

## **Cash crops and food security:**

### **Evidence from Ethiopian smallholder coffee producers**

#### **Supplementary material**

<b>1. Questions in the food security module of the survey .....</b>	<b>1</b>
Table S1: Questions in the food security module of the survey .....	1
<b>2. Considering the endogeneity of the total income .....</b>	<b>2</b>
Table S2: Replicating Table 5 but dropping total household income from the model .....	2
Table S3: Replicating Table 5 but replacing coffee income share with (log) coffee income.	3
<b>3. Redefining inheritance in IV regressions.....</b>	<b>3</b>
Table S4: Replicating Column 2 in Table 5 but redefining the instruments .....	4
<b>4. Using count models .....</b>	<b>4</b>
Table S5: Replicating Table 5 using Poisson models .....	5
<b>5. Adding controls for relative remoteness .....</b>	<b>5</b>
Table S6: Replicating Table 5 but adding remoteness measures to the model.....	6
<b>References.....</b>	<b>6</b>

## 1. Questions in the food security module of the survey

### Table S1: Questions in the food security module of the survey

1a. In the past 12 months, did you worry that your household would not have enough food?

1b. How often did you worry in the past 12 months?

2a. In the past 12 months, were you or any household members not able to eat the kinds of foods you preferred because of lack of resources?

2b. How often does this happen in the past 12 months?

3a. In the past 12 months, did you or any household members have to eat a limited variety of foods because of lack of resources?

3b. How often does this happen in the past 12 months?

4a. In the past 12 months, did you or any household members eat food that you did not want to eat because of lack of resources to obtain other types of food?

4b. How often does this happen in the past 12 months?

5a. In the past 12 months, did you or any household members have to eat a smaller meal than you felt you needed because there wasn't enough food?

5b. How often does this happen in the past 12 months?

6a. In the past 12 months, did you or any household member have to eat fewer meals in a day because there was not enough food?

6b. How often does this happen in the past 12 months?

7a. In the past 12 months, was there ever no food to eat at all in your household because there were no resources to get any food?

7b. How often does this happen in the past 12 months?

8a. In the past 12 months, did you or any household member go to sleep at night hungry because there was not enough food?

8b. How often does this happen in the past 12 months?

9a. In the past 12 months, did you or any household member go a whole day without eating anything because there was not enough food?

9b. How often does this happen in the past 12 months?

*The response options for a-questions were 'yes or no'.*

*The response options for b-questions were 'Rarely (1-2 times)', 'Sometimes (3-10 times)' or 'Often (more than 10 times)'.*

## 2. Considering the endogeneity of the total income

The total household income variable can also be endogenous for the same reasons as our coffee income share variable. This would then bias the estimate that measures the impact of the total income on food security, but it may also affect other estimated coefficients, especially if they are correlated with the total income variable. We explored the sensitivity of our estimates in two ways. In the first step, we excluded the total income variable from the model. In the second, we used (logged) coffee income instead of the share variable.

We see that by excluding the total income variable from the model yields nearly an identical estimate on the coffee-income share variable (Table S2 below). Similarly, using a coffee income variable instead of the share also gives a negative and significant coefficient in both the OLS and IV-models (Table S3 below).

**Table S2: Replicating Table 5 but dropping total household income from the model**

Dependent variable: Food insecurity Z-score	(1) OLS	(2) IV
Share of coffee income	-0.396*** (0.121)	-1.491** (0.693)
Controls? <sup>a)</sup>	yes	yes
Woreda dummies? <sup>a)</sup>	yes	yes
R <sup>2</sup>	0.327	-
Weak-identification tests:		
Cragg-Donald F-statistic	-	14.06
Kleibergen-Paap rk Wald F statistic (or Angrist-Pischke F-test of excluded instruments)	-	20.12
p-value of Angrist-Pischke F-test	-	0.000***
Over-identification test:		
Hansen-J	-	0.196
--- p-value	-	0.658
Observations	1,597	1,597

Source: Authors' calculations based on ESSP's coffee survey, 2014. Note: Standard errors clustered at the *kebele* level in parentheses. Statistical significance noted at \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. <sup>a)</sup> Coefficients omitted to preserve space.

**Table S3: Replicating Table 5 but replacing coffee income share with (log) coffee income**

Dependent variable: Food insecurity Z-score	(1)	(2)
	OLS	IV
(log) Coffee income	-0.100*** (0.025)	-0.427** (0.212)
(log) Total household income	-0.010 (0.046)	0.309 (0.209)
Controls? <sup>a)</sup>	yes	yes
Woreda dummies? <sup>a)</sup>	yes	yes
R <sup>2</sup>	0.333	-
Weak-identification tests:		
Cragg-Donald F-statistic	-	8.41
Kleibergen-Paap rk Wald F statistic (or Angrist-Pischke F-test of excluded instruments)	-	8.35
p-value of Angrist-Pischke F-test	-	0.001***
Over-identification test:		
Hansen-J	-	0.223
--- p-value	-	0.637
Observations	1,597	1,597

Source: Authors' calculations based on ESSP's coffee survey, 2014.

Note: Standard errors clustered at the *kebele* level in parentheses. Statistical significance noted at \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. <sup>a)</sup> Coefficients omitted to preserve space.

### 3. Redefining inheritance in IV regressions

In Section 5, we considered cases in which the household inherits land after the parent(s) pass away and cases in which parents allocate land to newlyweds so that they can start cultivating their own land.<sup>1</sup> The latter scenario poses a concern that the size of the inherited land is correlated with parents' capacity to support the newly formed households in times of need (e.g. during the lean season). This would invalidate our exclusion restriction. To check this, we re-estimated column 2 in Table 5 but refined our instruments so that they only considering land inheritance that took place

<sup>1</sup> Out of all parcels owned by the households, 34 % of was allocated by parents (pre-inheritance), 22 % was allocated by the government, 34 % was inherited after parents' decease and 10 % was purchased.

after the parent(s) passed away. Table S4 below shows the results. It is comforting that the estimated coefficient is similar in magnitude, though somewhat less precisely estimated.

**Table S4: Replicating Column 2 in Table 5 but redefining the instruments**

Dependent variable: Food insecurity Z-score	(1) (Column 2 in Table 5)	(2) Redefined IVs
Share of coffee income	-1.402** (0.678)	-1.559* (0.910)
Controls? <sup>a)</sup>	yes	yes
Woreda dummies? <sup>a)</sup>	yes	yes
Weak-identification tests:		
Cragg-Donald F-statistic	14.34	5.318
Kleibergen-Paap rk Wald F statistic (or Angrist-Pischke F-test of excluded instruments)	20.22	5.019
p-value of Angrist-Pischke F-test	0.000***	0.009***
Over-identification test:		
Hansen-J	0.204	0.058
--- p-value	0.652	0.810
Observations	1,597	1,597

Source: Authors' calculations based on ESSP's coffee survey, 2014.

Note: Standard errors clustered at *kebele* level in parentheses. Statistical significance at \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . <sup>a)</sup> Coefficients omitted to preserve space.

#### 4. Using count models

Our outcome variable is essentially a count (see Figure 1), and therefore the validity of using a linear model is questionable. The advantage of using the linear model is that it provides a host of specification tests that we can use to assess the validity of our IV-approach. Still, we assess the robustness of our findings using the Poisson model. For this exercise, we replace the z-score version of the dependent variable with the original count. This variable has a mean of 6.28 and standard deviation of 7.03 (see Table 3). Table S5 below shows that the estimated coefficients (marginal effects) are comparable to the ones obtained using the linear models in Table 5.<sup>2</sup> The

<sup>2</sup> The Pearson goodness-of-fit test presented at the bottom of Column 1 suggests that the conditional Poisson distribution does not provide a particularly good fit to our data. We replicated column 1 in Appendix E using a Negative Binomial Model which is less sensitive to over-dispersion (Winkelmann 2008, p. 134). While the estimated coefficient on the share of coffee income was somewhat larger, it was still within the 95-confidence interval (-14.77; -0.53) of the coefficient estimated with the Poisson model. These results are not reported in the paper but available upon request.

simple Poisson yields a coefficient of -1.96 which corresponds to -0.28 unit of standard deviations in the food insecurity score. This is slightly lower than the corresponding OLS estimate in column 1 of Table 4. The IV-Poisson estimates range between -7.63 and -13.04, corresponding to -1.09 and -1.86 units of standard deviations in the food insecurity score. The linear-IV estimate reported in column 2 of Table 5, -1.40, lies in the middle of this range.

**Table S5: Replicating Table 5 using Poisson models**

Dependent variable: Food insecurity (count)	(1)	(2)	(3)	(4)
estimator:	Poisson	Poisson IV-GMM	Poisson IV-GMM	Poisson IV control function
error structure:	-	additive	multiplicative	-
Share of coffee income in total household income	-1.955*** (0.681)	-7.642** (3.625)	-12.466* (6.905)	-13.042*** (5.051)
Controls? <sup>a)</sup>	yes	yes	yes	yes
Woreda dummies? <sup>a)</sup>	yes	yes	yes	yes
Pearson goodness-of-fit test	9623.0	-	-	-
--- p-value	0.000***	-	-	-
Over-identification test:	-			
Hansen-J $\chi^2$ -test	-	0.020	0.645	-
--- p-value	-	0.887	0.421	-
Observations	1,597	1,597	1,597	1,597

Source: Authors' calculations based on ESSP's coffee survey, 2014.

Note: Standard errors clustered at the *kebele* level in parentheses. The standard errors for the marginal effects at the means are computed using the delta-method. Statistical significance noted at \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. <sup>a)</sup> Coefficients omitted to preserve space.

## 5. Adding controls for relative remoteness

Relative remoteness could be correlated with reported food insecurity (Stifel and Minten 2017), but it may also be a factor in a household's decision to engage in cash crop cultivation (Fafchamps 1992). To check that this is not driving our results, we appended our model with variables that measure the distance of communities in which sample households reside to the nearest all-weather road and to the district capital. The results, reported in Table S6 below, are robust to including these variables in the model.

**Table S6: Replicating Table 5 but adding remoteness measures to the model**

Dependent variable: Food insecurity Z-score	(1)	(2)
	OLS	IV
(log) coffee income	-0.419*** (0.118)	-1.464** (0.684)
(log) distance to all season road	-0.044 (0.035)	-0.053 (0.037)
(log) distance to the woreda admin office	0.078 (0.054)	0.126* (0.070)
Controls? <sup>a)</sup>	yes	yes
Woreda dummies? <sup>a)</sup>	yes	yes
R <sup>2</sup>	0.333	-
Weak-identification tests:		
Cragg-Donald F-statistic	-	14.10
Kleibergen-Paap rk Wald F statistic (or Angrist-Pischke F-test of excluded instruments)	-	20.64
p-value of Angrist-Pischke F-test	-	0.000***
Over-identification test:		
Hansen-J	-	0.250
--- p-value	-	0.617
Observations	1,597	1,597

Source: Authors' calculations based on ESSP's coffee survey, 2014.

Note: Standard errors clustered at the *kebele* level in parentheses. Statistical significance noted at \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. <sup>a)</sup> Coefficients omitted to preserve space.

## References

- Fafchamps, Marcel. 1992. "Cash crop production, food price volatility, and rural market integration in the third world." *American Journal of Agricultural Economics* 74 (1):90-99.
- Stifel, David, and Bart Minten. 2017. "Market Access, Welfare, and Nutrition: Evidence from Ethiopia." *World Development* 90:229-241.
- Winkelmann, Rainer. 2008. *Econometric analysis of count data*. Verlag Berlin Heidelberg: Springer.