Growth and Labour Markets in Low Income Countries Programme

WP

GLM|LIC Working Paper No. 47 | May 2019

# Child Labor Measurement in Agricultural Households: Seasonality, Proxy Respondent and Gender Information Gaps in Ethiopia

**Jose Galdo** (Carleton University and IZA)
**Ana C. Dammert** (Carleton University and IZA)
**Degnet Abebaw** (EEPRI)

IZA Institute of Labor Economics
Initiated by Deutsche Post Foundation

UKaid
from the British people

GLM|LIC

**GLM|LIC Working Paper No. 47** | May 2019

# Child Labor Measurement in Agricultural Households: Seasonality, Proxy Respondent and Gender Information Gaps in Ethiopia

**Jose Galdo** (Carleton University and IZA)
**Ana C. Dammert** (Carleton University and IZA)
**Degnet Abebaw** (EEPRI)

**GLM|LIC**
c/o IZA – Institute of Labor Economics
Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-510

Email: glm-review@iza.org

IZA Institute of Labor Economics
Initiated by Deutsche Post Foundation

UKaid from the British people

**GLM|LIC Working Paper No. 47** | May 2019


# ABSTRACT

## Child Labor Measurement in Agricultural Households: Seasonality, Proxy Respondent and Gender Information Gaps in Ethiopia*

Measurement of child labor is critical for our understanding of its determinants and conditions, and for the design of social protection programs and policy. In this study, we report the findings of three survey design experiments implemented across Fairtrade coffee households in rural Ethiopia in three different agricultural seasons. Substantial variation in child labor participation ranging from 45 to 76 percent emerges depending on the timing of the survey. Random allocation of the survey to either the child or a head of household proxy respondent shows the work of girls in agricultural settings is systematically underreported by proxy respondents relative to the child's reports. Underreporting is explained by the child/proxy gender mismatch as differences in child labor reports ranges from 5 to 10 percentage points for girls when the proxy respondent is male. No reporting differences are found for boys across all seasons when the proxy respondent is male. Underreporting by the proxy respondent, relative to the child's own report, is not observed in households with homogenous child gender distribution. On the other hand, knowledge of Fairtrade standards and the degree and scope of effective commercial links between farmers and Fairtrade cooperatives do not have a systematic differential effect on the proxy reporting of child labor.



**JEL Classification:**
C8, J22, 012, Q12

**Keywords:**
survey design, child labor, gender



**Corresponding author:**
Jose Galdo
Department of Economics
Carleton University
1125 Colonel By Drive
Ottawa, Ontario K1S 5B6
Canada
E-mail: jose_galdo@carleton.ca


* The authors gratefully acknowledge the financial support from the IZA/DFID Growth and Labour Markets in Low Income Countries (GLM-LIC), grant agreement GA-C3-RA5-323. We thank seminar participants at the University of Maryland- College Park, 3ie-IFPRI Seminar, Ohio University, 2nd IZA/DFID GLM-LIC Research Network Conference, 2016 LACEA, and the 2017 Canadian Economics Association Meeting for comments and suggestions. This research is the result of a collaborative effort between Carleton University, the Partnership for Economic Policy (PEP), and the Ethiopian Economic Research Policy Institute (EEPRI). Any errors and omissions are our own.

## Introduction

Child labor continues to be an economic necessity for many households, especially poor households in developing countries. The International Labor Organization (ILO) estimates that 218 million children ages 5 to 17 were engaged in an economic activity around the world in 2016, mostly in developing countries. Over 73 million of these children were engaged in hazardous and exploitative forms of child labor. The agricultural sector accounts for by far the largest share of child labor (70 percent of those in child labor), where most children engage in family farm activities (ILO 2017). Accordingly, the measurement of child labor is essential to our understanding of the main factors that drive children to work and the development of sound policy in agricultural areas.

Despite the increasing availability of sources of information on child labor, and a large body of research on its main determinants and the effects of social protection policies (Edmonds 2009; Basu and Tzannatos 2003; Dammert et al 2017), there is little evidence on the validity of data collection methods for child labor with the only experimental evidence coming from Dillon et al 2012. Three overlooked features in child labor survey design and data collection deserve attention. Unlike adult work statistics, child labor statistics are mainly generated by proxy reporting. Proxy reporting, as opposed to self-reporting, could lead to attenuation bias in child labor statistics due to social and cultural values or asymmetries of information due to the gendered segmentation of labor markets. Moreover, the timing of the survey is important as seasonality of the household production function is a key component of the demand for child labor in agricultural settings. While there is a growing body of work on the impact of seasonality on rural labor markets (e.g. Beegle et al. 2017), there is scant evidence on how seasonality may affect the measurement of child labor statistics (e.g., Guarcello 2010). Furthermore, the gender stratification of work and social lives in poor economies could lead to important information gaps in child labor measurement. In East Africa, for instance, child's work is seen as an extension of, and subordinate to, women's work (e.g., Bass 2004), which could lead the work of girls to be unaccounted because their work is directed by women but reported by male heads of households.

This paper contributes to the literature by examining the extent to which seasonality and reliance on proxy respondents affect child labor statistics in rural Ethiopia. Specifically, our experimental survey design intervention consists of the random manipulation of the survey respondent in the application of the same survey instrument to 1200 Fairtrade households in three different seasons of the coffee production, which enable us to capture variation in the demand for child labor. A component of the household survey elicits information about the specific farm and non-farm labor activities of children. In some households, children are randomly selected as respondents (treatment group), while in others the head of households or spouses (control group) are selected as respondents. We conducted the same survey experiment during the *Meher* season (main rainy season), the *Belg* season (short rainy season) and the harvest season. By interviewing the same households in different seasons of the agricultural calendar, we incorporate into the analysis the fact that rural households' demand for child labor differs during the pre- and post-harvest seasons.

We focus on coffee-growing households engaged in Fairtrade activities, and as such, households whose income depends on seasonal agricultural coffee output. Fairtrade constitutes one of the fastest-growing niches in the global food market, with more than 5.5 billion Euros in registered sales in 2014, directly impacting more than 1.4 million small-scale producers in more than 74 developing countries (Fairtrade International 2014). Coffee is the second most traded commodity worldwide after oil and an intensive child labor cash crop (Kruger 2007). It is worth noting that to date there is no evidence of the magnitude and significance of child labor in the context of Fairtrade smallholder associations.

Our main results indicate substantial variation in child labor statistics at the extensive and intensive margins depending on the timing of the field survey. Indeed, the child labor participation rate changes from 45 percent in the main rainy season to 76 percent in the harvest season. The paper documents substantive and statistically significant underreporting of child labor by proxy respondents compared to the data reported by children themselves. Results show that these differences are sensitive to the agricultural season and the gender of the child. Underreporting is particularly salient for girls in the main rainy season (10 percentage points) and harvesting season (7 percentage points). Proxy underreporting is driven by the child/proxy gender mismatch, as differences in child labor statistics for girls emerge when the proxy respondent is the male head of household. No reporting differences are found for boys across all seasons when the proxy respondent is male. Variation in the knowledge of Fairtrade standards and the degree and scope of effective commercial link practices between farmers and Fairtrade cooperatives do not have a systematic differential effect on the proxy reporting of child labor status.

From a policy standpoint, these results suggest that one possible way to improve the accuracy of child labor measurement in contexts in which there is gender segmentation in women's participation in economic activities is to address child labor questions to both the household heads and their spouses. Results show a significant reduction in the child/proxy reporting gap in child labor statistics for girls when the proxy respondent is the spouse of the head of household.

Our paper contributes to the previous literature in several domains. First, our results contribute to the small literature on measurement of child labor outcomes in developing countries[1]. Two non-experimental studies have shown significant differences in child labor measures when comparing proxy versus self-reported information based on nationally representative household surveys (Dammert and Galdo (2013) for Peru and Janzen (2016) for Tanzania). On the contrary, Dillon et al. (2012) find no significant differences between child-based and proxy-based responses in a field survey experiment in Tanzania. There are several differences between our research design and that of Dillon et al (2012). Our sample is composed of a homogenous sample

[1] Measurement and survey design have recently received renewed attention in development economics. For example, the selected method of data collection provides different measures of malnutrition in Tanzania (Cayers et al 2012) and usage of loans for consumption purposes in Peru (Karlan and Zinman 2012). Likewise, self-reported number of doctor visits and per capita expenditures in India is sensitive to the reference period for which the outcome is reported (Das et al 2012).

of rural certified coffee farmers while Dillon et al.'s (2012) sample included urban and rural households, which might entail different opportunity costs and demand for children's activities. Moreover, we implemented survey design experiments during three different coffee agricultural seasons to determine whether households' allocation of child labor differ during the pre- and post-harvest seasons. Furthermore, we considered only the household heads as possible proxy respondents.[2] Dillon et al. (2012) randomly selected proxy respondents among all household members at least 15 years old, and as a result, parents made up 67 percent of the proxy respondents, while other relatives accounted for 33 percent. Finally, our sampling design was stratified by the gender of the household head, which allowed us to estimate child/proxy gender effects.

Second, we expand the growing body of work on the impact of seasonality on rural labor markets (e.g. Beegle et al. 2017). Our results highlight the importance of accounting for seasonal variation in labor demand over the production process. As informational constraints may be present in contexts where farms are mostly operated by families and monitoring is costly (Bharadwaj 2015), misreporting of children's activities can vary depending on the degree of complementarity of child's effort to adult labor, which is dependent on the seasonality of the activities.

Third, there is a consensus in the literature that women's work is poorly measured in developing countries (Mata-Greenwood 2000, Anker 1983). In fact, research on adult labor markets indicates that female labor statistics are affected by the gender division of farm labor across agricultural seasons (e.g. Beegle et al. 2017), and by gendered norms and beliefs about female employment when the proxy respondent is male (Reynolds and Wagner 2012, Lee and Lee 2012). We extend these results to the context of child labor statistics and assess the importance of the proxy-child gender match for the accuracy of child labor statistics.

Fourth, this study responds directly to the increasing call for assessing the 'external validity' of randomized interventions in developing countries (e.g. Banerjee 2015, Deaton and Cartwright 2016). By implementing the same design experiment in three different seasons, this study aims at providing 'credibility enhancing arguments' (Rodrick 2009) to our intervention. Had we implemented only one survey design experiment and picked the short rainy season for that purpose, we would have concluded that respondent type is not a source of variation in child labor statistics.

This study is organized as follows. Section 2 describes the context of this study, while Section 3 explains the sampling procedure and survey design intervention. Section 4 presents the main results and Section 5 assesses the role of differential treatment effects across a set of policy variables of interest. Finally, Section 6 presents conclusions and policy recommendations.

[22] Only in 2 percent of cases the proxy respondent was not a household head but a different household member.

## 2. Coffee Cultivation and Child Labor in Ethiopia

This study focuses on Ethiopia, one of the poorest countries in the world, with an estimated GDP per capita of US$500 dollars. Ethiopia is the second most populous country in Africa, with a total population close to 100 million, of which 80 percent live in rural areas and 60 percent are below 25 years of age (World Bank 2014). The agricultural sector accounts for 46 percent of the GDP, 85 percent of total employment, and 90 percent of export revenues (FAO 2015). Ethiopia is believed to have one of the highest rates of child labor in the world, with 54 percent of rural 5 to 14-year-old children are directly involved in economic activities according to one estimate, mostly as unpaid workers in family farms (Guarcello and Rosati 2007).[3]

Ethiopia is the birthplace of the Arabica coffee, a variety of coffee crop that is quite popular in international markets. It is estimated that over 4 million primarily smallholder households are engaged in coffee cultivation across the country (Minten et al. 2014), which implies that around 25 percent of the Ethiopian population depends directly or indirectly on coffee output and prices for their livelihood (Backman 2009). Indeed, coffee is Ethiopia's most important export crop, accounting for 22 percent of Ethiopia's commodity exports in 2014.

Coffee cultivation is a child labor-intensive crop due to the characteristics of the tasks associated with the pre-harvesting and harvesting production process (Kruger 2007). Before the harvest season, it has been documented that children participate actively in pruning, weeding, and fertilizing. At harvest time, coffee cherries must be picked immediately upon ripening to maximize their quality. Coffee producers employ children mainly as pickers of red coffee beans at harvest time (ILO 2004). ILO rapid assessments of child labor in coffee plantations show that child workers are either children of farm workers or the children of farmers residing near the plantations (Kifle et al 2005).

After the harvest, the red-cherry is dried using either the wet method or the dry method. The former involves drying the whole cherries on mats in the sun after which the beans are separated from the pulp. The latter requires specialized equipment and is usually done at a mill where the cherry skin is removed with a pressing machine and the bean is left to ferment in water in order to remove any remaining skin before the drying stage. In both methods, the dried beans, known as parchment coffee, are bagged and transported to processors.[4] All in all, two important features of small-scale coffee cultivation are its seasonal nature and high dependency on family work.

Fairtrade's presence in the Ethiopian coffee markets has consistently expanded in the past few years from a handful of certified primary cooperatives in early 2000 to 128 certified cooperatives

[3] In contrast to other developing settings, there are no social stigmas or negative perceptions of child labor in Ethiopia, as it is the case throughout eastern Africa. Child labor historically was viewed positively as a means of social reproduction and the useful training for children, which is explained by the cultural and historical influence of East Africa's triple heritage i.e., African, Islamic, and colonial (Bass 2004).

[4] Moisture content and the careful sorting of ripe cherries and dried beans from defective ones during the drying process are essential for quality (Minten et al 2014).

in 2014 (Minten et al 2015). Fairtrade labeling is a voluntary certification scheme that aims to respond to price uncertainty in agricultural markets and to improve the living conditions of smallholder farmers in developing countries (Dragusanu et al 2014)[5]. Certification is issued on behalf of cooperatives of smallholder producers after specific environmental, labor, and producer organization standards are met.[6]

Farmers associated with the Fairtrade cooperative benefit from two price mechanisms: a guaranteed Fairtrade minimum price and a price premium for their certified crop, both set by the Fairtrade Labelling Organization (FLO). Whenever the international market price of a commodity is below the Fairtrade minimum price, farmers receive the Fairtrade minimum price, which covers the average cost of sustainable production. In addition, Fairtrade farmers benefit from a price premium, which is directly paid to Fairtrade cooperatives to be invested in community development in health, education, and the environment. At least 25 percent of the Fairtrade premium must be invested in the improvement of productivity and quality of the Fairtrade coffee. Farmers are expected to choose democratically how the premium is spent.[7]

In contexts in which child labor is widespread, Fairtrade standards encourage cooperatives to include a mitigation and elimination plan to prevent children from being employed whenever child work is identified as a risky activity (i.e. work that jeopardizes schooling or the social, moral or physical development of the person). Child work on the household farm is allowed, provided the work is appropriate to the children's age and it takes place outside of school hours or during school holidays (Fairtrade International 2015). Theoretical papers that directly address the economic consequences of child labor free labelling yield ambiguous effects on child labor participation, as the final impact depends on institutional features such as the intensity of the monitoring (Basu et al. 2006), the share of certified producers (Baland and Duprez 2009), and competition among certification agencies (Brown 2006).

## 3. Experimental Design

### 3.1 Study Area and Sample

---

[5] The definition of small-scale farmer applies to "smallholders who do not depend on hired workers all the time but run their farm mainly by using their own and their family's labor" (Fairtrade International 2012).

[6] The Fairtrade Labelling Organization sets products and small producer organization standards based on environmental agricultural practices, labor regulations, the empowerment of women, and democratic management and participation of cooperatives and their associates. Monitoring of standards is implemented as a three-stage process: self-assessment of producers against Fairtrade standards, peer reviews from trading partners, and random external inspections. It should be noted that Fairtrade certification is not free of cost, as small-scale organizations pay a non-negligible initial membership fee. The yearly certification fee ranges from 1,430 Euros for an organization with less than 50 members to 3,470 Euros for an organization with more than 1,000 members (FLO-CERT 2011). Initial certification is provided for a specific period of time (usually three years), to which it is subject to renewal by FLO-CERT conditional on the inspection according to Fairtrade's standards and regulations in the field and the payment of a renewal certification fee.

[7] In the case of coffee, the 2015-2017 Fairtrade minimum price was set to $1.4 dollars/pound of certified conventional washed Arabica coffee, while the price premium was set to $0.20 dollars/pound.

Our survey design experiment is carried out in two different regions in Ethiopia, Jimma, and Sidama. These regions produce two different varieties of coffee with high demand in international markets, Limu and Sidama. As Figure 1 shows, Jimma and Sidama are situated in the west- and south-central part of the country, within the preeminent coffee-producing areas of the country (Minten 2015). These two regions also represent two different cultural and social settings within Ethiopia: Jimma is a Muslim populated region (90 percent of households in our sample), while Sidama is an Orthodox Christian region (97 percent).

Our population framework is based on 5,100 smallholder farmers who are active members of four Fairtrade cooperatives that are spread out over 12 different districts (Kebeles).[8] Administrative data from these cooperatives is used for sampling purposes. Within each selected region, we selected two representative Fairtrade Coffee Cooperatives, one characterized as of 'high' productivity and the second as of 'low' productivity to improve the external validity of the sample. In Sidama, we worked with two Fairtrade cooperatives, each with around 1,500 active smallholder farmer associates. One cooperative reported a yearly average production of 1,122 kilograms of coffee crops per associate, while the other reported an average production of 789 kilograms in 2014. In Jimma, we also worked with two Fairtrade cooperatives. The first has 800 active farmer associates and a yearly average production of over 1,600 kilograms of coffee per associate in 2014, while the second cooperative has 1,100 active farmer associates and a yearly average production of 600 kilograms of coffee per associate in 2014. Our population framework, therefore, includes variation across several dimensions such as geography, cultural environments, and levels of production across smallholder farmers.

Sample selection is based on a 2x2 stratified random design. Stratification is based on two variables of interest: the level of production and the gender of the household head. We split the population of farmers into high- and low-production groups according to whether they were above or below household median coffee production in 2014. This information is collected directly from the administrative records of each Fairtrade cooperative. As variation in household coffee production can entail different combinations of adult/child work, this approach yields a sample that is representative of low and high coffee production household units. Likewise, we stratified the population data according to the gender of the household head as there may be gender differences in preferences and attitudes toward child labor. Information about the name of the household head was collected from administrative records from Fairtrade cooperatives and was validated in the initial field survey operation in July 2015. Summing across the resulting sub-samples within all four cooperatives we obtained a representative sample of 1,203 households.

The selection of experimental treatment and control households is based on complete randomization of the stratified sample. One-third of the sample is randomly allocated to the treatment group (self-response survey design), while the remaining two-thirds to the control group (proxy-response survey design). We differ from Dillon et al (2012) in that the selection of the proxy respondent is limited to the household head or the spouse thereof. This statistical

[8] Wayicho, Degara, Moto, Shilicho, Babe, Shabe, Qeway, Tassano, Haro, Omo Boqo, and Omo Gurude.

design yields 401 and 802 household units in the treatment and control groups, respectively. A statistical power analysis of a two-sample mean difference based on a two-sided 5 percent-level test and an effect size of 10 points shows a statistical power above the conventional threshold of 80 percent.[9]

### 3.2 Survey Design Treatment

Treatment status in this study is determined through application of the same survey instrument to randomly selected respondents, children themselves in the treatment group and proxy respondents in the control group. In one component of the household survey, field surveyors ask respondents to answer a specific labor-market module about child labor activities in the last 30 days before the survey (a child is defined as 6-14 years of age). In the presence of widespread seasonal activities and a labor force composed predominantly of casual and/or temporary workers, the choice of the length of the reference period is important. A short length of the reference period (e.g., a day or a week) may not capture seasonal work depending on the precise timing of the survey work if labor inputs vary considerably across weeks (Arthi et al 2017), or if the chosen day or week is atypical due to religious holidays or community celebrations (Matta-Greenwood 2000, Comblon and Robilliard 2017). A long length for the reference period (e.g., 'last year'), on the other hand, can introduce bias in the measurement of variables due to recall error. We, therefore, ask respondents about child labor in the 30 days prior to the survey, although we also elicit information for a shorter period, 'last week', as a robustness check for the main results. To avoid the possibility of bias in the manner in which information is elicited, there is no distinction in the wording nor in the sequencing of the questions across treatment and control groups.

The advantages of using proxy-based reports rather than child-based reports are not clear. Child-reported information may be more accurate than proxy responses if a child knows best how she allocates her time. That said, children can misreport their own labor status if there are not fully knowledgeable about what "work" entails or because if it is difficult to track their activities and hours over time. On the other hand, a proxy respondent like the household head may be familiar with the children's activities since many child laborers in developing countries work on the family farm or enterprise. However, due to social and cultural values, institutional (Fairtrade) standards, or asymmetries of information due to the gendered segmentation of labor markets, a proxy respondent may tend to underreport these activities. In any case, since smallholder farmers typically employ mostly family labor, there are no written records nor wage income to anchor recall which may affect measurement error (Arthi et al 2017). Therefore, we do not test whether children provide more accurate information compared to the one provided by proxies since we do not have the true value of work status[10].

[9] The size effect considered for power analysis is based on a conservative approach to the survey design results found by Dammert and Galdo (2013) in Peru and Janzen (2016) in Tanzania. Both studies have found self/proxy treatment effects for child labor statistics above 10 percentage points.

[10] Administrative information, a validation study, or a respondent debriefing study would be required to know the true classification of children's work. In our context, such measures do not exist.

As the measurement of child labor statistics could be affected by how child labor definitions are operationalized in the survey design questionnaire (e.g., Bardasi et al. 2012), we employ a relatively long rather than a short questionnaire design. Specifically, the survey contains 12 questions that aim to elicit information about their specific farm and non-farm labor activities. The most important questions refer to work at the household farm, as this activity accounts for most child work in rural Ethiopia (Guarcello and Rosati 2007). A typical labor question asks: "*Did [name] work any time on the household farm in the last 30 days?*" Since the keyword ´work´ can have different meaning for respondents, we supplement this question with a detailed, standard explanation of the concept of work by using a set of typical farming activities which is read aloud to the respondents: planting, watering, weeding, mulching, seedling, fertilizing, handpicking cherry coffee, cattle herding. This question is followed by an 'intensive margin' question, "*how many hours in the last 30 days did [name] spend working on the household farm?*"

The subsequent questions use the same length of the reference period to capture information on work in non-farm household business, Fairtrade Coffee Cooperatives, coffee plantations, other households' farms, work as an exchange laborer in other farms, and non-farm wage work. A negligible percentage (2%) of both self and proxy responses report children working outside the domain of household farm and, thus, this information is not used in the empirical analysis due to lack of variation. It is worth noting that household chores such as fetching water and/or firewood, house cleaning, cooking, and providing child and elderly care are left out of these work categories and are explicitly explained as activities that do not belong to farm work. This approach aims to provide children and proxy respondents a clear delineation of what constitutes work activities at the farm.

Surveyors were selected among seasoned field officers with relevant experience in the specific areas of intervention and who were very familiar with the local languages and customs to minimize potential misreporting problems due to linguistic and cultural barriers.[11] Two additional features of this experimental design require acknowledgment. First, as the survey questionnaire includes multiple sections that cover additional topics related to gender empowerment, trust, and knowledge of Fairtrade regulations, among others, the sequencing of the survey modules becomes important, as this may influence or shape the responses given by the heads of households regarding child labor activities. In this respect, we placed the labor market module as the second section in the survey, immediately after household demographics questions. Thus, concerns about survey-design-induced responses are minimized. Likewise, due to the nature of the treatment under analysis, and in contrast to other types of intervention in labor markets (e.g., training), neither proxy- nor self-respondents are aware of the survey variation and/or the exogenous manipulation of the (survey) environment. This minimizes bias induced by the presence of 'John Henry' effects or, more generally, 'Hawthorne' effects (List and Rasul 2013).

### 3.3 Timeline and Sample Attrition

[11] The official national language is Amharic, which is spoken by 30 percent of the population. Oromo and Sidamo are two additional local languages spoken in the areas of intervention.

Fieldwork for this study took place from July 2015 to January 2017. As the variation in work activities across the year depend on the agricultural calendar, we implement the same survey design experiment over the same households in three different coffee seasons coinciding with Ethiopia's rainfall seasons:

i) First survey: July-August 2015, during the *Meher* or main rainy season. This is the period of final coffee fruit development and sowing of other crops (Moat et al 2017, USAID 2015). This period is also known as the lean season due to the relatively low agricultural activity and its negative impact on agricultural income. This is also the time of year during which children are out of school.

ii) Second Survey: during April-May 2016, the *Belg* season or short rainy season. This is the period that corresponds to coffee planting, seedlings, weeding, and early development of the coffee fruit. Land preparation for other crops takes place as well[12].

iii) Third Survey: during December 2016-January 2017, harvest season or dry season, the busiest agricultural season for coffee-growing households (Dercon and Krishnan 2000). Red cherry coffee crops are harvested and coffee processing and selling take place.

As shown in Table A.1 of the appendix, levels of attrition are very low. Out of 1,198 households surveyed in 2015, only one household was not surveyed in early 2016 and 10 households were not surveyed in late 2016. Regarding children aged 6 – 14 years, we surveyed 1,890 children in the first survey. We observe some children dropping out of the sample in the second and third surveys due to reaching the age of 15 during that year (100 children) and other children not present in the first survey but returning to the household in later surveys (46 children).

Comparability of labor statistics across these three surveys follows from the similarity of the survey design, data collection and processing procedures. We used the same group of field surveyors across all rounds of data collection and the same software and personnel for data entry and coding. The wording and type of questions, the length of labor modules and the reference period are the same across the three survey rounds[13].

### 3.4 Sample Characteristics

Out of the 1,200 households that form the experimental sample, a total of 1,198 were interviewed in July/August 2015, including 406 treatment and 792 control household units. Table 1 provides summary statistics of characteristics associated with the household, head of households, and children, along with the p-values of the t-test for the equality of means between

[12] This period is also known as the short rainy season. Few plots have a bi-annual cropping pattern of crop growing during the rainy and short rainy seasons (Minten et al 2017).

[13] Child labor questions related to non-farm remunerated work, non-farm household business work, remunerated work in Fairtrade cooperatives, coffee plantations, and other households' farm are not included in the second and third round of survey as the initial survey shows that only a negligible share of children aged 6-14 are involved in these types of activities.

the experimental treated and control groups. We present these statistics for the overall experimental sample and the sub-sample of households that have at least one child in the 6-14 age category, as the latter is the relevant sample used in the computation of treatment effects.

The average household is composed of six members. 57 percent of the households are Christian, and 43 percent are Muslim, which is consistent with the geographic distribution of the survey fieldwork. The average monthly household income is around 1200 birr (about US$52), which is somewhat higher with respect to income than the average farmer in Ethiopia. Nonetheless, our sample participants live in houses with poor infrastructures such as mud-based floors (70 percent), no electricity (78 percent), and no sanitary services (80 percent). Nearly two of every three households have access to mobile phones, which shows the high penetration of digital technologies in agricultural rural markets in Eastern Africa. The typical household head is male, 50 years of age, married and has 4 years of formal schooling. On average, there are 1.5 children aged 6 to 14 per household. These children are 10 years of age on average, with 2.5 years of formal schooling. These variables are balanced between the full and the restricted experimental treated and control groups as shown by the corresponding p-values.

Regarding the characteristics of household agricultural production and sales, one observes that, on average, the extent of the household plot is close to one hectare of which 58 percent is used for coffee production while the remaining land is mainly used for cultivation of enset and maize. On average, 430 kg of cherry coffee is bought by the Fairtrade cooperative at an average price of 9.5 birrs per kg. Farmers also process part of the coffee harvest at home and sell it as dried coffee (178 kg on average) mostly to private merchants at an average price of 24 birrs per kg. It is important to note that one kg of dried coffee is equivalent to 3 to 5 kg of red cherries (FAO 2014, Minten et al 2014) depending on quality.[14] The p-values reported in column 6 show that these variables are statistically balanced between the experimental treated and control groups.

In sum, observable characteristics are well balanced across the two groups (self- and proxy-respondents), which provides internal validity to our estimates as both groups are sampled from the same population.

### 3.5 Seasonality of Child Labor Statistics

Figures 2A and 2B display the extent of variation of child labor statistics in rural Ethiopia across three different agricultural seasons. Regardless of the type of respondent, and at the extensive margin, we observe that the proportion of children working in farming activities in the past 30 days increases from 45 percent in the main rainy season to 52 percent in the short rainy season to 76 percent in the harvesting season. Figure 2A also depicts important gender gaps in child labor participation with boys having higher participation rates than girls. Although these gender

[14] Standard conversion rates show that for their dried coffee, farmers receive an average price that ranges from 5 to 8 Birrs per cherry kg which is lower than the price they would have received if the coffee was sold as cherry coffee. This finding is in line with Macchiavello and Morjaria (2015)'s study on Rwandan coffee growers, where cherry coffee was sold, on average, at a 40% higher price than processed dried coffee.

differences are expected, they show strong variation across different agricultural seasons. In the main rainy season, the boy/girl child labor gap reaches 20 percentage points (55 vs 35), which decreases to 17 percentage points (60 vs 43) in the short rainy season and to 9 percentage points (80 vs 71) in the harvesting season. This indicates that intrahousehold time allocation varies strongly according to the seasonality of the household farm activities.

At the intensive margin, we observe the same patterns as the (unconditional) number of monthly hours of work increases from an average of 16 hours during the main rainy season to 20 hours during the short rainy season and to 35 hours during the harvest season. Similarly, the boy/girl gap in monthly worked hours changes from 11 hours (22 vs 11) in the main rainy season, to 12 hours (26 vs 14) in the short rainy season, and to 7 hours (39 vs 32) in the harvesting season. In sum, the evidence indicates that the timing of survey data collection is of paramount importance in agricultural settings for child labor statistics and policy recommendations based on those statistics.

How do these child labor statistics fare relative to national, representative estimates in rural Ethiopia? To answer this question, we rely on the 2015/2016 Ethiopian Socioeconomic Survey (ERSS) that was implemented by the Central Statistics Agency (CSA) as part of the Integrated Surveys on Agriculture program in close collaboration with the World Bank. The ERSS data is a representative survey of 4,954 households living in rural areas, small-town, and medium and large-sized towns. The reference period for the child labor questions is the last 7 days before the survey, the age cohort refers to children 7 to 14, and the data collection took place between February and April 2016, a timing that overlaps with our second survey intervention (short rainy season). The labor module is answered by children aged 10 or older, while guardians/caretakers answer the questionnaire on behalf of children age 7 to 9 years.[15] With these caveats in mind, we proceed to the comparison of child labor statistics after restricting the ERSS sample only to rural households. Results show that child labor statistics from a rural representative survey in Ethiopia are comparable to the numbers that emerge from our sample of Fairtrade coffee farmers interviewed during the short rainy season: 52.4 percent of children aged 7 to 14 works in the household farm. This data also reports important gender gaps, with 58 percent of boys and 46 percent of girls working on the household farm.

## 4. Survey Design Experimental Results

The estimation of survey design treatment effects for child labor measures is based on a reduced-form linear regression model for individual *i,* in Fairtrade cooperative *j* at season *t,*

$$y_{ijt} = \alpha + \beta D_{ij} + X_{ijt}'\gamma + c_j + \tau_t + \varepsilon_{ijt}, \qquad (1)$$

where the parameter of interest, $\beta$, represents the average impact of survey design on child labor status $y_{ijt}$ at the extensive margin, $D_{ij}$ denotes the treatment indicator that is fixed over

[15]Available at http://siteresources.worldbank.org/INTLSMS/Resources/3358986-1233781970982/5800988-1367841456879/9170025-1367841502220/ERSS_Manual_HH_AG_COM_APRIL_30_ENGLISH.pdf

the three seasons and equals 1 for individuals aged 6 to 14 whose labor status is self-reported and 0 for individuals aged 6 to 14 whose labor status is reported by the proxy respondent. $X_{ijt}$ denotes a set of socio-demographic covariates of children and household heads that are deemed important determinants of child labor in the literature (e.g., Basu and Van 1998, Edmonds 2009). A Fairtrade cooperative fixed-effect, $c_j$ , is also included in the regression models to control for potential managerial differences that could be correlated with the outcome of interest, while $\tau_t$ is the season fixed effects and $\varepsilon_{ijt}$ is the idiosyncratic mean-zero error term. A Tobit specification of Equation (1) is implemented at the intensive margin as the dependent variable 'hours of work' has limited support with a mass point of zero, which could bias the point estimates under standard OLS specifications (Angrist 2001)[16]. Since we target only household heads as proxy respondents, we excluded from the computation of the treatment effects other family members who acted as proxy respondents (3 percent of respondents).

Panels A and B of Table 2 report the experimental findings at the extensive and intensive margins. The first column reports the pooled regression-based estimates while columns 2, 3, and 4 report the estimates for each survey experiment separately. Robust standard errors are reported in parentheses.

Several findings emerge from this table. The pooled data in column 1 shows statistically significant variation by the type of respondent in child labor statistics. Self-reported measures yield 4 percentage points higher child labor participation rates compared to proxy-reported measures. This result, albeit smaller in magnitude, is in line with non-experimental studies from Dammert and Galdo (2013) and Janzen (2016), who showed large underreporting of child labor participation by proxy respondents relative to the child's own responses. Importantly, column 1 also shows that variation is child labor statistics is greatly affected by the gender of the child. While statistically significant self/proxy reporting gaps of 7.6 percentage points are observed for girls, negligible and statistically not significant impacts are observed for boys. This result suggests that the systematic underestimation of women's work in agricultural settings reported in the literature (e.g., Anker 1983, Mata-Greenwood 2000), is also observed when measuring the work of girls.

When turning our attention to the seasonality of the survey design in columns 2-4 of Table 3, one observes important variation in the self/proxy reporting of child labor statistics. While no statistically significant self/proxy effects are found for boys across all three seasons, large and statistically significant impacts for girls are found in the main rainy season (10.3 percentage points) and in the harvesting season (7.1 percentage points). In the short rainy season, however,

---

[16] We expect that the results at the extensive margin are less prone to measurement error, and therefore, we place more weight in our discussion on that set of estimates. One potential caveat with the results at the intensive margin is the possibility that the point estimates are affected by recall error since children and proxy respondents are asked about labor activities performed over the previous 30 days of the survey. Thus, results at the intensive margin should be interpreted with caution.

though the self/proxy gap for girls reaches 5 percentage points, this is not a statistically significant result. Table A.2 in the appendix shows that these results at the extensive margin hold when the length of the reference period is a week rather than a month. In the short rainy season, the child/proxy reporting difference is not statistically significant different from zero for boys and girls, while in the harvest season we observe statistically significant point estimates for girls and equal to 7.3 percentage points.[17]

These results highlight the importance of replicating survey design experiments in agricultural contexts in which the nature of the economic activity is highly seasonal. Had we implemented only one survey design experiment and picked the short rainy season for that purpose, we would have concluded that respondent type is not a source of variation in child labor statistics. Our approach is indeed an effort to provide 'credibility enhancing arguments' (Rodrick 2009) to survey design interventions to improve the (external) validity of our findings and inform policy.

The point estimates emerging from Table 2 suggest that child/proxy information gaps are observed in agricultural seasons in which monitoring of child labor is relatively more costly. In the main rainy season, children do not follow the daily school routine as the timing of the survey coincides with the school year vacation[18]. In addition, this period corresponds to the lean season in which farm income is scarce and, therefore, heads of household might allocate effort to other off-farm activities to supplement income. In the coffee harvest season, which only occurs once per year in Ethiopia, the household labor demand is at its peak (Dercon and Krishnan 2000). This is the time of the year in which household heads are actively engaged in picking, transporting, selling, storing, and drying coffee cherry crops. This period is also characterized by active social festivities in coffee towns as coffee farmers receive cash windfalls from selling their cherry-coffee crops. The short rainy season, on the other hand, is a relatively quiet period during which children are in school and the main agricultural activities consist of weeding, land preparation, and planting. These tasks imply routine steps with predictable timing and involve the participation of woman and children (Admassie and Bedi 2008).

These findings on the extensive margin can be further analyzed along the intensive margin. Panel B in Table 2 show that children self-report higher monthly hours worked compared to proxy respondents in our pooled data (2.1 monthly hours' difference), main rainy season (3.1 hours) and harvest season (2.9 hours). Consistent with results at the extensive margin, these positive survey design treatment effects are driven by the subsample of girls, while statistically not significant impacts are observed for boys across all seasons. Furthermore, and unlike results at the extensive margin, Table A2 in the appendix shows that the length of the reference period matters at the intensive margin as the point estimates show variation depending on whether one uses one week rather than a month as the reference period. While in the short rainy season all point estimates lack precision, the child/proxy reporting differences become statistically significant at the 1 percent level for boys and girls in the harvest season.

[17] This additional information was not collected in the main rainy season.
[18] 80 percent of children aged 6-14 attend formal school in our sample. No differences are observed for school enrollment between boys and girls.

As our results show that heads of households tend to underreport child work statistics relative to the child's own account (or, alternatively, that children tend to over-report their work status relative to that of proxy respondents), we next assess whether self/proxy differences in reports are driven by the child's age, since it is acknowledged in the literature that data quality based on self-reports may improve with the age of children due to the continuing development of cognitive and communicative skills (Borgers et al. 2000). Table 3 reports the point estimates for the pooled data at both the extensive and intensive margins. Results show that the same proxy underreporting is observed for younger and older children, with the point estimates showing similar statistically significant differences of 4.1 and 4.2 percentage points for children aged 6 to 9 and children aged 10 to 14. Consistent with the results at the extensive margin, there is little difference in the number of monthly hours worked between the subsamples of children aged 6 to 9 and 10 to 14, as the child/proxy gaps are 1.8 and 2.3 hours, respectively. Moreover, it is important to note that within the girls' subsample self/proxy differences are similar between the younger subsample and the older one at both the extensive (7.9 vs 7.5 percentage points) and intensive (2.6 hours) margins. Regardless of the age group of the children, it is the girl subsample that shows statistically significant survey desing impacts. All in all, these numbers do not support the idea that our self/proxy differences in child labor statistics are driven by the age of the children.

Is the proxy underreporting of girls' work by proxy respondents inherent to child labor statistics? To answer this question we implement a form of placebo test for estimating self/proxy reporting differences in two important child-related dimensions: household chores and schooling. Household chores (i.e. fetching water, firewood, house cleaning, cooking, laundry, childcare and elderly care), are performed mostly within the household premises and are therefore relatively easy to monitor as they involve routine activities regardless of the seasonality of agricultural production. Schooling is a well-defined activity that follows a formal yearly schedule with around 5 hours of daily classes during the school year in our rural areas of intervention. Table 4 reports the point estimates for school participation and household chores in the short rainy season and harvesting season[19]. The point estimates are close to zero for both boys and girls and no statistically significant differences between the child and the proxy reports emerge except in the harvest season for the subsample of girls for the school enrollment variable. These results suggest that child labor statistics derived from surveys are inherently affected by respondent type, particularly in settings where agricultural work has a highly seasonal component and where social norms regarding female work is a salient feature. The next section discusses and tests potential mechanisms that may explain gender differences in child labor statistics.

## 5. Understanding Child Labor Survey Design Variation in Agricultural Settings

### 5.1 Gender Division of Labor

[19] We did not collected information from both the proxy and the child on household chores and school enrollment in the first household survey (main rainy season) due to time and budgetary constraints as this initial survey included several additional modules related to other topics.

It is possible that gender differences in misreporting reflect different responsibilities and tasks that males and females perform on the family farm and the degree of complementarity of tasks performed by adults and children of the same gender (Foster and Rosenzweig 1996, Fafchamps et al 2009). In the context of coffee-producing households in Ethiopia, as reported by Fafchamps et al (2009) and Lim et al (2007), men are mainly responsible for coffee farming (harvesting and maintenance of coffee plants), while women contribute to coffee harvesting and are responsible for household chores.

This gender division of labor might be reflected in child work on the farm and evolve with the age of the child. Several studies indicate that younger children spend most of their time with their mothers and perform tasks that tend to be performed by women (Nkamleu and Kielland 2006). As children age, children increasingly take on the tasks that are typical for the adults of their gender. For instance, while picking and weeding are activities performed by both boys and girls, pruning and spraying demand more physical strength and the ability to endure long hours of work, and are thus performed mostly by boys (ILO 2004).

Motivated by this discussion in the literature and supported by the fact that we stratified our random sample by the gender of the household head, we assess the role of respondent gender in child labor statistics by pairing the gender of the proxy respondent and the child. Table 5 shows the child/proxy gap point estimates following the same specification and econometric details as before, but this time over four different subsamples: boy/male, girl/male, boy/female, and girl/female. A caveat of this analysis is the small share of female proxy respondents relative to male ones (1/10 ratio), which adds large variability to the results due to small sample sizes.

The salient feature that emerges from this table is that underreporting of child labor statistics by the proxy respondent, relative to the child' own report, is driven by the gender mismatch between child and proxy respondents. For instance, by looking at the pooled data, the difference in child labor statistics of girls is 7.3 percentage points when the proxy is male and 18.2 percentage points for boys when the proxy is female. No statistically significant differences are observed for either boys or girls when the proxy/child gender is matched.

When splitting the data by survey season, we observe more variability in the point estimates particularly for the female proxy subsamples due to small sample sizes, though the gendered role of the proxy respondent holds. In the main rainy season, we observe statistically significant differences in child/proxy reports of 33.9 percentage points for boys when the proxy is female and 10.5 percentage points for girls when the proxy is male. No statistically significant differences are observed whenever there is a match in the child/proxy gender. In the short rainy season, the point estimates are consistently higher for the gendered mismatch subsamples than that for the matched ones particularly for the boy/female subsample that show large and statistically significant impacts. In the harvest season, the busiest agricultural season for coffee-growing households, the estimates show a statistically significant difference (6.6 percentage points) in the reports between girls and their male proxies.

When turning our attention to the intensive margin in Panel B of Table 5, we observe the same pattern of results. Statistically, significant underreporting of child labor statistics by the proxy respondent, relative to the child' own report, is driven by the gender mismatch between child and proxy respondents. No statistically significant differences are observed, on the other hand, when the gender of the proxy respondent and the child is matched. Overall, and similar to the results on the extensive margin, the magnitude of the child/proxy reports gap is substantive for boys when the proxy respondent is a female.

In fact, and setting aside the gender of the child, an important insight of Table 5 is the consistent result that, relative to male proxy respondents, female household heads show higher child/proxy gaps in child labor statistics. The pooled data, for instance, shows underreporting of child labor statistics of 3.4 and 11.5 percentage points for the male and female proxy samples, both statistically significant at the 1 and 5 percent levels. At the intensive margin, these differences reach 1.6 and 7.9 hours respectively. This result is driven by the composition of the small female proxy sample that are exclusively composed by household heads who are mostly widowers, separated, or with husbands residing in different locations, and therefore, women in charge of overseeing alone all productive and non-productive activities within the household. This result is also consistent with the broad evidence that farm production is mainly the occupation of the male household heads who oversee and monitor work on the household farm more closely (Bass 2004). To support this argument, we elicit information on the weekly distribution of worked hours for both male and female proxy respondents in the first round of survey design intervention (main rainy season). On average, while male proxy respondents spend 25 hours per week in farm work, female proxy respondents spend only 14 hours. On the other hand, males only spend 9 weekly hours in household chores, while females spend 41 weekly hours. For other types of activities such as religious and social activities, non-farm household business, and non-farm work, there are no significant differences in the weekly time allocation between male and female proxy respondents.

While we do not have information about monitoring mechanisms and incentives, we investigate the offspring gender composition of the household and child/proxy information gaps. If monitoring or knowledge of child labor activities is affected by the gender stratification of work and social lives in Ethiopia, then one would expect to observe salient child/proxy reporting gaps within households that have mixed gender composition of children, relative to households that have homogenous gender composition, as the former would entail different intrahousehold allocation of child activities. We then split the data between these two types of households and report the results in Table 6 following the same specification model given in equation (1). Results for the pooled data in column 1 show statistically significant child/proxy information gaps in child labor statistics for girls (8.8 percentage points) but not for boys in households with mixed gender composition of children. By looking at the point estimates across different seasons in columns 2-4, we observe statistically significant effects for girls in the main rainy season (9.2 percentage points) and harvest season (10.9 percentage points). No statistically significant child/proxy gaps are uncovered for households with similar gender composition of children across all three survey seasons. Unreported similar pattern of results is observed on the intensive margin as well.

Furthermore, we asked household heads questions regarding time allocation decisions to ascertain the degree of control and information within the household[20]. We recorded whether the proxy respondent self-reported as being the only decision maker or strongly influencing the decisions about farm child labor and household chores. We then estimated the differential child/proxy treatment effects on child labor statistics with respect to the base category in which the proxy respondent self-reported sharing equally in decision-making with the spouse or not influencing these decisions at all. It has been documented that resources controlled by different members of the household affect labor allocation decisions within the households. For instance, Udry (1996) finds that child labor is used less intensively on plots controlled by women than on similar plots controlled by men in Burkina Faso. Oseni et al (2015) find that male managed plots are more likely to use male family labor compared to female-managed plots (and vice versa) in Nigeria. Thus, given the gender stratification of labor in Ethiopia, it is possible that asymmetries of information are exacerbated by the gender stratification of work and social lives in Ethiopia, as adult males may be able to observe boys' output and activities, while adult females may have more information about girls' work.

Table 7 shows the differential impact of household decision making on the child/proxy gaps in child labor statistics. Whenever the male head of household is the only decision maker or strongly influence the work time allocation of girls, underreporting of child labor statistics by the proxy respondent, relative to self-reporting, decreases significantly. It is likely that male proxy respondents are highly knowledgeable about a girl's schedule and activities, particularly when the former has sole responsibility for making decisions on the time allocation of the child. In fact, we observe that these negative differential impacts are significantly larger for girls than that for boys: we observe a differential reduction of the child/proxy gap of 20 and 10 percentage points when a male proxy respondent assigns the time allocation of girls to farm and household chores. Overall, these results point in one direction: variation in child labor statistics due to the type of respondent is significantly lower in agricultural settings whenever the (male) proxy respondent is actively engaged or more familiar with the girls' activities and work schedule. [21,22]

If the gender of the child and proxy respondent matters for child labor measurement, what does this suggest from a policy standpoint? One implication of our findings is that the large underreporting of girls' work by male proxy respondents, relative to the child' report, may decline

[20] This data is available only in the main rainy season survey.

[21] One alternative channel we assess is the impact of survey fatigue on the differences in child labor reports by randomly manipulating the order of an additional time discounting module in the survey design questionnaire. Female proxy respondents could be busy with household chore activities when receiving the visit of the surveyor at home and, thus, they may tend to rush through questions. One subset of the sample is randomly assigned to answer the time discounting module early in the main rainy season survey and right after the labor module is applied, while the remaining subsample answer the same time discounting module at the end of the household survey which contains more than 12 different modules. The survey fatigue random assignment is orthogonal to the type of respondent survey design as a cross-randomization strategy is used. Results shows negligible and statistical not significant impacts of survey fatigue in our sample.

[22] One may also wonder whether the head of household do not have information on children's activities due to temporarily migration. This is not the case in our sample of smallholder farmers since migration rate is very small (1 percent).

if the spouse of the head of household provides information on girls' activities as well. Thus, to further understand gender differences in child labor reports, we elicit information about farm labor participation for each child from two proxies, household heads and their spouses across all control group households (proxy-information).[23] It is worth noticing this new approach expands the sample of female proxy respondents across all households in the control group.

Results reported in Table 8 show that the child/proxy reporting gap in child labor statistics significantly decreases when the spouse of the household head provides the information on girl's employment in farm activities: the previously reported statistically significant child/proxy report gap of 7.1 percentage points for girls declines to 2.4 percentage points and loses statistical significance when the proxy respondent is the spouse of the head of household.[24]

### 5.2 The Role of Fairtrade Standards

There is no social stigma associated with child labor in Ethiopia, as is the case in all eastern Africa (see Bass 2004). Yet, by focusing on smallholder farmers that are active members of Fairtrade cooperatives, it is possible that our estimates might reflect a set of specific labor standards that affect behavior or incentives when reporting child work. For instance, Fairtrade does not ban children's engagement in the household farm but instead regulates it by allowing children to work in family farms as long it is outside school hours and free of risky activities. In this section, we investigate differential impacts of self- and proxy-based reports of child labor due to variations in farmers' knowledge of Fairtrade standards and the degree and scope of informal relational contracts between farmers and Fairtrade cooperatives. We acknowledge these Fairtrade variables can be correlated with the error term in the OLS estimation framework, and therefore, interpretation should proceed with caution.

We first assess whether knowledge about Fairtrade regulations and standards on the part of the proxy respondent has a differential impact on child labor reports. We gathered information about Fair Trade standards related to labor, environmental regulations and Fairtrade pricing and computed a knowledge index using principal component analysis. The higher the value of the index, the higher the overall knowledge of Fairtrade standards and regulations[25].

---

[23] Due to budgetary constraints, this additional information is collected for the harvest season survey only.

[24] One additional benefit of having two proxy measures of child labor status for every child in the proxy control subsample is that one can implement an ex-post survey random assignment that randomly selects one report for each child, either from the male proxy subsample or from the female proxy subsample. This strategy provides an opportunity to compare the child report against the male proxy report and the female proxy after randomly generating the male and female proxy groups. By repeating this ex-post random assignment 1000 times, one can estimate child/proxy treatment effects with the added value of randomly assigning the gender of the proxy respondent. These unreported set of estimates are quite similar the results shown in Table 8.

[25] Questions about Fair Trade knowledge were asked in the first survey round (June – July 2015). Households with response of "don't know" were treated as missing observations.

Table 9 shows the differential impacts after interacting the type of respondent treatment variable with the knowledge index variable. If proxy respondents answer the child labor module strategically, one would expect they would be more prone to underreport their children's work activities, and therefore, we would obtain positive and significant differential effects. The upper panel shows no differential treatment effects related to Fairtrade knowledge of regulations and standards in the pooled data. When splitting the data across agricultural seasons, the point estimates for the interaction term are uncertain: we observe positive significant impacts at the 10 percent level for boys in the main rainy season and negative significant impacts at the 5 percent level for girls in the harvest season.

Still, it is possible that Fairtrade farmers have limited knowledge about what Fairtrade entails even though they belong to Fairtrade certified cooperatives (Valkila and Nygren 2010), or that their self-reported knowledge of standards may be affected by recall error. Thus, we also assess instead the extent of effective 'relational contracts' between Fairtrade cooperatives and smallholder farmers as a source of variation in child labor reporting. The term 'relational contracts' refers to the degree and scope of effective trade links pre and post-harvest between cooperatives and their smallholder associates in the absence of contractual enforcement (MacLeod 2007, Macchiavello and Morjaria 2015). In our sample, for instance, the share of coffee that each farmer associate sells on average to Fairtrade cooperatives is 57 percent, only 6 percent of farmers are part of cooperative boards, while 37 percent of farmers did not exercise their right to vote in the last cooperative democratic election. Indeed, the extent of relational contracts is relevant as not every associate farmer benefits equally from Fairtrade cooperatives, given that there is considerable variation in the share of certified coffee production that each farmer associate sells to Fairtrade cooperatives (Janvry et al 2015).

The lower panel of Table 9 reports estimated coefficients after including a term interacting treatment status and the (normalized) relational contract index[26]. By looking at the point estimates, we only observe positive statistically significant differential effects for girls in the main rainy season.

All in all, these findings suggest that knowledge of Fairtrade standards and the degree and scope of effective commercial links between farmers and Fairtrade cooperatives do not have a systematic differential effect on the proxy reporting of child labor status relative to the child's own report.

## 6. Conclusions

This study addresses the measurement of child labor in agricultural settings with particular attention to seasonality, respondent type, and gender-based information gaps. In an effort to

[26] Specifically, we use in the computation of the relational contract index the following variables: share of red-cherry coffee sold to Fairtrade cooperatives (in kilos), membership to cooperative board, having voted in last cooperative election, attendance at cooperative meetings, walking distance (in minutes) from home to the closest cooperative collection center and to cooperative administrative center. For ease of interpretation, we normalized this relational contract index by using the mean and standard deviation of the control group.

provide 'credibility enhancing arguments' (Rodrick 2009), we implemented three similar survey design experiments among Fairtrade coffee-growing households in Ethiopia in three different agricultural seasons. The survey design experiments randomly assign each of 1200 households in our sample to one of two groups: the child (treatment) and the proxy (control) reporting groups. Stratification of the sample by the gender of the household head allows us to estimate child/proxy reporting gaps across four subsamples of interest: boy/male, girl/male, girl/male and girl/female. Regardless of the gender and type of respondent, this study uncovers substantial variation in farm child labor across three agricultural seasons with rates of participation ranging from 45 to 76 percent in the rainy and harvest season. This means that studies of child labor in agricultural settings should explicitly acknowledge and discuss the seasonality of their results for a better understanding of its determinants and conditions, and for the design of social protection programs and policy. Moreover, interventions in areas that are strongly affected by the seasonality of productive activities and income, should make efforts to incorporate this contextual reality in their design, implementation and in the analysis of results.

This study also uncovers statistically significant child/proxy reporting gaps in child labor statistics. Relative to the child's own report, underreporting by the head of household is particularly salient for girls but not for boys. The magnitude of this proxy underreporting for girls varies across different agricultural seasons: we found significant impacts in the main rainy season (9 percentage points) and harvest season (8 percentage points), yet insignificant effects in the short rainy season. This has a clear implication for the 'external validity' of survey design randomized interventions in developing settings: had we implemented only one survey design experiment and picked the short rainy season for that purpose, we would have erroneously concluded that respondent type is not a source of variation in child labor statistics. These survey design treatment effects are mainly explained by the gender mismatch between the child (girl) and the proxy (male) respondent. In fact, no significant impacts are observed for boys when the proxy respondent is male. This result suggests that the systematic underestimation of women's work in agricultural settings, commonly reported in the literature, also permeates the measurement of child labor statistics due to the gender stratification of work and social lives in East Africa that exacerbates asymmetries of information. Indeed, girls' work is seen as an extension of, and subordinate to, women's work**.**

Two additional pieces of evidence uncovered in this study also point in that direction. Firstly, the child/proxy measurement gaps are observed in the subsample of households with mixed gender composition of the children, while no statistically significant gaps are reported in the subsample of households with homogenous offspring gender composition. Secondly, the magnitude of child labor underreporting by the proxy respondent, relative to the child's report, decreases substantially when the proxy respondent is actively engaged in the allocation of the child's time schedule in farm work.

Since the sample of farmers in this study belongs to Fairtrade coffee cooperatives, a plausible competing hypothesis is that knowledge of Fairtrade standards by smallholder associates, or the presence of effective relational links between farmers and cooperatives, could lead to strategic behavior by proxy respondents when answering the survey questions. After all, though Fairtrade does not ban child labor, it does aim to regulate it. We provide evidence against these competing hypotheses.

From a policy standpoint, and taking together all of these results, the main challenge to address seems to be the underreporting of the work of girls in the agricultural sector by male household heads who typically answer household surveys in developing settings. We propose eliciting this information from their spouses. We implemented this recommendation in the last round of survey and results show that the child/proxy information gaps on girls' child labor status is halved and loses statistical significance. This study calls for an expanded role for female proxy respondents in the application of survey questions related to girls' outcomes in areas that are intrinsically affected by gender segmentation in work and social lives.

**References**


Admassie, Assefa, and Arjun Singh Bedi. 2008. "Attending School, Reading, Writing and Child Work in Rural Ethiopia." In *Economic Reform in Developing Countries*. Chapter 6. Edward Elgar Publishing.

Anker, R. 1983."Female labour force participation in developing countries: A critique of current definitions and data collection methods". *International Labour Review*, 1226):709-23.

Arthi, V, K. Beegle, J. De Weerdt, and A. Palacios-Lopez. 2018. "Not your average job: Measuring farm labor in Tanzania" *Journal of Development Economics* 130:160-172

Bäckman, Tora. 2009. "Fair Trade coffee and development, a field study in Ethiopia". Department of Economics at the University of Lund. Minor field study series no.188

Baland, Jean-Marie, and Cédric Duprez. 2009. "Are Labels Effective against Child Labor?" *Journal of Public Economics* 93 (11–12):1125–30.

Bardasi, Elena, Kathleen Beegle, Andrew Dillon, and Pieter Serneels. 2012. "Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania." *World Bank Economic Review* 25 (3):418–47.

Banerjee A. V. 2005. "New Development Economics and the Challenge to Theory." In *New Directions in Development Economics: Theory or Empirics? A Symposium in Economic and Political Weekly*, edited by Kanbur R., August 2005.

Bass, Loreta.2004. "Child Labor in Sub-Saharan Africa". Lynne Rienner Publishers. London.

Basu, Kaushik and Tzannatos, Zafiris.2003. "The global child labor problem: what do we know and what can we do?" *The World Bank Economic Review.* Vol. 17 (2), pp. 147-173.

Basu, Arnab, Nancy H Chau, and Ulrike Grote. 2006. "Guaranteed Manufactured without Child Labor: The Economics of Consumer Boycotts, Social Labeling and Trade Sanctions." *Review of Development Economics* 10 (3):466–91.

Beegle, Kathleen, Jessica Goldberg, and Emanuela Galasso. 2017. "Direct and Indirect Effects of Malawi's Public Works Program on Food Security." *Journal of Development Economics* 128:1–23.

Bharadwaj, P. 2015. "Fertility and rural labor market inefficiencies: Evidence from India" *Journal of Development Economics*, 115:217-232

Borgers, Natacha, Edith de Leeuw, and Joop Hox. 2000. "Children as Respondents in Survey Research: Cognitive Development and Response Quality." *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 66:60–75.

Brown, Drusilla. 2006. "Consumer Product Labels, Child Labor, and Educational Attainment." *The B.E. Journal of Economic Analysis & Policy* 5 (1):1–27.

Comblon, Virginie, and Anne-Sophie Robilliard. 2017. "Are Female Employment Statistics More Sensitive than Male Ones to Survey Design? Evidence from Cameroon, Mali and Senegal." Working Paper DT/2015-22, DIAL.

Dammert, Ana C, Jacobus De Hoop, Eric Mvukiyehe, and Furio Rosati. 2017. "The Effects of Public Policy on Child Labor: Current Knowledge, Gaps, and Implications for Program Design." Policy Research. *World Development,* forthcoming.

Dammert, Ana C, and Jose Galdo. 2013. "Child Labor Variation by Type of Respondent: Evidence from a Large-Scale Study." *World Development* 51 (11):207–20.

Deaton A. S. and N. Cartwright. 2016. "Understanding and Misunderstanding Randomized

Controlled Trials." NBER Working Paper No. 22595, NBER, Cambridge, MA.

Dercon, Stefan, and Pramila Krishnan. 2007. "Vulnerability, Seasonality and Poverty in Ethiopia." *Journal of Development Studies* 36 (6):25–53.

Dillon, Andrew, Elena Bardasi, Kathleen Beegle, and Pieter Serneels. 2012. "Explaining Variation in Child Labor Statistics." *Journal of Development Economics* 98 (1):136–47.

Dragusanu, R., D. Giovannucci, and N. Nunn. 2015. "The economics of Fair Trade", *Journal of Economic Perspectives*, 28(3): 217-236.

Edmonds, Eric. 2009. "Child Labor." In *Handbook of Development Economics*. Vol. 4. Elsevier Science, Amsterdam, North Holland.

Fafchamps, Marcel, Bereket Kebede, and Agnes Quisumbing. 2009. "Intrahousehold Welfare in Rural Ethiopia." *Oxford Bulletin of Economics and Statistics* 71 (4):567–99.

Fairtrade International. 2012. "Unlocking the Power: Annual Report."

———. 2014. "Strong Producers, Strong Future: Annual Report."

———. 2015. "Fairtrade and Forced Labour."

FAO. 2015. "FAOSTAT Country Profiles."

FLO-CERT. 2011. "Fairtrade Standard for Small Producer Organizations."

Foster, Andrew and Mark Rosenzweig. 1996. "Comparative Advantage, Information and the Allocation of Workers to Tasks: Evidence from an Agricultural Labour Market," *Review of Economic Studies*, 63, 347-374.

Guarcello, Lorenzo, Irina Kovrova, Marco Manacorda, Furio Rosati, and Scott Lyon. 2010. "Towards Consistency in Child Labour Measurement: Assessing the Comparability of Estimates Generated by Different Survey Instruments." UCW Working Paper 54.

Guarcello, Lorenzo, and Furio Rosati. 2007. "Child Labor and Youth Employment: Ethiopia Country Study." SP Discussion Paper 0704, The World Bank.

ICO. 2006. "Guidelines for the Prevention of Mould Formation in Coffee." International Coffee Organization.

ILO. 2004. "Safety and Health Fact Sheet Hazardous Child Labour in Agriculture: Coffee." International Programme on the Elimination of Child Labour.

———. 2017. "Global Estimates of Child Labour. Results and Trends 2012-2016." International Labour Organization, Geneva.

IPEC. 2004. "Child Labour Statistics: Manual on Methodologies for Data Collection through Surveys." International Labour Organization, Geneva.

Janvry, Alain de, Craig McIntosh, and Elisabeth Sadoulet. 2015. "Fair Trade and Free Entry: The Dissipation of Producer Benefits in a Disequilibrium Market." *Review of Economics and Statistics* 97 (3):567–73.

Janzen, Sarah. 2016. "Child Labor Measurement: Who Should We Ask?" *International Labour Review* forthcoming.

Kifle, Abiy, Getahun Belay, and Almaz Beyene. 2005. "The Rapid Assessment Study on Child Labor in Selected Coffee and Tea Plantations in Ethiopia." International Labour Organization, Geneva.

Kruger, Diana. 2007. "Coffee Production Effects on Child Labor and Schooling in Rural Brazil." *Journal of Development Economics* 82 (2):448–63.

Lee, Jungmin and S. Lee. 2012. "Does It Matter Who Responded to the Survey? Trends in the U.S. Gender Earnings Gap Revisited." *Industrial and Labor Relations Review* 65:148–60.

Lim, Sung Soo, Alex Winter-Nelson, and Mary Arends-Kuenning. 2007. "Household Bargaining Power and Agricultural Supply Response: Evidence from Ethiopian Coffee Growers," *World Development*, vol. 35(7), pages 1204-1220.

List, John, and Imran Rasul. 2011. "Field Experiments in Labor Economics." In *Handbook of Labor Economics*, 4a, Chapter 2:103–228.

Macchiavello, Rocco, and Ameet Morjaria. 2015. "Competition and Relational Contracts: Evidence from Rwanda's Coffee Mills."

MacLeod, Bentley. 2007. "Reputations, Relationships and Contract Enforcement." *Journal of Economic Literature* 45 (3):595–628.

Matta-Greenwood, A. 2000. "Incorporating gender issues in labour statistics. ILO.

Minten, Bart, Mekdim Dereje, Ermias Engeda, and Seneshaw Tamru. 2015. "Who Benefits from the Rapidly Increasing Voluntary Sustainability Standards? Evidence from Fairtrade and Organic Coffee in Ethiopia." IFPRI-EDRI, WP 71.

Minten, Bart, Seneshaw Tamru, Tadesse Kuma, and Yaw Nyarko. 2014. "Structure and Performance of Ethiopia's Coffee Export Sector." IFPRI-EDRI, WP 66.

Nkamleu, Guy and Anne Kielland. 2006. "Modeling farmers; decisions on child labor and schooling in the cocoa sector: a multinomial logit analysis in Cote d' Ivoire", *Agricultural Economics,* 35: 319-333

Oseni, Gbemisola, Paul Corral, Markus Goldstein, and Paul Winters. 2015. "Explaining Gender Differentials in Agricultural Production in Nigeria." *Agricultural Economics* 46 (3):285–310.

Rodrick, D. 2009 "The new Development Economics: We shall experiment, but how shall we learn?" In *What Works in Development? Thinking Big and Thinking Small*, edited by Easterly W., Cohen J. 24 -54. Washington, DC: Brookings Institution Press.

Rosenzweig, Mark, and Chris Udry. 2014. "External Validity in a Stochastic World: Evidence from Low-Income Countries." Manuscript, Yale University.

Reynolds, Jeremy, and Jeffrey Wenger. 2012. "He Said, She Said: The Gender Wage Gap According to Self and Proxy Reports in the Current Population Survey." *Social Science Research* 41:392–411.

Udry, Chris. 1996. "Gender, Agricultural Production, and the Theory of the Household." *The Journal of Political Economy* 104 (5):1010–46.

Valkila, Joni, and Anja Nygren. 2010. "Impacts of Fair Trade Certification on Coffee Farmers, Cooperatives, and Laborers in Nicaragua." *Agriculture and Human Values* 27 (3):321–33.

World Bank. 2014. "World Development Indicators." Washington D.C.

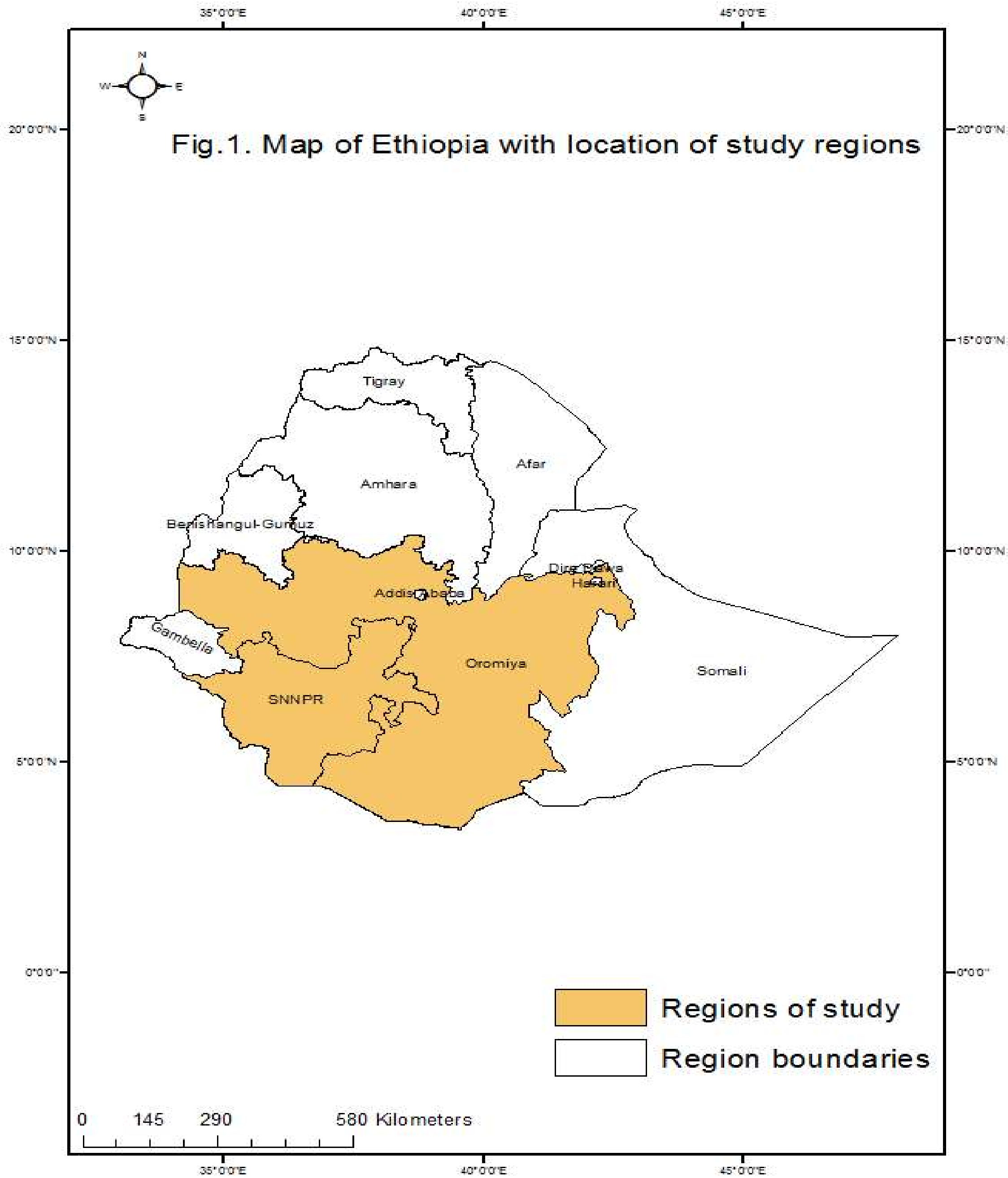
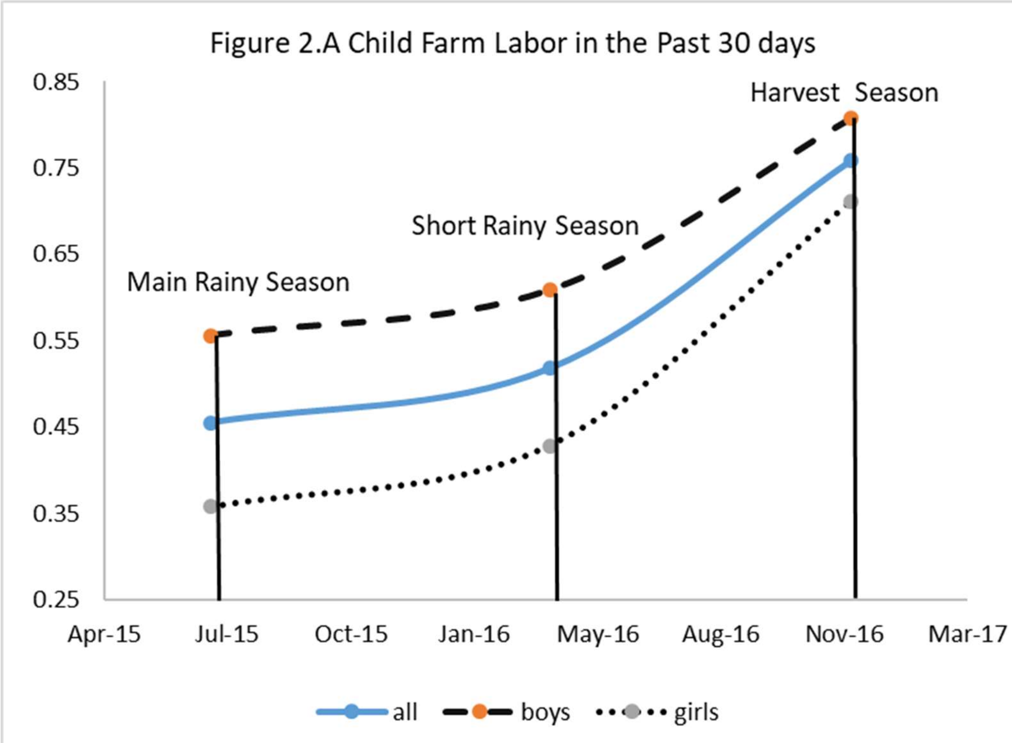



Fig.1. Map of Ethiopia with location of study regions

**Figure 2: Seasonality of Child Labor in Rural Ethiopia**



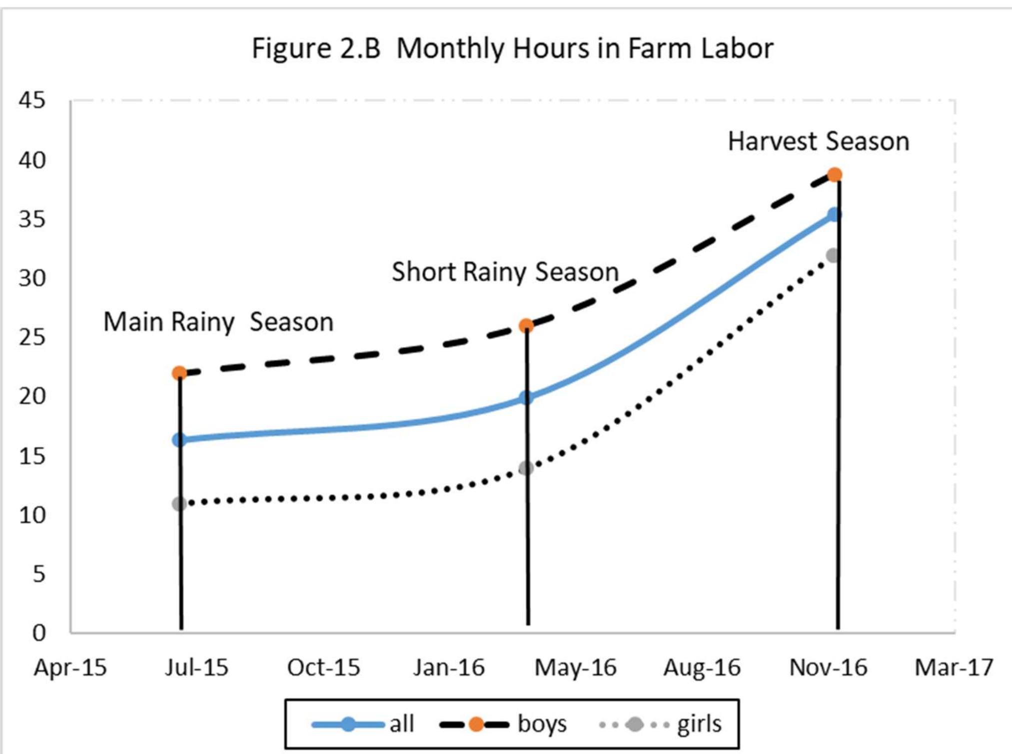



**Table 1: Balancing Test for Experimental Survey Assignment (July/August 2015 – Main Rainy Season)**

| | Full Sample | | | | Households w/ children aged 6-14 | |
|---|---|---|---|---|---|---|
| | Self-reported | Proxy-reported | Difference | p-value | Difference | p-value |
| **Panel A: Household socio-demographics** | | | | | | |
| Household size | 5.61 | 5.59 | 0.02 | 0.90 | -0.03 | 0.81 |
| Children aged 6-14 | 1.60 | 1.57 | 0.04 | 0.65 | 0.05 | 0.47 |
| Christian (%) | 0.57 | 0.55 | 0.02 | 0.57 | 0.01 | 0.81 |
| Muslim (%) | 0.43 | 0.44 | -0.01 | 0.66 | -0.01 | 0.88 |
| Main source for lighting is electricity/generator (%) | 0.23 | 0.22 | 0.01 | 0.63 | 0.03 | 0.33 |
| Mud floor (%) | 0.71 | 0.71 | 0.00 | 0.92 | 0.01 | 0.70 |
| Pit latrine not ventilated (%) | 0.81 | 0.77 | 0.03 | 0.20 | 0.07 | 0.02 |
| Owns a mobile phone (%) | 0.67 | 0.65 | 0.02 | 0.51 | 0.03 | 0.37 |
| Last month total income (Birr) | 880.13 | 819.83 | 60.30 | 0.61 | 114.04 | 0.44 |
| Yearly average monthly income (Birr) | 1266.41 | 1258.54 | 7.87 | 0.93 | -8.95 | 0.94 |
| Walking distance from house to closest prim school (min) | 22.39 | 21.74 | 0.65 | 0.47 | 0.80 | 0.44 |
| Walking distance from house to closest sec school (min) | 70.17 | 72.02 | -1.84 | 0.41 | 0.61 | 0.81 |
| **Panel B: Head of Household** | | | | | | |
| Gender (% Male) | 0.89 | 0.87 | 0.02 | 0.24 | 0.01 | 0.53 |
| Age | 50.22 | 49.65 | 0.57 | 0.53 | 0.41 | 0.65 |
| Years of schooling | 3.92 | 3.82 | 0.11 | 0.63 | 0.09 | 0.73 |
| Married | 0.86 | 0.84 | 0.02 | 0.44 | 0.01 | 0.59 |
| **Panel C: Children aged 6-14** | | | | | | |
| Gender (% Male) | | | | | 0.01 | 0.69 |
| Age | | | | | 0.14 | 0.25 |
| Years of schooling | | | | | 0.05 | 0.65 |
| Currently attending school | | | | | -0.01 | 0.44 |
| Illness/injury in the past 12 months | | | | | 0.00 | 0.99 |

**Note:** Sample means computed from the first survey design experiment carried out in July 2015. P-values refer to the null hypothesis of equality of means between self-reporting and proxy-reporting measures

**(continue) Table 1: Balancing Test for Experimental Survey Assignment (July/August 2015 – Main Rainy Season)**

| | Full Sample | | | | Households w/ children aged 6-14 | |
|---|---|---|---|---|---|---|
| | Self-reported | Proxy-reported | Difference | p-value | Difference | p-value |
| **Panel D: Household Agricultural Output** | | | | | | |
| Land size (hectare) | 1.11 | 1.06 | 0.05 | 0.49 | 0.04 | 0.63 |
| Coffee cultivated % total area | 0.58 | 0.58 | 0.00 | 1.00 | 0.00 | 0.79 |
| Share of red cherry sold to Fairtrade Coop (%) | 94.47 | 94.55 | -0.08 | 0.94 | 0.03 | 0.98 |
| Total sales of cherry coffee to Fairtrade Coop (kg) | 428.82 | 433.96 | -5.14 | 0.88 | -22.06 | 0.53 |
| Total sales of dried coffee to Fairtrade Coop (kg) | 182.68 | 171.72 | 10.96 | 0.66 | 4.52 | 0.88 |
| Yield of coffee per hectare | 3323.31 | 3273.56 | 49.75 | 0.83 | 100.09 | 0.72 |
| Selling price of cherry coffee (Birr per kg) | 9.53 | 9.50 | 0.03 | 0.93 | 0.04 | 0.92 |
| Selling price of dried coffee (Birr per kg) | 23.72 | 23.56 | 0.16 | 0.73 | -0.04 | 0.94 |
| | | | | | | |
| N Households | 406 | 792 | | | | |
| N Children aged 6-14 | 640 | 1240 | | | | |

**Note:** Sample means computed from the first survey design experiment carried out in July 2015. P-values refer to the null hypothesis of equality of means between self-reporting and proxy-reporting measures.

**Table 2: Survey Design Average Treatment Effects**

| | Pooled Experiments | Main Rainy Season (July - August 2015) | Short Raining Season (April - May 2016) | Harvest Season (Dec 2016 - Jan 2017) |
|---|---|---|---|---|
| **Panel A: Participation in Household Farm Activities over the Past Month** | | | | |
| All | 0.040*** | 0.065*** | 0.027 | 0.026 |
| | (0.013) | (0.023) | (0.021) | (0.020) |
| Boys | 0.005 | 0.022 | 0.004 | -0.016 |
| | (0.018) | (0.034) | (0.029) | (0.027) |
| Girls | 0.076*** | 0.103*** | 0.050 | 0.071** |
| | (0.019) | (0.032) | (0.031) | (0.029) |
| **Panel B: Hours Spent on Household Farm Activities (Tobit) over the Past Month** | | | | |
| All | 2.076*** | 3.159*** | -0.029 | 2.913** |
| | (0.744) | (1.155) | (1.116) | (1.431) |
| Boys | 1.183 | 1.860 | -0.980 | 3.003 |
| | (1.192) | (1.973) | (1.829) | (2.134) |
| Girls | 2.706*** | 3.636*** | 0.723 | 2.977 |
| | (0.907) | (1.262) | (1.301) | (1.896) |

**Notes:** Probability linear models on work status in the past 30 days from survey date (Panel A) and Tobit marginal effects on weekly hours worked in past 30 days from survey date (Panel B). Robust standard errors in parenthesis. Control covariates include child age and gender, gender of head of household, dummy variables for schooling of the head of household, household size, dummy variables for quartiles of household wealth, dummy variables for Fairtrade Cooperatives to which families belong. Pooled data estimation also includes dummy variables for specific survey season. Household wealth index is estimated by the first component of Principal Component Analysis which includes information from 13 variables related to house infrastructure (own house, per capita number of rooms, source of drinking water, type of sewage, type of floor, type of roof, electricity), and household assets (radio, TV, oven, bicycle, motorcycle, telephone). Sample sizes reported in Table A.1
*10%, **5%, ***1% statistical significant levels

**Table 3: Survey Design Average Treatment Effects by Age of Children (Pooled data)**

| | Work in household farm in past month | Monthly hours of work in household farm |
|---|---|---|
| **Panel A** | | |
| Age 6-9 | 0.041** | 1.821* |
| | (0.021) | (1.033) |
| Boys | 0.008 | 0.903 |
| | (0.031) | (1.716) |
| Girls | 0.079*** | 2.680** |
| | (0.028) | (1.212) |
| **Panel B** | | |
| Age 10-14 | 0.042*** | 2.334*** |
| | (0.016) | (1.040) |
| Boys | 0.008 | 1.785 |
| | (0.021) | (1.639) |
| Girls | 0.075*** | 2.626** |
| | (0.025) | (1.303) |

Notes: Probability linear model on work status in the past 30 days from survey data (Panel A) and Tobit marginal effects on monthly worked hours (Panel B). Robust standard errors in parenthesis. Control covariates are listed in Table 2. Sample sizes reported in Table A1.
*10%, **5%, and ***1% statistical significant levels.

**Table 4: Survey Design Average Treatment Effects, Household Chores and School Participation**

| | Participation in Household Chores | | Weekly Hours Spent on Household Chores | | School Enrollment | |
|---|---|---|---|---|---|---|
| | Short Raining Season (April - May 2016) | Harvest Season (Dec 2016 - Jan 2017) | Short Raining Season (April - May 2016) | Harvest Season (Dec 2016 - Jan 2017) | Short Raining Season (April - May 2016) | Harvest Season (Dec 2016 - Jan 2017) |
| All | -0.021 | 0.007 | 0.112 | -0.340 | 0.008 | 0.017 |
| | (0.014) | (0.014) | (0.431) | (0.397) | (0.015) | (0.013) |
| Boys | -0.015 | 0.009 | -0.079 | -0.180 | 0.029 | 0.000 |
| | (0.022) | (0.024) | (0.548) | (0.520) | (0.021) | (0.020) |
| Girls | -0.030 | 0.005 | 0.289 | -0.601 | -0.009 | 0.036** |
| | (0.018) | (0.016) | (0.665) | (0.589) | (0.021) | (0.017) |

**Notes:** Robust standard errors in parenthesis. Control covariates included are listed in Table 2. We collected information on household chores and schooling from both the proxy and the child in the second and third surveys. Sample sizes reported in Table A.1
*10%, **5%, ***1% statistical significant levels

**Table 5: Survey Design Treatment Effects by Gender of the Head of Household**

| | Pooled Survey Experiment | | Main Rainy Season (July - August 2015) | | Short Raining Season (April - May 2016) | | Harvest Season (Dec 2016 - Jan 2017) | |
|---|---|---|---|---|---|---|---|---|
| | Male | Female | Male | Female | Male | Female | Male | Female |
| **Panel A: Participation in Household Farm Activities In the Past Month** | | | | | | | | |
| All | 0.034*** | 0.115** | 0.055** | 0.165* | 0.016 | 0.117 | 0.024 | 0.085 |
| | (0.013) | (0.051) | (0.024) | (0.089) | (0.022) | (0.077) | (0.021) | (0.086) |
| Boys | -0.007 | 0.182** | -0.007 | 0.339*** | -0.006 | 0.222* | -0.014 | -0.013 |
| | (0.019) | (0.077) | (0.035) | (0.123) | (0.030) | (0.136) | (0.028) | (0.123) |
| Girls | 0.073*** | 0.083 | 0.105*** | 0.014 | 0.039 | 0.075 | 0.066** | 0.169 |
| | (0.020) | (0.072) | (0.034) | (0.133) | (0.033) | (0.096) | (0.031) | (0.132) |
| **Panel B: Monthly Hours Spent on Household Farm Activities (Tobit)** | | | | | | | | |
| All | 1.6230** | 7.911*** | 2.432** | 9.851** | -0.594 | 5.860 | 2.914** | 9.488 |
| | (0.778) | (3.347) | (1.197) | (5.137) | (1.198) | (3.870) | (1.471) | (7.170) |
| Boys | 0.393 | 14.642*** | 0.579 | 13.671* | -1.737 | 18.506*** | 2.516 | 17.953 |
| | (1.238) | (5.870) | (2.066) | (7.823) | (1.945) | (6.347) | (2.132) | (12.991) |
| Girls | 2.462*** | 3.917 | 3.131*** | 5.401 | 0.249 | 0.937 | 3.349* | 4.678 |
| | (0.953) | (3.647) | (1.277) | (5.561) | (1.434) | (2.939) | (1.993) | (8.298) |

**Notes:** Probability linear models on work status in the past 30 days from survey date (Panel A) and Tobit marginal effects on weekly hours worked in past 30 days from survey date (Panel B). Robust standard errors in parenthesis. Control covariates included are listed in Table 2. Sample sizes of children reported in Table A.1. *10%, **5%, ***1% statistical significant levels

**Table 6: Children Gender Composition and Survey Design Treatment Effects (Extensive Margin)**

| | Pooled Survey Experiments | Main Rainy Season (July - August 2015) | Short Raining Season (April - May 2016) | Harvest Season (Dec 2016 - Jan 2017) |
|---|---|---|---|---|
| **Panel A: Households with Mixed Gender Composition of Children** | | | | |
| All | 0.051*** | 0.081*** | 0.024 | 0.041* |
| | (0.016) | (0.030) | (0.027) | (0.025) |
| Boys | 0.009 | 0.064 | -0.006 | -0.034 |
| | (0.023) | (0.044) | (0.038) | (0.034) |
| Girls | 0.088*** | 0.092** | 0.048 | 0.109*** |
| | (0.023) | (0.041) | (0.040) | (0.037) |
| **Panel B: Households with Similar Gender Composition of Children** | | | | |
| All | 0.019 | 0.034 | 0.024 | 0.002 |
| | (0.021) | (0.037) | (0.034) | (0.033) |
| Boys | -0.004 | -0.041 | 0.022 | 0.001 |
| | (0.028) | (0.053) | (0.046) | (0.043) |
| Girls | 0.044 | 0.098 | 0.024 | 0.011 |
| | (0.032) | (0.065) | (0.050) | (0.051) |

**Notes:** Probability linear models on work status in the past 30 days. Robust standard errors in parenthesis. Control covariates included are listed in Table 2. Sample sizes included in table A1.
*10%, **5%, ***1% statistical significant levels.

**Table 7: Differential Treatment Effects by Male Head of Household Decision-Making Power (July/August 2015-Main Rainy Season)**

| | Farm work | Household chores |
|---|---|---|
| All | | |
| Treatment | 0.065*** | 0.066*** |
| | (0.026) | (0.026) |
| Decision power | -0.022 | 0.026 |
| | (0.042) | (0.054) |
| T*Decision power | -0.079 | -0.123 |
| | (0.079) | (0.081) |
| Boys | | |
| Treatment | -0.011 | 0.006 |
| | (0.038) | (0.037) |
| Decision power | -0.006 | 0.080 |
| | (0.061) | (0.077) |
| T*Decision power | 0.026 | -0.144 |
| | (0.101) | (0.117) |
| Girls | | |
| Treatment | 0.132*** | 0.116*** |
| | (0.037) | (0.036) |
| Decision power | -0.042 | -0.039 |
| | (0.058) | (0.075) |
| T*Decision power | -0.199** | -0.098 |
| | (0.094) | (0.113) |

**Notes:** Probability linear models on work status in the past 30 days from survey date. Robust standard errors in parenthesis. Control covariates included are listed in Table 2. Decision-making power is a dummy variable that takes the value 1 if the male head of household is the only decision maker or strongly influence the decision made, 0 otherwise. Sample sizes reported in Table A.1 . 10%, **5%, ***1% statistical significant levels

**Table 8: Survey Design Treatment Effects for Both Proxy Head of Household and Proxy Spouse (Dec 2016 / Jan 2017 - Harvest Season)**

| | **Proxy is** | | |
|---|---|---|---|
| | **Head of Household** | **Spouse** | **p-value** ($\hat{\beta}_{head} = \hat{\beta}_{spouse}$) |
| **Panel A: Participation in Household Farm Activities** | | | |
| All | 0.026 | 0.010 | 0.129 |
| | (0.020) | (0.021) | |
| Boys | -0.016 | -0.008 | 0.679 |
| | (0.027) | (0.028) | |
| Girls | 0.071*** | 0.024 | 0.001 |
| | (0.029) | (0.031) | |

**Notes:** Probability linear models on work status in the past 30 days from survey date. Robust standard errors in parenthesis. Control covariates for the head of household are listed in Table 2. Spouse covariates include age and years of education. Sample sizes reported in Table A.1
*10%, **5%, ***1% statistical significant levels

**Table 9: Survey Design Treatment Effects and Fair Trade**

| | Pooled Survey Experiment | | | Main Rainy Season (July - August 2015) | | | Short Raining Season (April - May 2016) | | | Harvest Season (Dec 2016 - Jan 2017) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All | Boys | Girls | All | Boys | Girls | All | Boys | Girls | All | Boys | Girls |
| **Panel A: Fairtrade Knowledge Standardized Index** | | | | | | | | | | | | |
| Treatment | 0.035*** | -0.002 | -0.009*** | 0.050*** | 0.002 | 0.092*** | 0.014 | -0.001 | 0.032 | 0.041* | -0.010 | 0.096*** |
| | (0.014) | (0.019) | (0.014) | (0.025) | (0.036) | (0.034) | (0.022) | (0.030) | (0.032) | (0.021) | (0.029) | (0.031) |
| FT Index | -0.009 | -0.018 | -0.009 | -0.033** | -0.056*** | -0.01406 | -0.012 | -0.013 | -0.009 | 0.019 | 0.014 | 0.023 |
| | (0.009) | (0.012) | (0.009) | (0.015) | (0.023) | (0.021) | (0.014) | (0.019) | (0.020) | (0.014) | (0.018) | (0.020) |
| T*FT | 0.016 | 0.021 | 0.016 | 0.042* | 0.058* | 0.033 | 0.041 | 0.016 | 0.060 | -0.042* | -0.018 | -0.070** |
| | (0.014) | (0.019) | (0.014) | (0.025) | (0.035) | (0.035) | (0.022) | (0.029) | (0.032) | (0.022) | (0.029) | (0.033) |
| **Panel B: Relational Contract Standardized Index** | | | | | | | | | | | | |
| Treatment | 0.029** | -0.003 | 0.062*** | 0.043* | 0.002 | 0.079*** | 0.018 | 0.002 | 0.033 | 0.026 | -0.016 | 0.070** |
| | (0.013) | (0.018) | (0.019) | (0.024) | (0.035) | (0.033) | (0.021) | (0.030) | (0.031) | (0.020) | (0.028) | (0.030) |
| FT Index | -0.002 | -0.003 | 0.000 | 0.001 | -0.003 | 0.004 | -0.031** | -0.020 | -0.041* | 0.028 | 0.019 | 0.039* |
| | (0.009) | (0.012) | (0.014) | (0.016) | (0.024) | (0.023) | (0.016) | (0.022) | (0.023) | (0.014) | (0.017) | (0.022) |
| T*FT | 0.027* | 0.022 | 0.031 | 0.061** | 0.048 | 0.077*** | 0.029 | 0.012 | 0.042 | -0.012 | 0.000 | -0.028 |
| | (0.015) | (0.021) | (0.021) | (0.026) | (0.039) | (0.036) | (0.027) | (0.038) | (0.038) | (0.022) | (0.028) | (0.032) |

**Notes:** Probability linear models on work status in the past 30 days from survey date. Robust standard errors in parenthesis. Control covariates included are listed in Table 2. Fairtrade knowledge index is measured using a principal component analysis model that includes seven standardized variables: head of household knowledge on FT price-setting, FT certification, FT standards on child labor, gender equality, environmental practices and democratic participation on Cooperative practices. Fairtrade relational index is measured principal component analysis model that includes six variables related to share of total coffee production sold to FT Coops, walking distances from house to FT coop collection and administrative centers, whether head of household is part of FT Coop Board, whether head of household vote in last FT Coop elections, whether head of household made participated in FT Coop meetings in past 12 months. Estimation includes the same set of control covariates as reported in main Table 2. Sample sizes reported in Table A.1 *10%, **5%, ***1% statistical significant levels

**Table A1: Number of Observations in Each Survey Design Experiment**

| | Main Rainy Season (July - August 2015) | Short Raining Season (April - May 2016) | Harvest Season (Dec 2016 - Jan 2017) |
|---|---|---|---|
| Number of Households | 1198 | 1197 | 1188 |
| Number of Households with Children 6-14 | 909 | 891 | 874 |
| Self-Reported | 307 | 305 | 298 |
| Proxy-Respondent | 602 | 586 | 576 |
| Head of Household Proxy | 583 | 568 | 561 |
| HH Male Proxy | 515 | 505 | 500 |
| Number of Children 6-14 | 1890 | 1829 | 1793 |
| Out of main rainy season | ------ | 20 | 26 |
| Out of short rainy season | 40 | ------ | 87 |
| Out of harvest season | 24 | 36 | ------ |
| # children 6- 14 with complete information | 1850 | 1813 | 1765 |
| # boys 6- 14 with complete information | 911 | 893 | 875 |
| # girls 6- 14 with complete information | 939 | 920 | 890 |
| # children aged 6-9 with complete information | 767 | 753 | 735 |
| # children aged 10-14 with complete information | 1083 | 1060 | 1030 |

**Table A.2: Survey Design Treatment Effects by Gender of the Proxy Respondent: Work Status in the Past 7 days**

| | Short Raining Season (April - May 2016) | Harvest Season (Dec 2016 - Jan 2017) |
|---|---|---|
| **Panel A: Participation in Household Farm Activities** | | |
| All | 0.005 | 0.027 |
| | (0.021) | (0.020) |
| Boys | -0.008 | -0.016 |
| | (0.030) | (0.027) |
| Girls | 0.017 | 0.0736** |
| | (0.031) | (0.030) |
| **Panel B: Weekly Hours Spent on Household Farm Activities (Tobit)** | | |
| All | -0.252 | 1.403*** |
| | (0.366) | (0.397) |
| Boys | -0.532 | 1.525*** |
| | (0.589) | (0.598) |
| Girls | -0.042 | 1.363*** |
| | (0.439) | (0.517) |

**Notes:** Probability linear models on work status in the past 7 days from survey date. Robust standard errors in parenthesis. Control covariates included are listed in Table 2. Sample sizes reported in Table A.1.
*10%, **5%, ***1% statistical significant levels