week, with similar values reported across age groups and by gender.

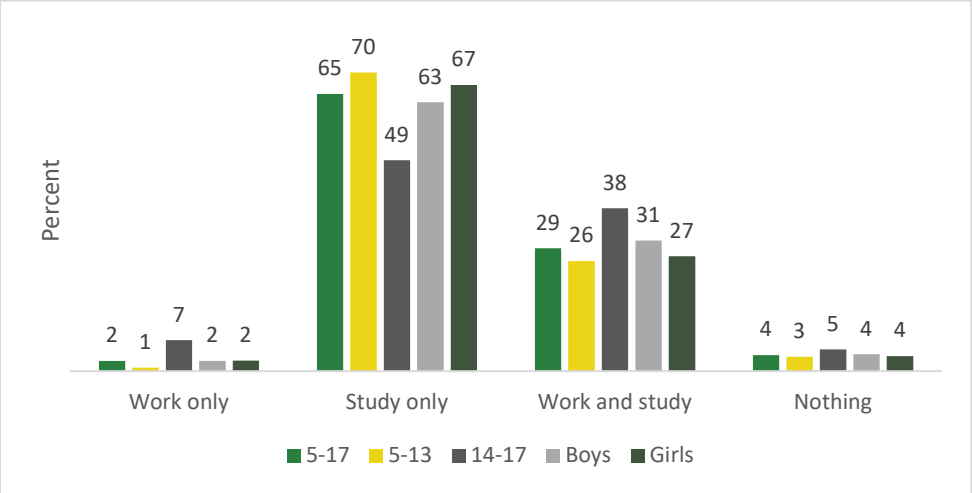
Table 4. **Weekly hours worked in agriculture by age, sex and sector of employment**

	Agricultural sector
5-17	7.8
5-13	7.5
14-17	8.1
Boys	7.9
Girls	7.5

Children's distribution across four mutually exclusive activity categories (i.e., work only, study only, work and study, nothing) is reported in Figure 2. As shown, the distribution across the four categories differs considerably by age group: 7% of children aged 14-17 work without attending school, compared to 1% of children aged 5-13. On the other hand, younger children are more likely to attend school without working (70%) than their older counterparts (49%). About 26% of

the younger children and 38% of older children combine school and work. The share of children neither attending school nor working is quite low for both age ranges and for both boys and girls[4].

Figure 2. **Type of activity by age and sex (% of child population)**



Child labour

As mentioned above, we can identify child labourers only within the subgroup of those who answered to Question 5. These children are all working in agriculture, but might not be necessarily representative of all children working in agriculture[5]. Nonetheless, with this caveat in mind, we will refer to them as children working in agriculture to simplify exposition.

As shown in Table 5, 88.04% of children working in agriculture are in child labour. Children in rural areas (88.49%) are more likely to be involved in child labour than their peers in urban areas (85.46%).[6] There are no remarkable differences between boys and girls. The child labour rate is significantly higher among younger children (91.89%) than older children (82.35). 0 shows that the majority of children in child labour are involved in non-wage activities (82.56%). The share of children who are employees (9.15%) is relatively small.

[4] For a discussion on children neither working nor studying see Biggeri M., Rosati F., Lyos S. Guarcello, L. (2003). "The puzzle of 'idle' children: neither in school nor performing economic activity: evidence from six countries" UCW working paper series, at www.ucw-project.org.

[5] Follow up questions to better capture working children were not asked after Question 5.

[6] The urban-rural definition here is based on classification of the settlement by respondents in the household survey.

Table 5. **Children in child labour by region, age and sex**

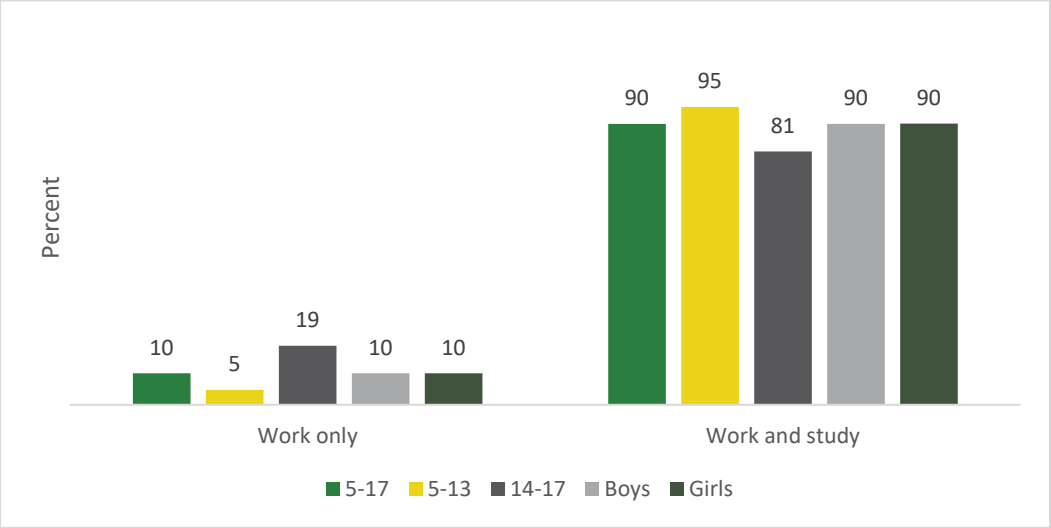
	Residence		Age group		
	Urban	Rural	5-13	14-17	5-17
All	64,149	381,667	277,413	168,403	445,816
Boys	34,660	222,952	151,663	105,949	257,613
Girls	29,489	158,714	125,749	62,454	188,203
% of child population in agriculture					
All	85.46	88.49	91.89	82.35	88.04
Boys	82.72	88.13	90.29	83.48	87.36
Girls	88.92	89.00	93.90	80.50	88.99

Table 6. **Child labour status, by age and sex**

	Wage employee	Self-employed	Contributing family worker	Other
5-17	9.15	0.12	82.44	8.29
5-13	8.32	0.10	82.64	8.94
14-17	10.53	0.14	82.11	7.22
Boys	10.89	0.18	79.42	9.51
Girls	6.77	0.04	86.57	6.62

Figure 3 shows that the majority of children in child labour attend school, independently of the gender. In fact, only 10% of children aged 5-17 are involved in child labour without attending school. The share of children only in child labour is lower for the younger group (5%) compared to the older group (19%).

Figure 3. **Type of activity by age and sex**



3 ECONOMETRIC APPROACH

A Risk class approach

Figure 4, which plots the child employment rate by community together with the standard errors, highlights that most communities have very similar level of child employment rates, which are statistically indistinguishable from each other. This can be noticed also looking at Figure 5, showing that the variability is quite small and that 75% of communities have an incidence of child employment lower than 40%: a very narrow band considering also the standard errors.

Owing to these characteristics, and also on the basis of some preliminary testing, we decided to identify different classes of child labour risk for the communities rather than to predict the expected incidence rate by single community. In other words, our approach seeks to identify the different classes of risk to which the communities belong and develop an econometric model able to predict the class membership (e.g. high risk, medium risk, low risk) of each community.

Figure 4. **Mean and SD of child employment in the study communities**

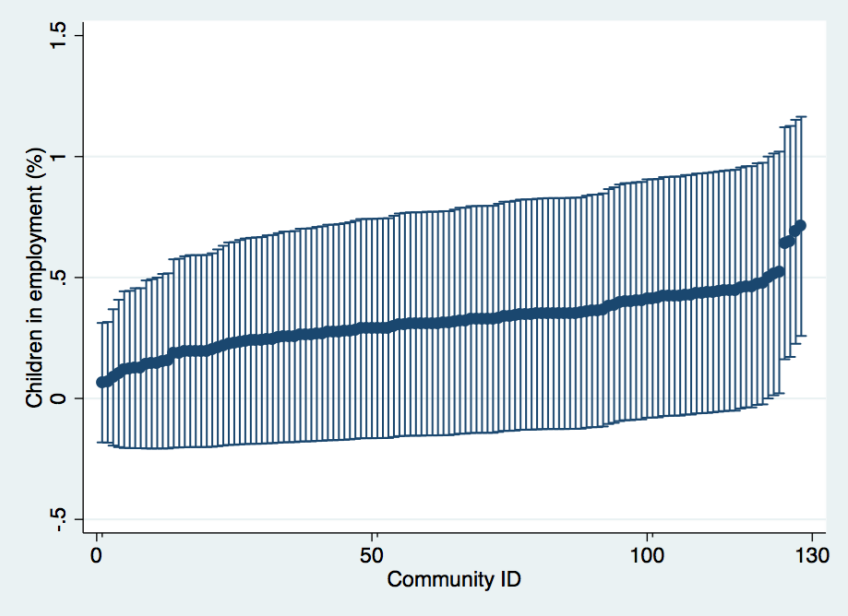
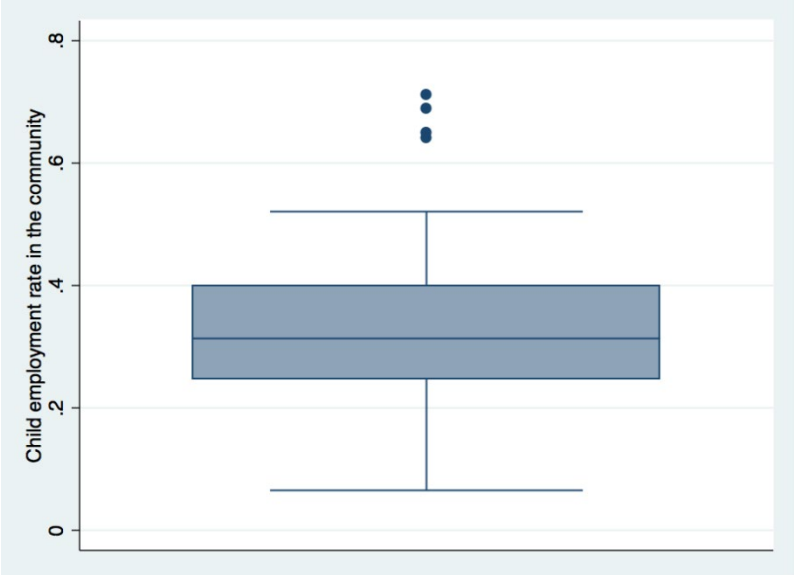
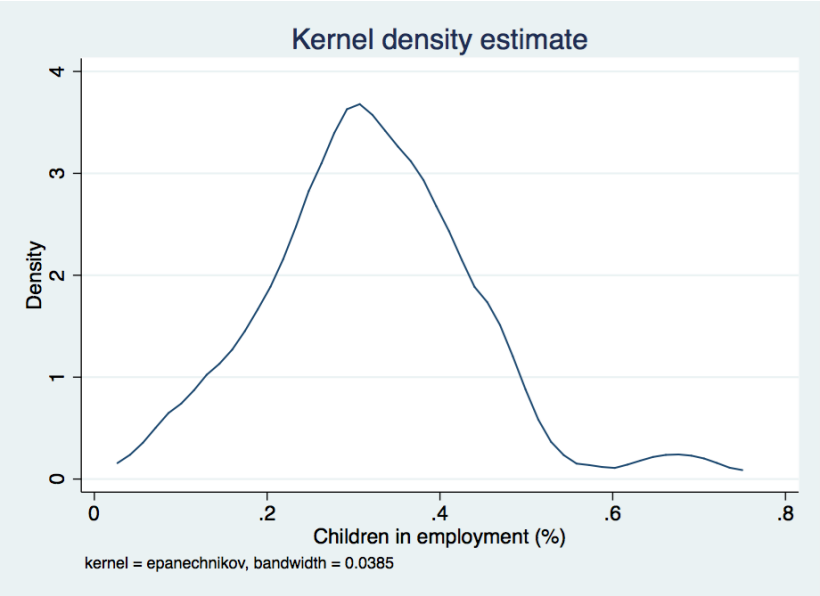


Figure 5. **Box plot of child employment**



We have first followed the finite mixture model approach (McLachlan and Peel 2000), adopted also to identify the classes of child labour risk in a companion paper relative to Côte d'Ivoire. However, given the low level of heterogeneity in the data, the model allows to identify only one latent class, i.e. one class of child labour risk. This is not surprising, especially if we look at the density of the child employment incidence, as approximated by the kernel density presented in Figure 6. It is easy to see that the density has one peak, indicating the low level of heterogeneity in the distribution of child employment incidence and supporting the idea that child employment rates are largely concentrated around the mean.

Figure 6. **Kernel density of child employment**



Ad hoc class identification

Given that the data driven approach does not allow us to identify different risk classes, we have followed an ad hoc approach and subdivided the sample in three terciles on the base of the community incidence of child employment.

The three terciles represent three class of risk of child employment (and child labour): from the lowest to the highest. We then estimate a multinomial logit model, to identify and, subsequently predict, the probability of belonging to one of the three risk class of child labour.

The dependent variable, y_i, can take three values: 1 if the $i-th$ community belongs to the first tercile (low risk), 2 if the $i-th$ community belongs to the second tercile (medium risk), 3 if the $i-th$

community belongs to the third tercile (high risk). The probability of each $i-th$ community to be in the $j-th$ risk class is given by:

$$p_{ij} = \Pr[y_i = j] = \frac{e^{x_i' \beta_j}}{\sum_{l=1}^{m} e^{x_i' \beta_l}}, \quad j = 1, \dots, m \quad (1)$$

Since $\sum_{j=1}^{m} p_{ij} = 1$, the restriction needed to identify the model is $\beta_1 = 0$ (Cameron and Trivedi 2005), i.e. the first risk class is the reference risk class. Therefore, the coefficients β_2 and β_3 measure the change relative to the first class of risk of child employment. The estimation of the parameters of interest is conducted through the maximum likelihood method. Once the multinomial logit model is estimated, the predicted probabilities of belonging to the three class of risk can be computed and we assign each community to the class of risk with the highest predicted probability.

As a robustness check of risk class identification, we also estimate a multinomial logit model with three classes computed considering the four quartiles of child employment. In particular, the first class includes communities in the first quartile (bottom 25% of the sample), the second class includes communities in the second and in the third quartiles (central 50% of the sample) and the third class includes communities in the fourth quartiles (top 25% of the sample). The three classes represent respectively low, medium and high risk of child labour. We have considered the three groups, based on the quartiles, to focus on the tails of the child employment distribution, given the low level of heterogeneity across the entire distribution.

Variable selection

The PCCF survey contains information on a very large set of variables. The first step, therefore, in order to estimate the model, outlined in the previous section, is to identify a subset of relevant variables. In doing that, we need to take into consideration the fact that the sample size of the household survey was selected in order to allow for a set of around 10 explanatory variables. We have considered all the sections of the PCCF and divided the available variables in four categories: infrastructure, farming, education and women empowerment and child protection. Given the large number of variables, it is difficult to choose a priori which one should be included in order to predict the risk of child employment. Therefore, we use the stepwise regression model to gather information on which variable is more informative with respect to the child employment rate in each community. In

particular, we estimate both the backward-stepwise and the forward-stepwise methods, which lead us to similar results. [7]

The stepwise procedure was carried out separately for each of the four categories described above.

Table 7 present the results of the procedure indicating the variables considered and the one that resulted significant (detailed results are available on request).

Table 7. **Variable selection: stepwise regression analysis**

Variable	Description	Significant	Variable	Description	Significant
Infrastructure					
Kindergarten	Dummy variable, kindergarten is in the community		Kindergarten No.	Number of kindergartens in the community	
Primary school	Dummy variable, primary school is in the community		Primary school No.	Number of primary schools in the community	Yes
Junior secondary school	Dummy variable, junior secondary school is in the community	Yes	Junior secondary school No.	Number of junior secondary schools in the community	
Health centre	Dummy variable, primary health centre is in the community		Health centre No.	Number of health centre in the community	
Electricity	Dummy variable, connection to electricity network in the community		Kindergarten distance	Distance from kindergarten	
Mobile	Dummy variable, connection to mobile network in the community	Yes	Primary school distance	Distance from primary school	
Internet	Dummy variable, connection to internet network in the community		Health centre distance	Distance from health centre	
Road	Dummy variable, community reachable by road		Senior secondary distance	Distance from senior secondary	Yes
Road surface	Dummy variable, community road surface		Vocational distance	Distance from vocational school	

[7] For additional details see Draper and Smith 1981

Table 7. **Variable selection: stepwise regression analysis (continued)**

Variable	Description	Significant	Variable	Description	Significant
Road accessible	Dummy variable, road accessible all year	Yes	Birth certificate	Percentage of children with birth certificate (0-40, 41-69, 70-100)	
Farming					
Buying company	Dummy variable, licence buying company in the community		Cocoa land size	Cocoa farm size per farmer in the community (acres)	
Cocoa organization	Dummy variable, cocoa farmer organization in the community	Yes	Cocoa farmers	Number of cocoa farmers in the community	
Extension services	Dummy variable, extension services in the community		Cocoa production	Cocoa production per year in the community (ton)	
Input available	Dummy variable, farming inputs available in the community		Farmers trained by Ext. Serv.	Number of farmers trained by ext. services in the community	
Casual labour available	Dummy variable, adult casual labour available in the community		Share of households cultivating cocoa	Percentage of households cultivating cocoa in the community	Yes
Input affordable	Dummy variable, farming inputs affordable in the community				
Agr. Services	Dummy variable, agricultural services in the community				
Education					
Toilet facilities in primary	Dummy variable, toilet facilities in primary school		Enrolment rate	Percentage of children 5-17 enrolled in school	
Scholarship in secondary	Dummy variable, scholarship in secondary school		Children enrolled in kindergarten	Number of children enrolled in kindergarten	
Feeding program in primary	Dummy variable, feeding programme in primary school	Yes	Children enrolled in primary	Number of children enrolled in primary school	

Table 7. **Variable selection: stepwise regression analysis (continued)**

Variable	Description	Significant	Variable	Description	Significant
			Children enrolled in junior secondary	Number of children enrolled in junior secondary school	
			Children enrolled in senior secondary	Number of children enrolled in senior secondary school	
			Teachers No.	Number of teachers paid by the government	
Women empowerment and child protection					
Community Action Plan	Dummy variable, Community Action Plan in the community		Female lead farmers	Number of female lead farmers in the community	
Community Child Protection Committee	Dummy variable, Community Child Protection Committee in the community		Female leadership positions	Percentage of leadership positions occupied by females in the community	Yes
Regulations to protect children	Dummy variable, regulations to protect children in the community		Women education	Main education level reached by women in the community	
Remediation services	Dummy variable, remediation services for children in the community		Women in livelihood activities	Percentage of women engaged in livelihood activities (no cocoa) in the community	Yes

A limited number of variables are statistically significant in explaining rates of children's involvement in employment. According to the results of the stepwise regression, the following variables are included in the estimation of model: the availability of accessible road throughout the year, the presence of a junior secondary school, the number of primary schools, the distance to the nearest senior secondary school, mobile connection (from the infrastructure section); the presence of cocoa organizations and the percentage of households cultivating cocoa (from the farming section); the availability of a feeding school program (from the education section); the percentage of female leadership positions and the percentage of women in livelihood activities (from the women's empowerment and child protection section). Even if not identified as significant in the stepwise regression, we include some additional variables that are potentially of interest in predicting child

employment, i.e.: the availability of casual adult labour (from the farming section); the percentage of women that can read in the community (from the women empowerment and child protection section). The inclusion of these additional variables does not invalidate the estimation procedure.

Community characteristics

Table 8 shows the descriptive statistics for the community characteristics included in the estimate of the empirical model. As reported in the table, there are on average almost 4 primary schools in the communities, 79% of the communities have a junior secondary school, and the distance to the nearest senior secondary school is about 8km on average over all communities. The majority of the communities have a road accessible all year (65%) and has a connection to the mobile network (89%).

The farming background is characterized by several variables: the share of households cultivating cocoa (75%), availability of casual adult labour (87%) and the presence of cocoa organizations (55%).

From the education section, it emerges that 29% of communities have a feeding programme in primary schools.

As far as women empowerment is concerned, in the majority of communities women are involved in livelihood activities (not connected with cocoa production), while in half of communities the majority of women are unable to read. Finally, the percentage of leadership positions occupied by women is 18%.

Table 8. **Community characteristics**

Variable	Obs.	Mean	SD	Min	Max
Child employment in the community (No.)	128	6646.61	13780.01	168.97	113940.00
Children 5-17 (No.)	128	20850.19	39865.13	377.71	324729.00
Child employment in the community (%)	128	0.32	0.12	0.07	0.71
Primary in the community (No.)	128	3.89	5.62	0.00	46.00
Junior High School in the community	128	0.79	0.41	0.00	1.00
Distance to the nearest Senior High School	128	8.36	11.30	0.00	68.00
Community road accessible all year	128	0.65	0.48	0.00	1.00
Community connected to mobile network	128	0.89	0.31	0.00	1.00
Share of households cultivating cocoa	128	0.75	0.27	0.00	1.00
Cocoa farmer organization in the community	128	0.55	0.50	0.00	1.00
Casual adult labour available in the community	128	0.87	0.34	0.00	1.00
Feeding programme at Primary school	128	0.30	0.46	0.00	1.00
Women in livelihood activities (%): 0-30	128	0.24	0.43	0.00	1.00
Women in livelihood activities (%): 31-60	128	0.30	0.46	0.00	1.00
Women in livelihood activities (%): 61-80	128	0.30	0.46	0.00	1.00
Women in livelihood activities (%): 81-100	128	0.15	0.36	0.00	1.00
Female leadership positions (%)	128	0.18	0.18	0.00	0.83
Women can read (%): 0-30	128	0.49	0.50	0.00	1.00
Women can read (%): 31-60	128	0.38	0.49	0.00	1.00
Women can read (%): 61-80	128	0.12	0.32	0.00	1.00
Women can read (%): 81-100	128	0.01	0.09	0.00	1.00

4 IDENTIFYING RISK CLASSES

Child employment

As mentioned, in order to estimate the risk of child labour, we split the sample into three terciles of the percentage of child employment in each community. Table 9 shows the distribution of communities across the three classes. The average incidence of child employment in the low risk class is about 20%. While in the medium and high risk class the child employment rates are respectively 32% and 45%.

Table 9. **Terciles of child employment and average child employment rate**

Class	Obs.	%	Employment Rate
1-low risk	43	33.59	0.20
2-medium risk	43	33.59	0.32
3-high risk	42	32.81	0.45
Total	128	100.0	

Figure 7 presents graphically these results indicating the range of variation around the median.

Figure 7. **Box plot of child employment by class of risk**

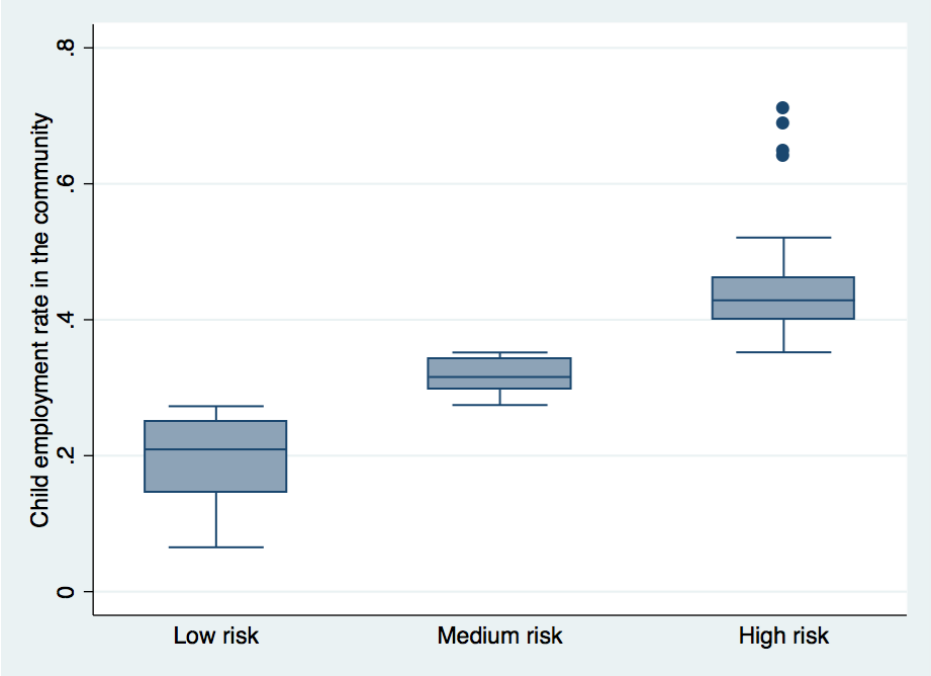


Table 10 shows estimation results of the multinomial logit model from equation (1), where the coefficients (the multinomial log odds) are reported in terms of the base category, which in this case is the first class, i.e. the group with lowest risk of child employment. The marginal effects derived from the multinomial logit model are reported in Table 11.

Table 10 shows that only few coefficients are statistically significantly different from zero. In particular, the relative probability of belonging to the medium risk class of child labour rather than to the low risk class decreases for communities with a road accessible all the year than for communities without an accessible road, while it increases if there is a junior secondary high school in the community.

As the number of primary schools in the community increases, communities are more likely to move from high risk class of child labour to low risk class. Finally, as the distance to the nearest senior secondary high school increases, communities are more likely to belong to the high risk class, rather than to the low risk class.

Table 10. **Multinomial logit, child employment - parameter estimates**

		Classes	
VARIABLES	1	2	3
Primary in the community (No.)	-	-0.011	-0.345**
	-	(0.048)	(0.122)
Junior High School in the community	-	1.845*	1.218
	-	(0.859)	(0.729)
Distance to the nearest Senior High School	-	0.041	0.108**
	-	(0.036)	(0.038)
Community road accessible all year	-	-1.770**	-0.464
	-	(0.618)	(0.697)
Community connected to mobile network	-	0.094	-1.784
	-	(1.168)	(1.035)
Share of households cultivating cocoa	-	1.360	-0.466
	-	(1.045)	(1.039)
Cocoa farmer organization in the community	-	0.040	-0.307
	-	(0.558)	(0.580)
Casual adult labour available in the community	-	0.506	1.862
	-	(0.841)	(0.964)
Feeding programme at Primary school	-	0.752	0.552
	-	(0.557)	(0.627)
Women engaged in livelihood activities (%)	-	-0.272	-0.352
	-	(0.282)	(0.307)
Female leadership positions (%)	-	-1.849	-2.828
	-	(1.535)	(1.703)
Women can read (%)	-	0.198	-0.160
	-	(0.383)	(0.443)
Constant	-	-1.620	1.427
	-	(1.836)	(1.764)
Observations	128	128	128

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Marginal effects in Table 11 show that an additional primary school in the community increases the probability of belonging to the low risk class by 3 percentage points. Moreover, the probability of belonging to the low risk class is on average 27 percentage points higher for communities with connection to mobile network than for communities without connection. As the distance to the nearest senior secondary high school increases by one unit, the probability to be in the low risk class

decreases by about 2 percentage points. However, communities with junior secondary high school are less likely to be in this class by 37 percentage points with respect to communities without the school.

The probability of belonging to the medium risk class increases by 3 percentage points for each additional primary school and by about 30 percentage points for communities with a junior secondary high school. Communities with accessible road have a probability to be in the medium risk class lower by 37 percentage points than communities without an accessible road.

Turning to the high risk class, an additional primary school in the community decreases the probability to be in this class by 7 percentage points, while as the distance to Senior Secondary High School increases by one unit, the probability to be in this class increases by about 2 percentage points. Communities are less likely to be in the high risk class, by 42 percentage points, if there is a connection to the mobile network. Finally, the availability of casual adult labour increases the probability of belonging to the high risk class by 23 percentage points, indicating possible complementarities between child and adult work and/or high level of dependency on agriculture.

Table 11. **Multinomial logit, child employment - marginal effects**

	Classes		
VARIABLES	1-Low risk	2-Medium risk	3-High risk
Primary in the community (No.)	0.033*	0.033*	-0.066***
	(0.013)	(0.014)	(0.019)
Junior High School in the community	-0.369*	0.287*	0.082
	(0.150)	(0.125)	(0.104)
Distance to the nearest Senior High School	-0.015*	-0.001	0.017**
	(0.007)	(0.006)	(0.005)
Community road accessible all year	0.265**	-0.368**	0.103
	(0.102)	(0.113)	(0.095)
Community connected to mobile network	0.188	0.233	-0.420*
	(0.150)	(0.144)	(0.179)
Share of households cultivating cocoa	-0.141	0.372	-0.231
	(0.196)	(0.231)	(0.195)
Cocoa farmer organization in the community	0.023	0.041	-0.064
	(0.110)	(0.115)	(0.099)
Casual adult labour available in the community	-0.218	-0.008	0.226**
	(0.191)	(0.184)	(0.084)
Feeding programme at Primary school	-0.145	0.120	0.024
	(0.104)	(0.114)	(0.104)
Women engaged in livelihood activities (%)	0.069	-0.029	-0.040
	(0.057)	(0.059)	(0.053)
Female leadership positions (%)	0.507	-0.149	-0.358
	(0.306)	(0.328)	(0.300)
Women can read (%)	-0.012	0.064	-0.051
	(0.080)	(0.079)	(0.074)
Observations	128	128	128

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

After the estimation of the multinomial logit, we predict the probability of communities of belonging to the three risk classes. Then, each community was assigned to the risk class with the highest predicted probability. In Table 12 the predicted class membership is compared to the true class membership (the terciles) to analyse the goodness of fit of the adopted model.

Table 12. **Class membership, child employment**

	Predicted class			
Class (terciles)	1	2	3	Total
1	29	6	8	43
2	11	25	7	43
3	8	10	24	42
Total	48	41	39	128

As mentioned in the previous section, we conduct a robustness check for risk class identification, estimating the multinomial logit model with three classes divided by thresholds that are further from the sample median. Specifically, we split the sample by quartiles and consider the first quartile as the first class, the second and the third quartiles aggregated as the second class and the fourth quartile as the third class.

Table 13 and Figure 8 shows the distribution of communities across the three classes. The average incidence of child employment in the low risk class is 18%. While in the medium and high risk class the child employment rates are respectively 32% and 47%.

Table 13. **Robustness check: Quartile-based classes of child employment and average child employment rate**

Class	Obs.	Average Employment Rate
1-low risk	32	0.18
2-medium risk	65	0.32
3-high risk	31	0.47
Total	128	

Figure 8. **Box plot of child employment by quartile-based class of risk**

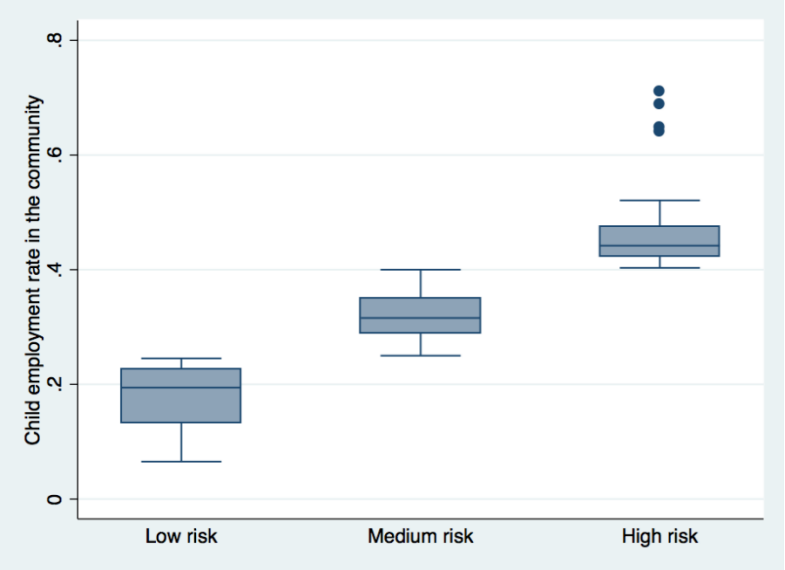


Table 14 shows estimation results of the multinomial logit model using the alternative definition of risk class, where the coefficients (the multinomial log odds) are reported in terms of the base category, which in this case is the first class, i.e. the group with lowest risk of child employment. The marginal effects derived from the multinomial logit model are reported in Table 16.

Table 14 shows results quite similar to those obtained with the terciles (Table 10) both in terms of direction and magnitude of coefficients relative to statistically significant variables. However, looking at the marginal effects in Table 16, we obtain a different picture: only the availability of an accessible road in the community seems to significantly influence the probability of belonging to the low risk class (positively) and to the high risk class (negatively). None of the covariates significantly influences the probability of belonging to the medium risk class.

Table 14. **Multinomial logit, child employment - parameter estimates (quartile-based risk classes)**

	Classes		
VARIABLES	1	2	3
Primary in the community (No.)	-	-0.051	-0.349*
	-	(0.051)	(0.144)
Junior High School in the community	-	1.372	0.474
	-	(0.762)	(0.786)
Distance to the nearest Senior High School	-	0.050	0.100*
	-	(0.041)	(0.045)
Community road accessible all year	-	-1.348*	-0.243
	-	(0.652)	(0.818)
Community connected to mobile network	-	-15.482	-17.444
	-	(1700.121)	(1700.121)
Share of households cultivating cocoa	-	1.670	-0.128
	-	(0.952)	(1.127)

Table 15. **Multinomial logit, child employment - parameter estimates (quartile-based risk classes) (continued)**

VARIABLES	1	2	3
Cocoa farmer organization in the community	-	-0.589	-0.629
	-	(0.560)	(0.669)
Casual adult labour available in the community	-	0.588	2.332*
	-	(0.800)	(1.175)
Feeding programme at Primary school	-	0.148	-0.069
	-	(0.552)	(0.716)
Women engaged in livelihood activities (%)	-	-0.262	-0.336
	-	(0.274)	(0.342)
Women can read (%)	-	0.121	-0.028
	-	(0.369)	(0.481)
Constant	-	14.950	16.403
	-	(1700.122)	(1700.122)
Observations	128	128	128

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 16. **Multinomial logit, child employment - marginal effects (quartile-based risk classes)**

		Classes	
VARIABLES	1	2	3
Primary in the community (No.)	0.006	0.041	-0.047
	(0.943)	(1.208)	(0.265)
Junior High School in the community	-0.079	0.213	-0.133
	(12.337)	(6.854)	(5.486)
Distance to the nearest Senior High School	-0.003	-0.005	0.008
	(0.505)	(0.480)	(0.025)
Community road accessible all year	0.051	-0.196	0.145
	(8.503)	(5.408)	(3.097)
Community connected to mobile network	0.243***	0.223	-0.466**
	(0.056)	(0.181)	(0.176)
Share of households cultivating cocoa	-0.066	0.329	-0.263
	(10.986)	(5.983)	(5.009)
Cocoa farmer organization in the community	0.030	-0.017	-0.012
	(4.934)	(3.988)	(0.951)
Casual adult labour available in the community	-0.054	-0.131	0.185
	(8.481)	(9.631)	(1.152)
Feeding programme at Primary school	-0.005	0.037	-0.032
	(0.871)	(0.390)	(0.499)
Women engaged in livelihood activities (%)	0.014	0.000	-0.014
	(2.334)	(1.973)	(0.365)
Women can read (%)	-0.005	0.027	-0.022
	(0.762)	(0.384)	(0.390)
Observations	128	128	128

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Finally, we predict the probability of communities belonging to the three risk classes, assigning each community to the risk class with the highest predicted probability. In Table 17 the predicted class membership is compared to the true class membership to analyse the goodness of fit of the adopted model. The table shows a fairly good predictive power of the model as the majority of observations lies along the main diagonal. However, constituting risk classes based on quartiles rather than terciles

on the distribution of child labour incidence does not lead to a significant improvement of the predictive power of the model.

Table 17. **Class membership, child employment (quartile-based risk classes)**

	Predicted class			
Class	1	2	3	Total
1	13	18	1	32
2	8	51	6	65
3	3	11	17	31
Total	24	80	24	128

Child labour in agriculture

Also for child labour in agriculture, we split the sample into three terciles of the percentage of child labour in each community. Table 18 shows the distribution of communities across the three classes. The average incidence of child labour in the low risk class is about 8%. While in the medium and high risk class the child labour rates are respectively 18% and 32%.

Table 18. **Terciles of child labour and average child labour rate**

Class	Obs.	%	Employment Rate
1-low risk	43	33.59	0.08
2-medium risk	43	33.59	0.18
3-high risk	42	32.81	0.32
Total	128	100.0	

Figure 9 presents graphically these results indicating the range of variation around the median.

Figure 9. **Box plot of child labour by class of risk**

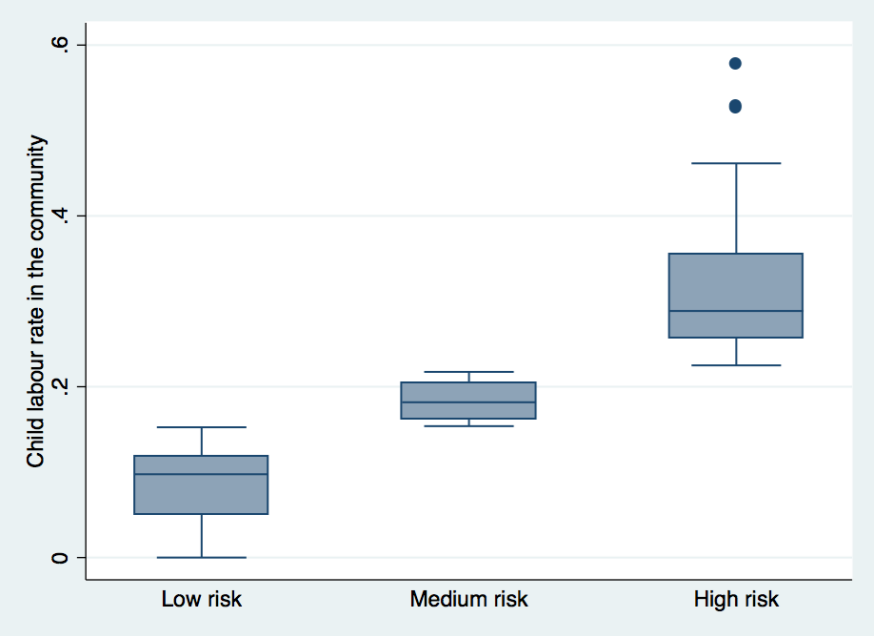


Table 19 shows estimation results of the multinomial logit model from equation (1), where the coefficients (the multinomial log odds) are reported in terms of the base category, which in this case is the first class, i.e. the group with lowest risk of child labour. The marginal effects derived from the multinomial logit model are reported in Table 20.

Table 19 shows that only few coefficients are statistically significantly different from zero. In particular, the relative probability of belonging to the medium risk class of child labour rather than to the low risk class decreases for communities with a road accessible all the year than for communities without an accessible road. As the number of primary schools in the community increases, communities are more likely to move from high risk class of child labour to low risk class, as well as if communities have available connection to mobile network. Finally, as the distance to the nearest senior secondary high school increases, communities are more likely to belong to the high risk class, rather than to the low risk class.

Table 19. **Multinomial logit, child labour - parameter estimates**

		Classes	
VARIABLES	1	2	3
Primary in the community (No.)	-	-0.066	-0.380**
	-	(0.060)	(0.125)
Junior High School in the community	-	-0.662	-0.195
	-	(0.764)	(0.779)
Distance to the nearest Senior High School	-	-0.005	0.069*
	-	(0.031)	(0.031)
Community road accessible all year	-	-1.318*	0.549
	-	(0.557)	(0.683)
Community connected to mobile network	-	-0.630	-2.542*
	-	(1.275)	(1.205)
Share of households cultivating cocoa	-	0.697	0.741
	-	(0.954)	(1.057)
Cocoa farmer organization in the community	-	0.014	-0.106
	-	(0.530)	(0.569)
Casual adult labour available in the community	-	0.082	1.511
	-	(0.762)	(0.934)
Feeding programme at Primary school	-	0.124	-0.185
	-	(0.518)	(0.596)
Women engaged in livelihood activities (%)	-	-0.212	-0.217
	-	(0.259)	(0.297)
Female leadership positions (%)	-	0.166	-0.578
	-	(1.472)	(1.585)
Women can read (%)	-	0.297	0.253
	-	(0.356)	(0.416)
Constant	-	1.720	1.072
	-	(1.809)	(1.916)
Observations	128	128	128

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

From the marginal effects reported in Table 20, it emerges that an additional primary school in the community increases the probability of belonging to the low risk class by 4 percentage points. Moreover, the probability of belonging to the low risk class is on average 28 percentage points higher for communities with connection to mobile network than for communities without connection. The probability of belonging to the medium risk class is significantly influenced only by the availability of an accessible road throughout the year. In fact, communities with an accessible road have a probability to be in the medium risk class lower by 36 percentage points than communities without an accessible road. Turning to the high risk class, an additional primary school in the community decreases the probability to be in this class by 6 percentage points, while as the distance to senior secondary high school increases by one unit the probability to be in this class increases by 1 percentage point. Surprisingly, communities are more likely to be in the high risk class by 22 percentage points if there

is a road accessible throughout the year, and less likely to be in this class by 50 percentage points if there is a connection to the mobile network. Finally, the availability of casual adult labour increases the probability of belonging to the high risk class by 20 percentage points, indicating possible complementarities between child and adult labour.

Table 20. **Multinomial logit, child labour - marginal effects**

	Classes		
VARIABLES	1	2	3
Primary in the community (No.)	0.042**	0.022	-0.064***
	(0.014)	(0.016)	(0.018)
Junior High School in the community	0.107	-0.142	0.034
	(0.141)	(0.153)	(0.108)
Distance to the nearest Senior High School	-0.005	-0.008	0.013**
	(0.006)	(0.006)	(0.005)
Community road accessible all year	0.148	-0.363***	0.215**
	(0.104)	(0.105)	(0.082)
Community connected to mobile network	0.282*	0.206	-0.488**
	(0.119)	(0.135)	(0.158)
Share of households cultivating cocoa	-0.162	0.093	0.069
	(0.191)	(0.209)	(0.180)
Cocoa farmer organization in the community	0.007	0.014	-0.021
	(0.107)	(0.111)	(0.094)
Casual adult labour available in the community	-0.112	-0.090	0.202*
	(0.175)	(0.170)	(0.081)
Feeding programme at Primary school	-0.002	0.048	-0.046
	(0.108)	(0.109)	(0.091)
Women engaged in livelihood activities (%)	0.049	-0.029	-0.019
	(0.054)	(0.054)	(0.048)
Female leadership positions (%)	0.027	0.098	-0.125
	(0.298)	(0.312)	(0.264)
Women can read (%)	-0.064	0.046	0.018
	(0.075)	(0.074)	(0.068)
Observations	128	128	128

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

On the basis of the estimates of the multinomial logit, we predict the probability of communities of belonging to the three risk classes. Each community is assigned to the risk class with the highest predicted probability. In Table 21, the predicted class membership is compared to the true class membership (the terciles) to analyse the goodness of fit of the adopted model. The table shows that the goodness of fit of the model is relatively good as the majority of observations lies along the main diagonal.

Table 21. **Class membership, child labour**

	Predicted class			
Class (terciles)	1	2	3	Total
1	28	9	6	43
2	11	22	10	43
3	8	5	29	42
Total	47	36	45	128

5 RESULTS BASED ON INDIVIDUAL AND HOUSEHOLD LEVEL DATA

Given the availability of individual and household level data, we analyse the determinants of child employment exploiting this information, in order to check the consistency with results obtained using community level data.

We consider the following individual level covariates: sex and age of the child. At the household level, we include: sex and education of the household head, a dummy variable indicating whether the household is in an urban or rural area, a dummy variable indicating whether the household cultivates cocoa (which was also included in the community level analysis and averaged at the community level) and monthly household income. Finally, we include also the covariates at the community level (from the PCCF questionnaire) used for the community level estimates.

We first estimate a multinomial logit relative to four mutually exclusive categories: study only, work only, work and study, nothing. The results are shown in Table 22 in terms of marginal effects. We also estimate a probit model for the probability to be in employment and the results are shown in Table 23 in terms of marginal effects.

Both the multinomial logit and the probit model estimates show that at the individual level age is an important determinant of child employment, while there are not significant gender effects. Also the area of residence influences children's activities, as children in rural areas are less likely to attend school without working and more likely to work (both with and without attending school). As the education of the household head increases, the probability of children of working only and of doing nothing significantly decreases. Finally, a remarkable determinant of children's activities is the

involvement of the household in cocoa production. The two tables indicate that children in households producing cocoa are more likely to be in employment (both with and without attending school) and, in particular, less likely to attend school without working and more likely to do neither activity.

Consistently with the results obtained in the previous sections, the community level covariates on the availability of primary and junior secondary schools, cocoa farmer organizations, primary school feeding programme and female leadership positions significantly reduce the probability of child employment.

Table 22. **Determinants of child activities - individual level**

	Study only	Work only	Work and study	Nothing
Male	-0.023	-0.001	0.029	-0.004
	(0.024)	(0.004)	(0.023)	(0.005)
Age	-0.060**	-0.008	0.085*	-0.017*
	(0.023)	(0.005)	(0.022)	(0.004)
Age^2	0.001	0.001**	-0.003**	0.001*
	(0.001)	(0.000)	(0.001)	(0.000)
Rural	-0.091*	0.010*	0.078**	0.002
	(0.027)	(0.005)	(0.027)	(0.005)
Male head	0.004	-0.002	0.000	-0.002
	(0.030)	(0.005)	(0.029)	(0.006)
Head education	0.001	-0.001**	0.001	-0.001*
	(0.001)	(0.000)	(0.001)	(0.000)
HH in cocoa	-0.074*	0.012*	0.054**	0.008**
	(0.022)	(0.004)	(0.021)	(0.004)
Monthly income (log)	0.026**	-0.008*	-0.011	-0.007**
	(0.013)	(0.002)	(0.013)	(0.002)
Primary in the community (No.)	0.003	0.000	-0.004*	0.001*
	(0.002)	(0.000)	(0.002)	(0.000)
Junior High School in the community	-0.075*	-0.002	0.087*	-0.011**
	(0.023)	(0.003)	(0.022)	(0.004)
Distance to the nearest Senior High School	-0.003*	0.000**	0.002**	0.000
	(0.001)	(0.000)	(0.001)	(0.000)
Community road accessible all year	0.02	0.007**	-0.039*	0.012**
	(0.023)	(0.004)	(0.023)	(0.005)
Community connected to mobile network	0.036	0.006	-0.054*	0.012*
	(0.031)	(0.005)	(0.031)	(0.006)
Cocoa farmer organization in the community	0.042**	-0.003	-0.041**	0.001
	(0.019)	(0.004)	(0.019)	(0.004)
Casual adult labour available in the community	-0.043	0.012**	0.02	0.011*
	(0.028)	(0.005)	(0.028)	(0.006)
Feeding programme at Primary school	0.045*	0.006	-0.064**	0.013**
	(0.023)	(0.004)	(0.022)	(0.005)
Women engaged in livelihood activities (%)	-0.002	-0.001	0.004	-0.001
	(0.009)	(0.002)	(0.009)	(0.002)
Female leadership positions (%)	0.111**	-0.01	-0.100**	-0.001
	(0.051)	(0.009)	(0.051)	(0.010)
Women can read (%)	-0.026*	0.000	0.028*	-0.003
	(0.015)	(0.003)	(0.014)	(0.004)

* 0.10 ** 0.05 * 0.001; SE in parenthesis; marginal effects from multinomial logit estimation

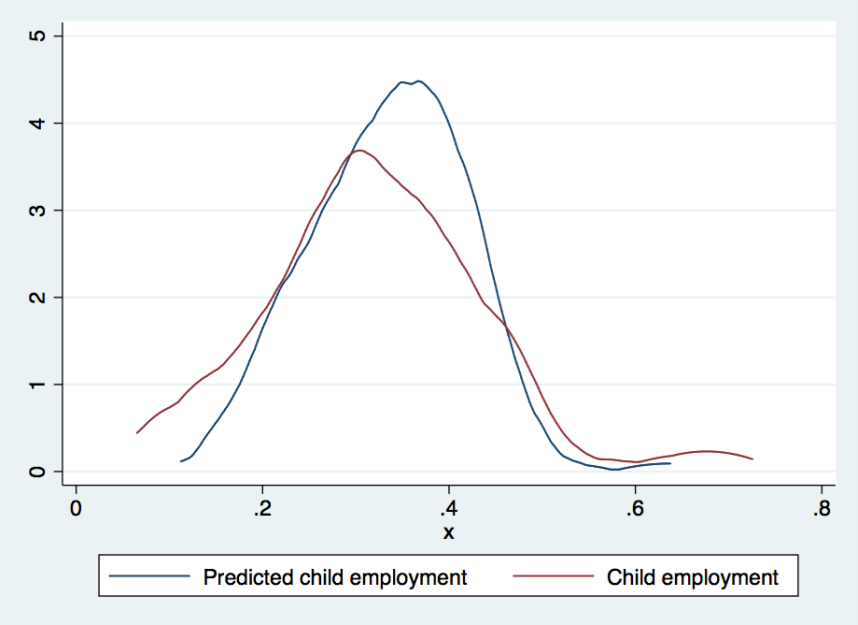
Table 23. **Determinants of child employment - individual level**

	Child employment
Male	0.024
	(0.024)
Age	0.077*
	(0.023)
Age^2	-0.002**
	(0.001)
Rural	0.092*
	(0.028)
Male head	0.001
	(0.030)
Head education	0.000
	(0.001)
HH in cocoa	0.062**
	(0.022)
Monthly income (log)	-0.02
	(0.013)
Primary in the community (No.)	-0.004*
	(0.002)
Junior High School in the community	0.086*
	(0.023)
Distance to the nearest Senior High School	0.003*
	(0.001)
Community road accessible all year	-0.031
	(0.023)
Community connected to mobile network	-0.051
	(0.031)
Cocoa farmer organization in the community	-0.044**
	(0.020)
Casual adult labour available in the community	0.03
	(0.029)
Feeding programme at Primary school	-0.056**
	(0.023)
Women engaged in livelihood activities (%)	0.004
	(0.009)
Female leadership positions (%)	-0.105**
	(0.053)
Women can read (%)	0.028*
	(0.015)

* 0.10 ** 0.05 * 0.001; SE in parenthesis; marginal effects from probit estimation.

The following figure shows the predicted child employment obtained with the probit model (at the individual level) and averaged at the community level. We compare the predicted values with the observed child employment at the community level looking at the kernel densities.

Figure 10. **Predicted child employment from individual level probit estimation**



The probit model is estimated also using a dummy indicator for child labour as dependent variables and the same set of covariates as independent variables. The results, reported in the following table, show some differences with results obtained considering child employment. In fact, we observe that very few covariates are statistically significant determinants of child labour, i.e. age, involvement of household in cocoa production, distance to senior secondary school and primary school feeding programme.

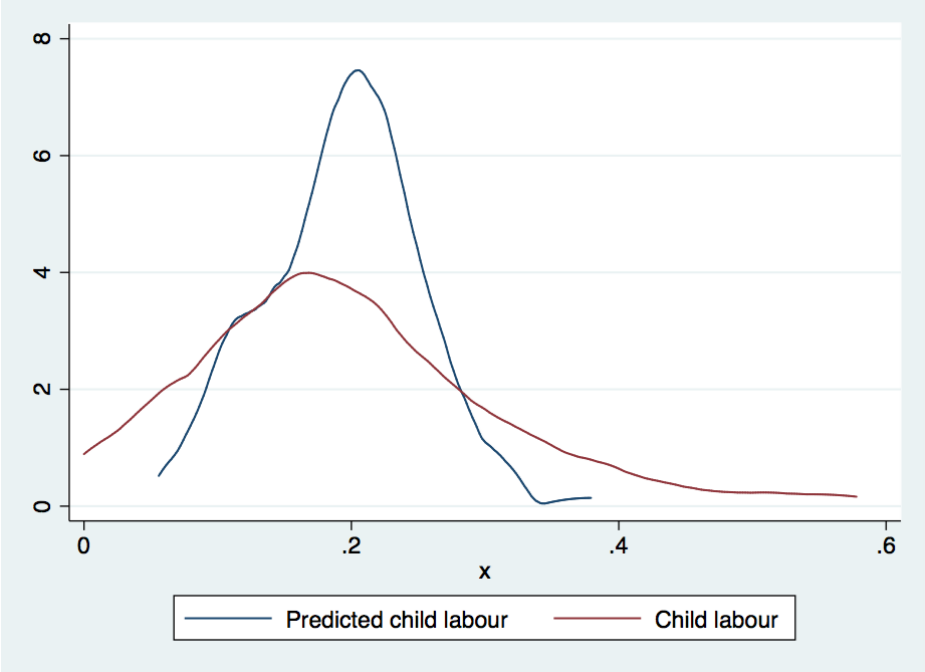
Also the comparison between the predicted value of child labour and the observed child labour, averaged at the community level, in Figure 11 leads to different picture with respect to what observed with predicted child employment in the previous figure. In this case, the kernel density of predicted child labour seems to be quite different from the kernel density of observed child labour. This result could also be given to the selected nature of the sample used for child labour.

Table 24. **Determinants of child labour - individual level**

	Child labour
Male	0.03
	(0.018)
Age	0.034*
	(0.018)
Age^2	-0.001
	(0.001)
Rural	0.071**
	(0.023)
Male head	-0.03
	(0.024)
Head education	0.000
	(0.001)
HH in cocoa	0.073*
	(0.016)
Monthly income (log)	-0.006
	(0.010)
Primary in the community (No.)	-0.002
	(0.002)
Junior High School in the community	0.013
	(0.017)
Distance to the nearest Senior High School	0.002**
	(0.001)
Community road accessible all year	0.022
	(0.019)
Community connected to mobile network	-0.014
	(0.023)
Cocoa farmer organization in the community	-0.004
	(0.015)
Casual adult labour available in the community	0.01
	(0.024)
Feeding programme at Primary school	-0.044**
	(0.018)
Women engaged in livelihood activities (%)	0.001
	(0.007)
Female leadership positions (%)	-0.02
	(0.040)
Women can read (%)	0.008
	(0.012)

* 0.10 ** 0.05 * 0.001; SE in parenthesis; marginal effects from probit estimation.

Figure 11. **Predicted child labour from individual level probit estimation**



6 CONCLUSION

In this paper we have built a "risk" indicator for the presence of child labour at the community level in Ghana. In particular, the study identifies classes of risk of child labour based on community characteristics. Different techniques are available to support the identification of community level classes of risk of the presence of child labour that do not establish a priori the number of classes, i.e., Latent Class or Finite Mixture models. However, the heterogeneity in terms of child labour incidence at community level was not substantial enough to apply these methods. Therefore, a different approach was followed. Three classes of risk were identified a priori by dividing the sample in three terciles on the basis of the incidence of child labour at community level. The three terciles represent three classes of risk of child labour: low, medium and high risk. A multinomial logit model was then estimated to predict the probability of belonging to any of the identified risk class.

The study makes use of two primary data sources. The 2017 Ghana Child Labour Prevalence Survey and the 2017 Protective Cocoa Community Framework (PCCF). The Child labour prevalence survey is an ad-hoc survey carried out in 2017 to measure the incidence and characteristics of child labour in the cocoa growing areas of Ghana. The PCCF was also conducted in 2017 and in the same communities selected for the Child labour prevalence survey. It is a community assessment tool, incorporating key indicators and proxies related to community development, community empowerment, education, child protection, gender and livelihoods in cocoa-growing communities in Ghana.

We first consider the full sample and the individual child employment information averaged at community level to estimate the model and classify the communities in classes of risk. The average incidence of child employment in the low risk class is about 20%. While in the medium and high risk class the child employment rates are respectively 32% and 45%.

The study finds that the most statistically significant community characteristics influencing child employment risk classification are the presence of a primary school, presence of junior secondary high school, distance to the nearest senior secondary high school, availability of a connection to a mobile network, access to infrastructure (i.e., access to road) and availability of casual adult labour.

The results show that an additional primary school in the community increases the probability of belonging to the low and medium risk class by 3 percentage points and reduces the probability of being in the high risk class by about 7 percentage points. Communities with the presence of junior secondary high school are less likely to belong to the low risk class and more likely to belong to the

medium risk class. The effect is not different from zero regarding the probability of belonging to the high risk class of child labour.

As the distance to the nearest senior secondary high school increases by one unit, the probability of a community to be in the low risk class decreases by about 2 percentage points, while the probability of being in the high risk class decreases by again 2 percentage points.

Access to infrastructure matters. Communities are less likely to be in the high risk class, by 42 percentage points, if there is a connection to the mobile network, and more likely to be in the low risk class by 27 percentage points in presence of accessible roads compared with communities with no or limited access to mobile network and accessible roads throughout the year. Finally, the availability of casual adult labour increases the probability of belonging to the high risk class by 23 percentage points, indicating possible complementarities between child and adult work and/or high level of dependency on agriculture.

ANNEX I: EXTENDED REFERENCES AND BIBLIOGRAPHY

Abou, Edouard Pokou (2014). A Re-examination of the determinants of child labour in Côte d'Ivoire. Nairobi: African Economic Research Consortium.

Boas, Morten; Huser, Anne (2006). Child labour and Cocoa Production in West Africa: The Case of Côte d'Ivoire and Ghana. FAFO Report 522, 2006.

Cameron A. C. and Trivedi P. K. (2005). Microeconometrics, Methods and Applications. Cambridge University Press.

Draper, N., and H. Smith (1981), Applied Regression Analysis. 2d ed. New York: Wiley

ICI 2016. *Researching the Impact of Increased Cocoa Yields on the Labour Market and Child Labour Risk in Ghana and Côte d'Ivoire*. ICI Labour Market Research Study, 2016

Kolavalli, S., & Vigneri, M. (2011). Cocoa in Ghana: Shaping the success of an economy. Yes, Africa can: success stories from a dynamic continent, 2011

McLachlan, G., Peel, D. (2000). Finite Mixture Models. New York: Wiley.

Understanding Children's Work (UCW) Programme (2016). Not just cocoa. Child labour in the agricultural sector in Ghana, Understanding Children's Work Programme Country Report Series (Rome)

Understanding Children's Work (UCW) Programme (2014). Le double défi du travail des enfants et de la marginalisation scolaire dans la région de la CEDEAO. Understanding Children's Work Programme Country Report Series (Rome)

ANNEX 2: QUESTIONS USED TO DEFINE CHILDREN'S EMPLOYMENT AND CHILD LABOUR

In what follows we detail the questions of the survey instrument used to define child employment and child labour.

Children's Employment. We define a child, aged 5 to 17 years, to be in employment if the following questions were affirmatively answered:

- Did you engage in any work for at least one hour during the last 7 days? (as employee, self-employed, employer or unpaid family work) ?
- During the last 7 days, did you do any of the following activities, even for only one hour? We excluded fetch water or collect firewood for household use.
- Even though you did not do any of these activities in the past week, do you have a job, business, or other economic or farming activity that you will definitely return to?
- Work in agriculture involves work on your own or the household's plot or any other farm, food garden, or help in growing farm produce or in looking after animals for the household. Did you work in agriculture during the last 7 days?

Child Labour. Following the international definition, children are classified in child labour on the basis of the following criteria:

- children aged 5-11 years in employment; and
- children aged 12-14 years in (i) hazardous occupation or (ii) working from 14 to 42 hours per week; and
- children aged 15-17 years working in (i) hazardous occupations or (ii) working more than 42 hours per week

(i) **Hazardous occupation.**

Working children aged 12 to 17 years were considered to be involved in hazardous occupations if involved in:

(1) *Hazardous activities,* identified using the following questions: *Did you experience any of the following environmental hazards and other difficulties while working in agriculture in the last 7 days?*

Dust or dangerous fumes ; Fire, gas or flames ; Long hours in the direct sun ; Work at dangerous heights (high up on a tree, etc.) ; Climbing trees higher than 3 meters to cut mistletoe with cutlass ; Work in water, lake, pond or river ; Spraying of pesticides, insecticides ; Being present or working in the vicinity of farm during pesticide spraying ; Reentering a sprayed farm within less than 12 hours of spraying ; Working alone on the farm in isolation (i.e. beyond visible or audible range of nearest adult) ; Going to or returning from the farm alone or working on farm between 6.00 p.m. and 6.00 a.m. ; Withdrawn from school during cocoa season to do farm work ; The sale, transport, or handling of agro-pharmaceutical products (insecticides, herbicides, fungicide, chemical fertilizers, etc.);

(2) *Experience of injuries,* identified using the following question: *Did you experience any of the following health consequences from exposure to these hazards when performing work in agriculture in the last 7 days?*
Experienced injuries include: I was in very bad pain ; I felt very sick or tired ; I did not feel well for a long time ; I had to receive treatment at a health center ; I had to receive treatment at a hospital ; I could not continue working ; I could not go to school ; other ;
(3) *Use of dangerous tools*, identified using the following question : *When working in agriculture did you use any of the following tools, equipment or machinery in the last 7 days?*
Dangerous tools include : Machete ; Tractor ; Bulloc ; Hoe ; Sprayer ; Sickle

(ii) working from 14 to 42 hours per week and working more than 42 hours per week.

Average weekly working hours. We consider the number of hours per days worked during the week prior to the survey reported under the following questions: *During each day of the last 7 days how many hours did you do this work?*