Storing a staple crop for own consumption: Linkages to food security

L.Y. Khor; M. Zeller

University of Hohenheim, , Germany

Corresponding author email: lykhor@uni-hohenheim.de

Abstract:

We examine in our study the effect on food security from storing more of a harvested staple crop for own consumption. The analysis is based on survey data from more than 1,800 households in rural Guatemala that harvested beans in 2015, with crop production data from two planting seasons. By including a full factorial of three variables of wealth and the harvest usage decision from two different seasons, we examine the effect of storing more beans across different wealth levels and harvest storage amounts. We find an interesting contrast between the impact on the household's sense of food security and its actual food security. Even though storing more beans does not reduce the likelihood of a household running out of food, it does help to reduce the worry among poorer households that such an incident might occur. We also analyze the effect of storage decisions on different categories of food security using generalized ordered logit, and then compare whether the effect on food security is different between adults and children using multinomial logit. We find that increasing storage could lead to adults having less variation in their food consumption patterns, but the effect is less pronounced among children.

Acknowledegment: This work was supported by the German Federal Ministry for Economic Cooperation and Development (BMZ).

JEL Codes: Q18, O13

#1449

Storing a staple crop for own consumption: Linkages to food security

Abstract

We examine in our study the effect on food security from storing more of a harvested staple crop for own consumption. The analysis is based on survey data from more than 1,800 households in rural Guatemala that harvested beans in 2015, with crop production data from two planting seasons. By including a full factorial of three variables of wealth and the harvest usage decision from two different seasons, we examine the effect of storing more beans across different wealth levels and harvest storage amounts. We find an interesting contrast between the impact on the household's sense of food security and its actual food security. Even though storing more beans does not reduce the likelihood of a household running out of food, it does help to reduce the worry among poorer households that such an incident might occur. We also analyze the effect of storage decisions on different categories of food security using generalized ordered logit, and then compare whether the effect on food security is different between adults and children using multinomial logit. We find that increasing storage could lead to adults having less variation in their food consumption patterns, but the effect is less pronounced among children.

Keywords: Staple crop, food security, household consumption, bean, Guatemala

1 Introduction

Earlier research on the storage of staple crops for own consumption focuses on how price fluctuations affect households' storage decisions. Empirical results have been inconclusive on the direction of the effect. Toquero et al. (1975) find that the amount of rice harvested has a positive and significant impact on how much households decide to store for own consumption, but the price of rice does not have a significant impact on that decision. They suggest that the reason is that rice is a staple crop that fulfills basic subsistence needs. However, Medani (1975) shows that higher prices lead to a smaller amount being stored for own consumption and Renkow (1990) finds that farmers keep more for consumption when the price of a staple crop rises. According to a study by Chinn (1976), both higher production and price lead to a higher storage of staple crop for a household.

The impact of agricultural commercialization is another common research topic related to staple crops, specifically farmers' decisions to switch from subsistence crops to cash crops and how that affects their income and food security (Kennedy and Cogill, 1988; von Braun, 1995; Pereira et al., 2014; Muriithi and Matz, 2015; Van den Broeck and Maertens, 2016). A recent study by Carletto et al. (2017) find little evidence that agricultural commercialization, measured as the percentage of crop sold from the amount produced, leads to a higher household nutritional status. Rather than the choice between cash and subsistence crops, our research focuses on the allocation decision for a staple crop after harvest. How much does a household decide to keep for its own consumption and does this decision affect its food security? Previous research has shown that seasonality matters when analyzing food security (Abdullah and Wheeler, 1985; Branca et al., 1993; Garrett and Ruel, 1999; Hillbruner and Egan, 2008); food security should be compared between the harvest and lean seasons. In our study, we focus on the duration of the effect by examining the impact of the two recent harvest storage decisions.

The dataset for our analysis is collated from a household survey containing information on the household production and food security of more than 1,800 farmers in Guatemala who harvested beans in 2015. The production data include decisions about the amount of harvested beans that were stored from the past two planting seasons. The two harvests were six months and one year before the survey was conducted, respectively. We examine how the harvest usage decisions from these two most recent planting seasons affect the food security of households.

2 Study Region and Dataset

The biofortification of staple crops has gained popularity as a potentially cost-effective and sustainable method for reducing micronutrient deficiency in the rural areas of developing countries (Bouis and Welch, 2010; Meenakshi et al., 2010; Hotz et al., 2012a,b; Saltzman et al., 2013). As part of the effort, HarvestPlus introduced high-iron beans in Guatemala in August of 2016. Anemia is a public health problem in Guatemala. Based on the National Health Survey of 2009, 47.7% of children under the age of six and 21.4% of women of childbearing age suffer from anemia (Mujica-Coopman et al., 2015). A baseline survey was conducted in June of 2016, before the introduction of the beans, to allow for impact analysis in the future. This study is based on the data collected during that survey. As the main purpose of the project was to reduce the anemia rate among the population, the release of the beans was targeted to areas of the country with high levels of iron deficiency and ubiquitous consumption of the staple crop, beans. Seven municipalities from Chiquimula, Jalapa, and Jutiapa were purposively selected for this study. A community listing exercise was conducted in 2015 to identify the households that produce beans and that have girls between 10 and 15 years of age. All these households form the sampling frame of the study (Birol et al., 2015).

Communities were randomly selected into treatment and control groups, however, all participating households were included in the baseline survey, which is the basis of this article. A socio-economic survey was conducted in June of 2016 on all households; the questionnaire included sections on household characteristics and household crop production, with detailed questions about bean production, bean consumption patterns, housing conditions, assets, risk preferences, health knowledge, and a series of food security self-assessment questions. Immediately after the second survey round, a third round, the nutrition survey, was conducted on the same households to collect data on health conditions and nutrition, including food consumption data from the preceding seven days.

In the baseline study, 99% of farmers who harvested beans in 2015 kept some harvest for own consumption. The average harvest was 36 kg per year per capita, with 18 kg being stored. The storage amount was slightly lower than the average bean consumption of 23 kg per year per capita. Many of the farmers (about 63% of the total) supplemented their stock by purchasing beans from the market. For the outcome variable, households were asked during the baseline survey about their food security situation over the preceding three months. In terms of the timeline of the variables, the food security questions covered the period between seven and nine months after the first harvest storage decision was made and between one and three months after the second harvest storage decision was made (cf. Table 1). A list of 16 food security self-assessment questions were asked to the household

Table 1: Timeline of the main variables collected from the baseline survey in June of 2016

2015											2016				
Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
Harvest (first planting season) →						Harvest (second planting season) →							← Food security →		

head's spouse. Two questions refer to the perception and actual food availability at the household level. There are seven questions about food insecurity of adults and children in the household. Each of these seven questions is asked twice: first referring to adults who are above the age of 18, then referring to children below the age of 18. About 90% of respondents

Table 2: List of questions to capture various levels of food insecurity

Food insecurity	Question levels	Questions
Mild	Household as a whole	Worried that household will run out of food
↓		There is no food in the household
	Each question is asked twice, once for adults ($\geq$ 18 year-old) and once for children ($<$ 18 year-old)	Always eat the same type of food
		Do not have a healthy and diverse diet
		Reduce the amount of food served
		Eat less than one should
		Feel hungry but do not eat
		Skip one meal
Severe		Stop eating the whole day

said they were worried the food in their household would run out. Due to a lack of resources, 46% of households had to reduce the amount of food served to adults and 40% had to reduce the amount served to children. At the severe end of the scale, 15% of households had at least one adult who stopped eating the whole day and 12% answered yes to the same question referring to children.

3 Methodology

As the dependent variables for food insecurity are dummy variables, various forms of logit regression are used to determine the effect of harvest storage decisions on food insecurity. In the case of the household-level food insecurity questions, there is only one dummy variable and a simple logit regression is used. For the seven other food insecurity questions that are asked separately for adults and children, there are two dummy variables for each question. This produces four possible scenarios for each question: the whole household is food secure, the adults are not food secure, the children are not food secure, the whole household is not food secure. In this case, we can generate an ordinal outcome variable with three categories ranging from whole household is food secure, to either adults or children are not food secure, to both adults and children are not food secure. A common model used to analyze an ordinal outcome variable is the ordered logit model. However, it has a constraint that is often violated (Long and Freese, 2014), as it assumes that the effect of the independent variable is constant as we move across the different categories. Williams (2016) suggests relaxing the proportional odds assumption by using a generalized ordered logit model. Following this model, the outcome for a dependent variable with N categories is

$$P(Y_i > j) = \frac{\exp(\alpha_j + X_i\beta_j)}{1 + \left[\exp(\alpha_j + X_i\beta_j)\right]}, \qquad j = 1, 2, ..., N-1. \tag{1}$$

If we assume that $\boldsymbol{V_i}$ is a list of independent variables where the effect on Y_i is different across categories, and $\boldsymbol{Z_i}$ is a list of independent variables with a constant effect, Equation (1) for our analysis with three food security categories becomes

$$P(Y_i > j) = \frac{\exp(\alpha_j + \boldsymbol{V_i}\boldsymbol{\beta_j} + \boldsymbol{Z_i}\boldsymbol{\gamma})}{1 + \left[\exp(\alpha_j + \boldsymbol{V_i}\boldsymbol{\beta_j} + \boldsymbol{Z_i}\boldsymbol{\gamma})\right]}, \qquad j = 1, 2. \tag{2}$$

The empirical analysis is carried out using the Stata command *gologit2* from Williams (2006). It involves an iterative process that tests whether the effect is the same when moving from the first category to the second, and when moving from the second category to the third. If the test shows no statistically significant difference for a coefficient, then the proportional odds assumption is imposed on that variable and there is only one coefficient for that variable. If the test rejects the assumption, then there will be more than one coefficient for that variable, as it means that the marginal effect that causes the change in the category of the outcome variable depends on which category the change refers to. In this case, the effect that leads to the change from the first category to the second category is different from the effect that corresponds to the change from the second category to the third category.

In our study, the main independent variables are the decisions made on the use of bean harvest from the two most recent planting seasons. As shown in Table 1, the baseline survey was conducted in June of 2016 with the harvest of the two previous planting seasons being in July of 2015 and January of 2016. As the food insecurity questions cover the three-month period between March and May of 2016, the main analysis is about the effect on food insecurity from the harvest use of two different time periods. First is the effect on food insecurity from the harvest storage decision made between seven and nine months earlier (longer-term) and the second is the effect from the decision made between one and three months (shorter-term) before the reference period of the food insecurity questions. Household wealth is likely to be one of the main determinants of food security. In addition, the effect on food insecurity from the harvest storage of one planting season could be affected by the amount of stored harvest from another season. Therefore, we include in the regression the full factorial of the three variables with W_i being household wealth, S_i being harvest storage from the first planting season, and T_i being harvest storage from the second planting season: W_i, S_i, T_i, $(W_i \times S_i)$, $(W_i \times T_i)$, $(S_i \times T_i)$, and $(W_i \times S_i \times T_i)$.

Household wealth is captured by an index constructed from principal component analysis (PCA). An index is used to reflect wealth because the inclusion of multiple variables makes it a more stable indicator over time and reflects the different aspects of wealth, such as assets and housing conditions. A list of potential wealth variables from the baseline survey is filtered according to the methodology of Zeller et al. (2006). PCA is run on these variables and only the variables with loading factors above a pre-determined value of 0.3 are kept (Henry et al., 2003). Table 3 shows the final list of variables used to construct the wealth index and their corresponding loading factors. We also check for the validity of the index by examining its eigenvalue and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. It has an eigenvalue of 3.62 and a KMO score of 0.83, fulfilling the criteria recommended by Henry et al. (2003).

Table 3: List of variables in the wealth index

Description	Mean	Loading
Household has a gas or electric stove (1 = yes, 0 = no)	0.17	0.6056
Household has a refrigerator (1 = yes, 0 = no)	0.29	0.6798
Household has a color television (1 = yes, 0 = no)	0.63	0.5355
Household has a truck (1 = yes, 0 = no)	0.07	0.4536
Household has a motorcycle (1 = yes, 0 = no)	0.07	0.4117
Household has a savings account at a bank (1 = yes, 0 = no)	0.16	0.4133
Floor of the house is earth (1 = yes, 0 = no)	0.39	−0.6414
House is connected to the water distribution network (1 = yes, 0 = no)	0.67	0.4898
Water tap is inside the house (1 = yes, 0 = no)	0.36	0.4470
Toilet is inside the house (1 = yes, 0 = no)	0.37	0.6743
No indoor or outdoor toilet (1 = yes, 0 = no)	0.35	−0.6464
Number of rooms in the house, excluding kitchen and toilet	2.19	0.4834

In comparison to linear models, interpreting the effect of interaction terms is more complicated in nonlinear models, such as logit (Norton et al., 2004). There are three common methods used for this purpose. First is to examine the marginal effect of one variable while holding the other interaction terms at their respective mean values. Second is to calculate the change in probability of the outcome for each individual observation by modifying only the value of one variable and then taking the average of these changes in probability for all observations. As explained by Williams (2012), both methods are quite restrictive, as only one estimate is produced, and it could potentially mask the interesting insight that could be obtained by observing the change in marginal effects across the different values of other variables. Williams suggests the use of a third method, which is to look at the marginal effects at representative values for the variables of greatest interest. We follow this recommendation in the presentation of our results in Section 4. As there are three variables included in the interaction, an ideal way is to look at how the marginal effect for one of the three variables changes, as we move across all the possible values of the other two variables. However, this would require a three-dimensional presentation of the results and might not be that clear in a paper format. Therefore, for interpreting the effect on food insecurity from the three variables in the interaction terms, we examine the change in the marginal effect of one variable across all the possible values of the second variable while holding the third variable constant at its 10th percentile, 50th percentile, and 90th percentile, respectively. We then repeat the procedure but reverse the position of the second and third variables. This allows us to present the results clearly in two-dimensional graphs and cover the range of the variables at the lower end, the middle, and the higher end.

In addition to the abovementioned interaction variables that are the focus of our analysis, we include in the regression a list of controls based on commonly used variables from previous studies and put them within the framework of Scoones (1998), which covers four main categories of capital in rural livelihoods: financial, natural, human, and social. Table 4 shows the summary statistics of these variables. We use a wealth index and ownership of

Table 4: Descriptions of variables

Variable	Mean	Standard deviation
Harvested bean in first season of 2015 for own consumption (kg per capita)	3.72	8.46
Harvested bean in second season of 2015 for own consumption (kg per capita)	14.54	24.03
Household has livestock (1=yes, 0=no)	0.85	0.36
Total area of farm land (hectares)[a]	1.86	2.72
Average distance from house to farm land (minutes)	15.82	13.60
Total bean harvested in first season of 2015 (kg)	40.02	124.84
Total bean harvested in second season of 2015 (kg)	198.14	390.80
Age of household head (years)	46.73	11.85
Education of household head (years)	2.75	2.79
Gender of household head (1=male, 0=female)	0.74	0.44
Risk preference of household head[b]	2.18	0.65
Household size	6.60	2.23
Frequency of sharing agricultural information with friends or neighbors[c]	2.01	0.91
Frequency of sharing agricultural inputs or seeds with friends or neighbors[c]	1.59	0.93

Note: Summary statistics are calculated from the 1,829 observations with data in all regression variables.

[a] (10000/628.8) tarea = 1 hectare

[b] Based on a question about willingness to try out new agricultural technology before observing the results. There are four categories: 0 = strongly disagree, 1 = disagree, 2 = agree, 3 = strongly agree.

[c] Four categories: 0 = never, 1 = rarely, 2 = occasionally, 3 = frequently

livestock to capture the financial capital of a household. The latter variable helps to improve food security by providing an additional food source to the household. As we do not have the local weather data for the natural capital, we include dummy variables that represent the municipalities in the dataset to control for the municipality-level fixed effects, with the assumption that the weather pattern does not vary much within the same municipality. In addition, the condition of farm land (Garrett and Ruel, 1999; Feleke et al., 2005; Mallick and Rafi, 2010) is represented by the total farm area to which the household has access, the farm location as measured by the average distance from house to farm land, and the bean harvest in the two planting seasons, which helps to control for the total amount of harvested beans that is available to the households before a decision is made on whether to sell them or to keep them for own consumption. Human capital is captured by the characteristics of the household head (Kennedy and Cogill, 1988; Babatunde and Qaim, 2010; Mallick and Rafi, 2010) and household size (Kennedy and Cogill, 1988; Garrett and Ruel, 1999; Feleke et al., 2005; Babatunde and Qaim, 2010). The characteristics include age, education, gender, and risk preference. Some of these variables, such as risk preference and household size, could also reflect how likely the household is to interact with their community. Grootaert and Narayan (2004), Wetterberg (2007), and Wossen et al. (2016) show the importance of social capital in improving household welfare. In a study on food insecurity among women of households receiving charitable food assistance, Tarasuk (2001) finds that hunger and food insecurity are much higher among those who think of themselves as socially isolated. In our regression analysis, the additional variables that we use to control for social capital include whether the household frequently shares agricultural information and inputs, such as seeds, with friends or neighbors.

For testing the goodness-of-fit (GOF) of nonlinear regression models, a common choice is the Hosmer-Lemeshow test (Hosmer and Lemeshow, 1980). However, as this test statistic is very sensitive to the number of groups chosen and the choice of this number is arbitrary, Allison (2014) suggests the use of a standardized Pearson test, which is one of the tests that do not require the grouping of data. The test statistic is calculated based on the asymptotic distribution of Windmeijer (1990) using the code from Weesie (2002). The null hypothesis for this test is that the model is correctly specified. A high test statistic or a low p-value rejects the model. We include the GOF test output in the results table.

4 Results and Discussion

The results are separated according to the type of food insecurity questions shown in Table 2. There are two questions that cover the food insecurity situation of the household as a whole. Section 4.1 presents the results for these two variables on household-level food insecurity. The other seven questions are asked twice, once referring to the group of adults in the household and once referring to the group of children. Section 4.2 includes the outcome captured by these seven food insecurity questions. As the analysis includes a full factorial of three independent variables, there are many potential outcome scenarios. Table 5 shows the scenarios that are analyzed for the three variables in the interaction terms under each of the food insecurity questions in Table 2. It is also possible to examine the other scenarios

Table 5: Scenarios analyzed for every food insecurity question

Scenario	Change in marginal effect of ...	Across full range of ...	Holding ...	at ...
1	S_i	W_i	T_i	10^{th} percentile
2	S_i	W_i	T_i	50^{th} percentile
3	S_i	W_i	T_i	90^{th} percentile
4	T_i	W_i	S_i	10^{th} percentile
5	T_i	W_i	S_i	50^{th} percentile
6	T_i	W_i	S_i	90^{th} percentile
7	T_i	S_i	W_i	10^{th} percentile
8	T_i	S_i	W_i	50^{th} percentile
9	T_i	S_i	W_i	90^{th} percentile

S_i Bean harvest storage from the first planting season
T_i Bean harvest storage from the second planting season
W_i Household wealth index

from the interaction terms, such as changes in the marginal effect of wealth across different storage decisions. We do not carry out these analyses because wealth is a very significant determinant in the regression before adding the interaction terms. Therefore, any new insight gained will probably be minimal, as wealth is likely to be significant across all the possible values of storage decisions.

In Section 4.1, where food insecurity at the household level is captured by a dummy variable, we analyze the marginal effect of the harvest storage decision on food insecurity by examining all the potential scenarios shown in Table 5. With two household-level food insecurity questions in this section, there are 18 possible outcome scenarios. The number is even greater in Section 4.2, as each of the seven questions in Table 2 produces a dependent variable of household food insecurity that is represented by three or four categories. The analyses from the generalized ordered logit and multinomial logit regressions produce 189 and 252 potential outcome scenarios, respectively. Due to space limitations, we do not present the output of all the scenarios. We start by providing detailed results from the first food insecurity question in Table 2. For the rest of the questions, we summarize the output in words, supplemented by graphs from a few of the more interesting scenarios.

4.1 Food insecurity at the household level

In a linear regression model, heteroskedasticity leads to an inefficient estimator although it is still unbiased and consistent. However, heteroskedasticity in a nonlinear model causes the maximum likelihood estimator to be inconsistent (Greene, 2003, p. 679). Before running the analysis on the first food insecurity question, we check for potential model misspecification by testing for heteroskedasticity. We compare the coefficients from the homoskedastic probit model with those from the heteroskedastic probit model using the Stata *hetprobit* command. Results for four different specifications are shown in Table 6. The first column is the basic model without any interaction terms. The second column is the basic model with the bean harvest from both seasons in logarithmic form. The third column is the basic model with the full factorial of the three independent variables: household wealth, harvest storage from

Table 6: Test for heteroskedasticity in four model specifications

Variables	(1)	(2)	(3)	(4)
First season bean storage (S_i)	−0.0002	−0.0010	0.0029	0.0013
	(0.0054)	(0.0073)	(0.0065)	(0.0086)
Second season bean storage (T_i)	−0.0029*	−0.0018	−0.0042*	−0.0021
	(0.0017)	(0.0017)	(0.0022)	(0.0025)
Wealth index (W_i)	−0.1745***	−0.1788***	−0.2352***	−0.2184***
	(0.0450)	(0.0449)	(0.0547)	(0.0559)
$S_i \times T_i$			−0.0004*	−0.0004*
			(0.0002)	(0.0003)
$S_i \times W_i$			0.0062	0.0050
			(0.0057)	(0.0056)
$T_i \times W_i$			0.0013	0.0002
			(0.0013)	(0.0015)
$S_i \times T_i \times W_i$			0.0004*	0.0004*
			(0.0002)	(0.0002)
Livestock ownership	−0.0377	−0.0118	−0.0399	−0.0194
	(0.1141)	(0.1153)	(0.1146)	(0.1157)
Total farm area	−0.0027	−0.0037	−0.0043	−0.0039
	(0.0144)	(0.0137)	(0.0142)	(0.0139)
Distance to farm land	0.0058*	0.0062*	0.0059*	0.0060*
	(0.0032)	(0.0032)	(0.0032)	(0.0032)
Bean harvest in first season	−0.0001		−0.0002	
	(0.0003)		(0.0003)	
Bean harvest in second season	−0.0003**		−0.0002*	
	(0.0001)		(0.0001)	
Dummy for harvest in first season		−0.1985		−0.2469
		(0.2418)		(0.2473)
ln(Bean harvest in first season)		−0.0308		−0.0330
		(0.0639)		(0.0654)
Dummy for harvest in second season		−0.4657**		−0.4914**
		(0.2053)		(0.2125)
ln(Bean harvest in second season)		−0.1195***		−0.1085***
		(0.0378)		(0.0396)
Age of household head	0.0005	0.0006	0.0014	0.0014
	(0.0036)	(0.0036)	(0.0037)	(0.0037)
Education of household head	0.0189	0.0210	0.0218	0.0228
	(0.0159)	(0.0159)	(0.0161)	(0.0161)
Gender of household head	−0.2872***	−0.2801***	−0.3025***	−0.2950***
	(0.1004)	(0.1010)	(0.1011)	(0.1016)
Risk preference of household head	−0.0696	−0.0626	−0.0575	−0.0526
	(0.0592)	(0.0593)	(0.0595)	(0.0596)
Household size	0.0139	0.0210	0.0127	0.0191
	(0.0186)	(0.0191)	(0.0187)	(0.0193)
Sharing agricultural information	0.0790	0.0805*	0.0845*	0.0827*
	(0.0484)	(0.0487)	(0.0486)	(0.0489)
Sharing agricultural inputs and seeds	0.0352	0.0317	0.0353	0.0307
	(0.0473)	(0.0477)	(0.0475)	(0.0478)
Observations	1,829	1,829	1,829	1,829
LR test of homoskedasticity (p-value)	0.0000	0.0000	0.0005	0.1377

Note: Asterisk (*), double asterisk (**), and triple asterisk (***) denote significance at 10%, 5%, and 1%, respectively. All the regressions above include the dummy variables for municipality.

the first planting season, and harvest storage from the second planting season. The fourth column is the basic model with the harvest in logarithmic form and full factorial of the three variables mentioned above. As the heteroskedastic probit model does not converge in the case of the first and second columns, the iteration limit is set at 100 for those two specific regressions. Test results show that the difference in the coefficients between the homoskedastic and heteroskedastic probit models is statistically significant at the 5% level for the models in the first three columns. The difference is insignificant for the model in the fourth column. This supports the inclusion of the full factorial of the wealth and harvest storage variables.

As there are farmers with a harvest in only one season, taking the logarithmic form of harvest presents some issues in the season when the harvest amount is zero. In this case, we follow the recommendation of Battese (1997) by replacing the values of zero with one, so that the values become zero in the logarithmic form. A dummy variable is then added into the regression to account for this switch. This new dummy variable takes the value of one for all the cases in which the value of the harvest has been replaced. In addition to the heteroskedasticity test, we compare also the results from both probit and logit regressions. Although there are differences in the magnitude of the coefficients, the sign and the statistical significance of the coefficients are consistent across both. Therefore, we report only the logit version of the results due to the possibility of interpreting the coefficients in the form of odd ratios.

Table 7 shows the regression output from the first two food insecurity questions of Table 2, including the goodness-of-fit statistic from a standardized Pearson test (Windmeijer, 1990; Weesie, 2002). Note that the questions refer to food insecurity in the household. Therefore, a positive coefficient means that the variable increases food insecurity (or lowers food security). The null hypothesis for the Pearson-Windmeijer goodness-of-fit test is that the model fits well. Test results show that there is no evidence of poor fit in either regression; the p-value

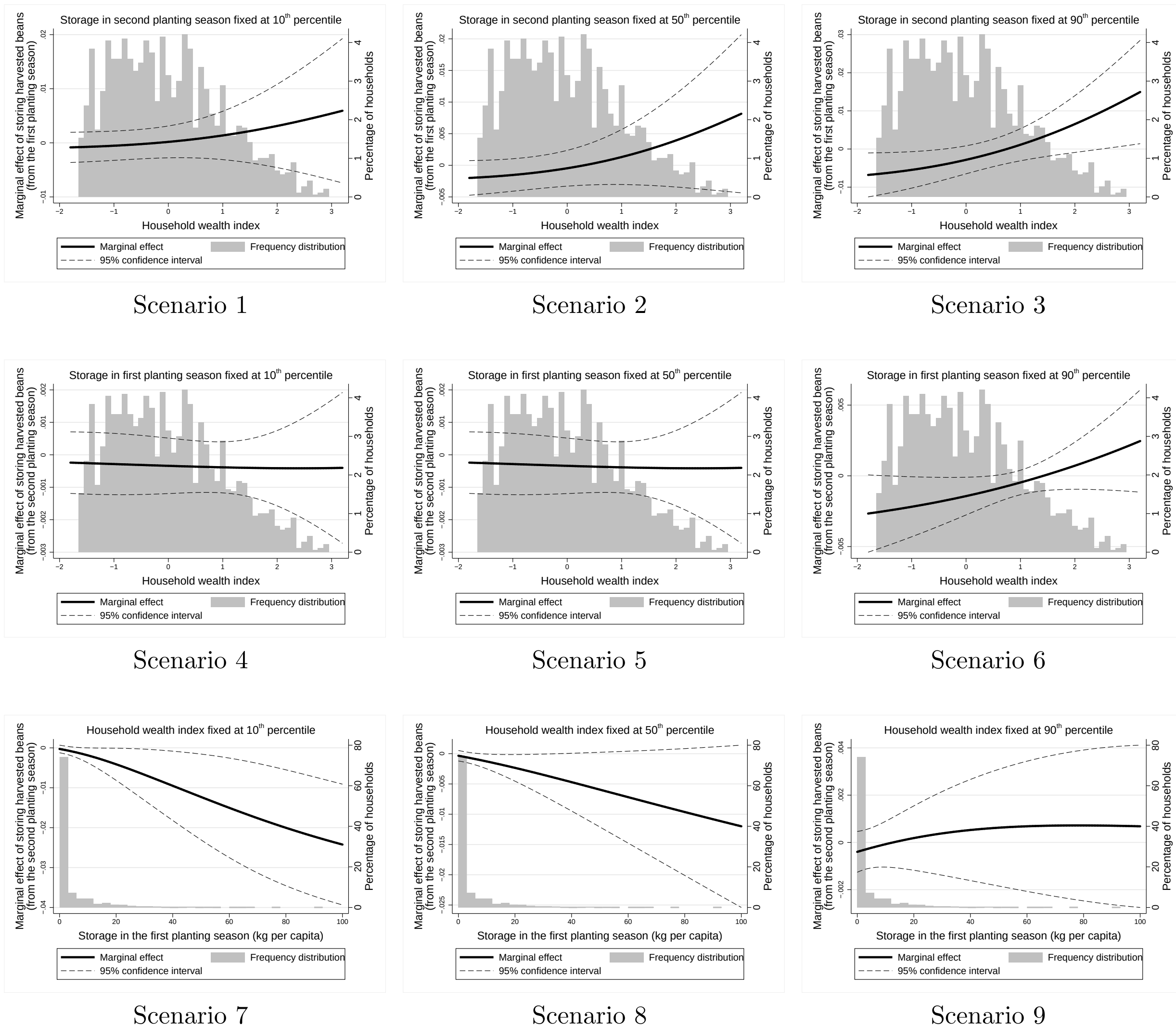
Table 7: Regression output for household-level food insecurity

Variables	Food insecurity (Question 1)	Food insecurity (Question 2)
First season bean storage (S_i)	0.0018	0.0044
	(0.0156)	(0.0106)
Second season bean storage (T_i)	−0.0034	0.0005
	(0.0043)	(0.0036)
Wealth index (W_i)	−0.3953***	−0.3026***
	(0.1040)	(0.0721)
$S_i \times T_i$	−0.0008*	0.0003
	(0.0004)	(0.0004)
$S_i \times W_i$	0.0091	0.0052
	(0.0102)	(0.0070)
$T_i \times W_i$	0.0004	0.0015
	(0.0026)	(0.0022)
$S_i \times T_i \times W_i$	0.0007*	−0.0001
	(0.0004)	(0.0003)
Livestock ownership	−0.0439	−0.2470*
	(0.2200)	(0.1382)
Total farm area	−0.0050	−0.0682**
	(0.0229)	(0.0302)
Distance to farm land	0.0114*	0.0053
	(0.0062)	(0.0037)
Dummy for harvest in first season	−0.4788	−0.4848*
	(0.4817)	(0.2936)
ln(Bean harvest in first season)	−0.0617	−0.0952
	(0.1224)	(0.0800)
Dummy for harvest in second season	−0.9687**	−0.9092***
	(0.4139)	(0.2547)
ln(Bean harvest in second season)	−0.2114***	−0.2594***
	(0.0747)	(0.0516)
Age of household head	0.0014	0.0015
	(0.0069)	(0.0046)
Education of household head	0.0420	−0.0019
	(0.0303)	(0.0204)
Gender of household head	−0.5457***	0.1041
	(0.1975)	(0.1197)
Risk preference of household head	−0.1050	−0.0195
	(0.1107)	(0.0767)
Household size	0.0325	0.0879***
	(0.0366)	(0.0242)
Sharing agricultural information	0.1446	0.2261***
	(0.0900)	(0.0631)
Sharing agricultural inputs and seeds	0.0650	−0.1007*
	(0.0890)	(0.0606)
Observations	1,829	1,829
Pearson-Windmeijer goodness-of-fit test (p-value)	0.5511	0.9990

Note: Asterisk (*), double asterisk (**), and triple asterisk (***) denote significance at 10%, 5%, and 1%, respectively. All the regressions above include the dummy variables for municipality.

of the test is high so we cannot reject the null hypothesis. As can be seen in Table 2, the difference between the two dependent variables is that the first question captures the worry about running out of food, while the second question refers to having no food in the household. Regression results show that the harvest amount in the second season is highly significant in reducing both cases of food insecurity, whereas the harvest amount in the first season is insignificant. This shows the relative importance of the harvest amount in the most recent season (one to three months before the food insecurity measure was taken), rather than the previous season (seven to nine months before the food insecurity measure was taken). Owning livestock also reduces the likelihood that a household will run out of food. The odds ratio, $\exp(\beta)$, is 0.78, meaning that the odds of being food insecure are lower for households that own livestock. Their odds are about 0.78 times the odds for the households without livestock.

Table 5 explains the nine possible scenarios that are analyzed for each household-level food insecurity question from the full factorial of the three variables of wealth and harvest storage decisions for both seasons. We plot in Figure 1 the results from all the possible scenarios for the first food insecurity question. The first two rows of graphs illustrate the change in the average marginal effects of the harvest storage decision on food insecurity across the full range of possible wealth levels. Only Scenario 3 and (marginally) Scenario 6 have a statistically significant range (i.e., the 95% confidence interval is fully above or below zero) at the lower end of the wealth index. The effect is insignificant otherwise. This means that the harvest storage decisions from both seasons complement each other. Only when the harvest storage is high in one season, can an increase in harvest storage in another season help to reduce the worry that the food would run out, and this applies only to the farmers with less wealth. This is reflected in the third row of graphs, in which only Scenario 7 contains a range that is statistically significant. This row captures the change in the average marginal effects of the storage decision in the second season across the full range of storage amounts in the first season. Scenario 7 shows the situation in which wealth is held constant at the 10^{th} percentile, while Scenarios 8 and 9 capture the changes when wealth is held at



Scenario 1 Scenario 2 Scenario 3

Scenario 4 Scenario 5 Scenario 6

Scenario 7 Scenario 8 Scenario 9

Figure 1: Changes in average marginal effects of storage decision on being worried about food insecurity across different wealth levels and different storage amounts in another season (the scenarios mentioned here correspond to the scenarios in Table 5).

the 50$^{\text{th}}$ and 90$^{\text{th}}$ percentiles, respectively. We see that in Scenario 7, only the farmers who have high storage in the first season feel more food secure when they increase their food storage in the second season. Bearing in mind that about 90% of the households said that they were worried the food in the household might run out during the three-month reference period of the food insecurity measures, it is not surprising that only higher harvest storage in both seasons (and not one season) has an effect on improving the sense of security. This effect is only seen among the poorer farmers, as the wealthier farmers probably have other means to alleviate food insecurity problems and do not have to rely on the amount stored from harvest.

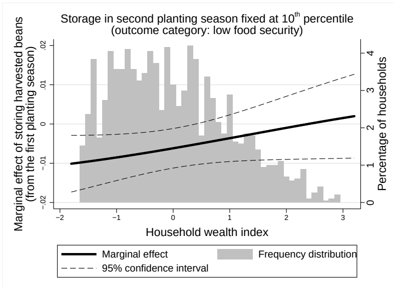
The graphs are not shown for the second food insecurity question because the average marginal effects of storing beans are insignificant across the whole range of wealth and storage amounts. Results from this question, which captures the situation in which food runs out in the household, present an interesting contrast to the results from the first food insecurity question, which is about the sense of food insecurity. Harvest storage in this case does not have a significant impact on whether or not the food actually runs out, not even for the group of farmers who are poor. This is probably because those who choose to sell the beans, instead of keeping them for own consumption, can use the income to purchase other food. However, even though a higher amount of stored harvested beans does not affect the level of food availability in the household, it does help to make the poorer farmers feel more food secure.

4.2 Food insecurity of adults and children

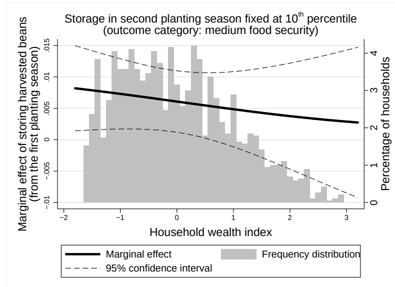
In addition to the two household-level food insecurity measures analyzed above, seven of the nine questions in Table 2 are asked twice for each household: once referring to adults above the age of 18 and once referring to children under the age of 18. The answers to each of these questions are dummy variables. It means that for each household, we have a set of two dummy variables for each question. We create a new ordinal variable with three categories from these answers: household is food secure, the adult or children group is not food secure, both the adult and children groups are not food secure. The analysis of the food storage decision on this ordinal food insecurity variable is carried out using the generalized ordered logit model, the results of which are presented in Section 4.2.1. In cases where the average marginal effects are statistically significant, we analyze the scenarios in more detail by examining whether there are any differences in the effects between the adult and children groups. In this case, we use multinomial logit for the regression with the outcome variable consisting of four categories: household is food secure, only the adult group is not food secure, only the children group is not food secure, both the adult and children groups are not food secure. The results for this more detailed analysis on the differences between the adult and children groups can be found in Section 4.2.2.

4.2.1 Generalized ordered logit

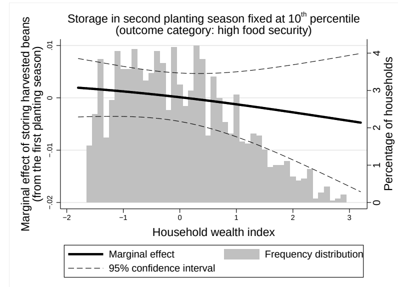
As there are three possible categories for every household from each of the seven food insecurity questions that ask about the situation of the adults and children separately, we have 189 potential scenarios in this section. Due to space limitations, the results are summarized in words with some graphs used to further the explanation. Out of the seven food insecurity questions applicable to this section, the harvest storage decision has a significant impact on only one of them: lack of food variation. Figure 2 shows the impact of the harvest storage decision on food variation. Note that this question focuses on whether the subject always eats the same food, and not necessarily on diet diversity, as a household could have a diet of diverse food groups but still consume the same food every day. The outcome variable of food variation has three categories: low, medium, and high. Low means both adults and children have low food variation. Medium means either the adults or children have low food variation. High means both adults and children have high food variation. Each graph shows the effect of storing more beans on the likelihood of being in a specific food variation category, and how that effect changes across the range of different wealth levels. If the 95% confidence interval is fully above (or below) zero, it means that storing more beans increases (or decreases) the likelihood of being in that specific food variation category. For example, Figure 2(a) shows that a higher bean storage decreases the likelihood of a household being in the low food variation category. However, this effect is only statistically significant for the



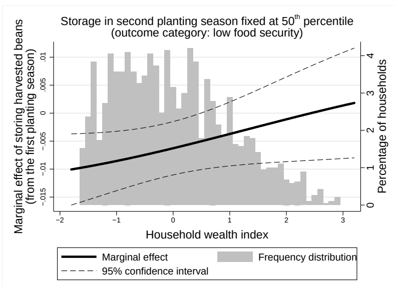
(a) Food variation: low



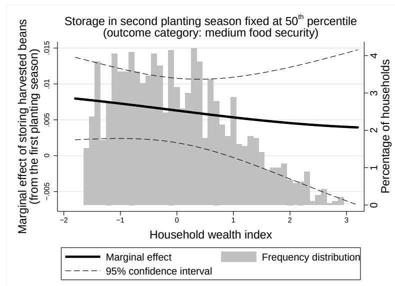
(b) Food variation: medium



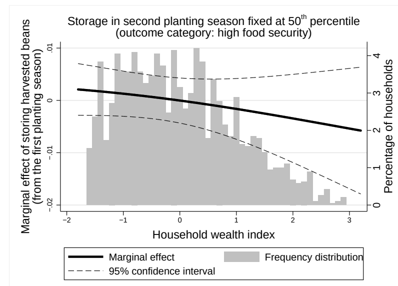
(c) Food variation: high



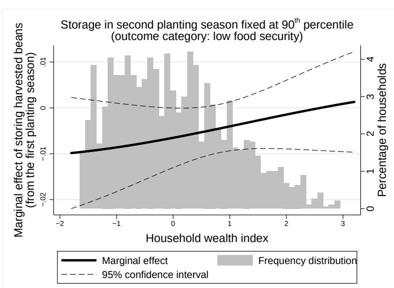
(d) Food variation: low



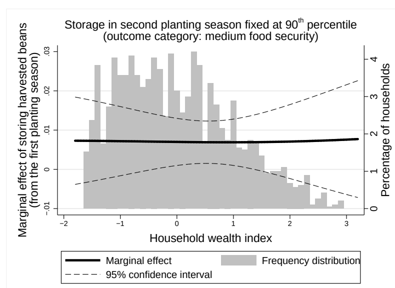
(e) Food variation: medium



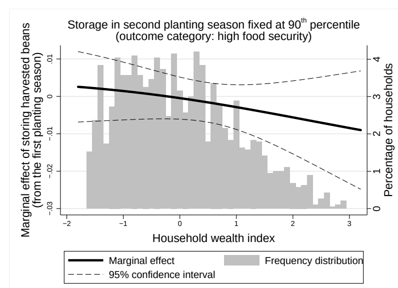
(f) Food variation: high



(g) Food variation: low



(h) Food variation: medium



(i) Food variation: high

Figure 2: Changes in average marginal effects of storage decision on food variation across different wealth levels and different storage amounts in another season, estimated with generalized ordered logit regression.

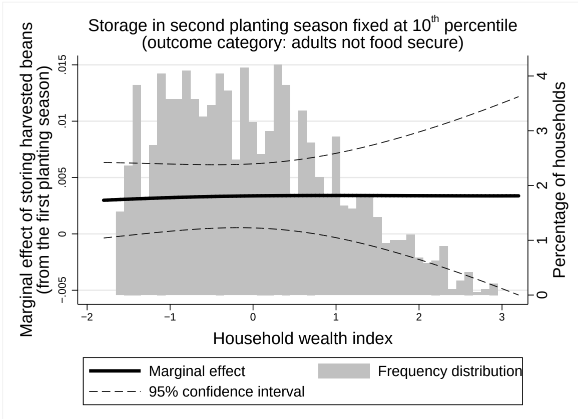
farmers with less wealth. Figure 2(b) shows that storing more beans increases the likelihood of a household being in the medium food variation category, and this effect is also only seen among the farmers with less wealth. The storage decision has no impact on whether a household is in the high food variation category, as illustrated by Figure 2(c). The graphs in the second and third rows show how the effects change when we modify the level of bean storage of another season from the 10^{th} to the 50^{th} percentile, and then from the 50^{th} to the

90^{th} percentile.

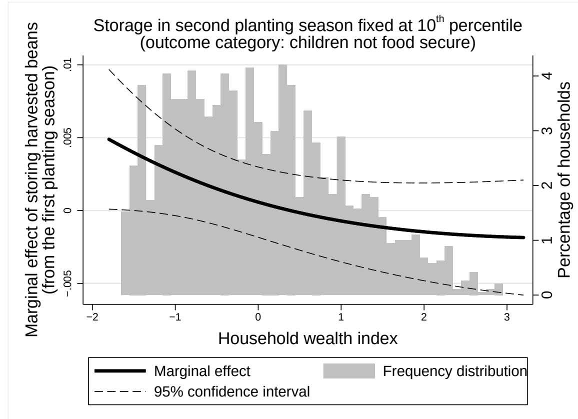
To summarize the results from these graphs in words, they show that a household with higher bean storage is less likely to have both adults and children always eating the same food. However, the household is more likely to have either the adults or the children eating the same food regularly. These effects disappear for the farmers with more wealth or when the amount of bean storage from another season is high. The regression output does not tell us whether it is the adults or the children who are more likely to have low food variation. For that, we will need to carry out another analysis in which we split the adults and children into two separate categories. We will perform this in Section 4.2.2 when we run the regression with multinomial logit. We cannot use ordered logit after we split the children and adults into two separate categories because it is no longer possible to rank the groups in an ascending or descending order when the four categories are: household is food secure, adults are not food secure, children are not food secure, and household is not food secure. The benefit of using multinomial logit is that we can examine the effect of bean storage on the likelihood of being in each category even when we cannot rank the categories, such as in the case of low food variation for children and low food variation for adults. However, it loses the valuable information that ordered categories could provide, ranging from low level of food security to high level of food security.

4.2.2 Multinomial logit

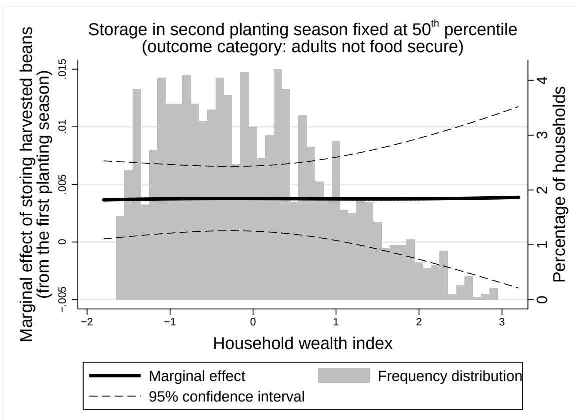
Similar to the situation in Section 4.2.1, but with one more category for the outcome variable, there are 252 potential scenarios in this section. We focus on the food insecurity question about the lack of food variation that is found to be significant in Section 4.2.1, and analyze it further to examine whether there are any differences in the effect of harvest storage on food insecurity between adults and children. There are four categories for the outcome variable: both adults and children have low food variation, adults have low food variation, children have low food variation, both adults and children have high food variation. In the category with high food variation for the whole household, the results are similar to those from the previous section, that is the bean storage decision has no significant impact on the outcome across all the possible values of the other two variables in the interaction terms. The results are also consistent in the category with low food variation for the whole household. Higher bean storage reduces the likelihood of a household being in that category for the farmers with less wealth. Therefore, we present in Figure 3 only the results for the other two categories, the categories that refer specifically to the adults and the children. Graphs in the top two rows show that storing more beans increases the likelihood of low food variation for both adults and children. The effect disappears for the farmers with more wealth and almost disappears for the farmers with a higher storage amount from another season. It affects almost 80% of the adults but only 10% of the children, especially those from the households of least wealth. A reason could be that among poorer households, the adults are more likely to consume food from own production. If they stored more of the harvest, they are more likely to eat the same food regularly. In comparison, the farmers who sell more of their harvest and buy their food mainly from the market are able to have more variation in their pattern of food consumption.



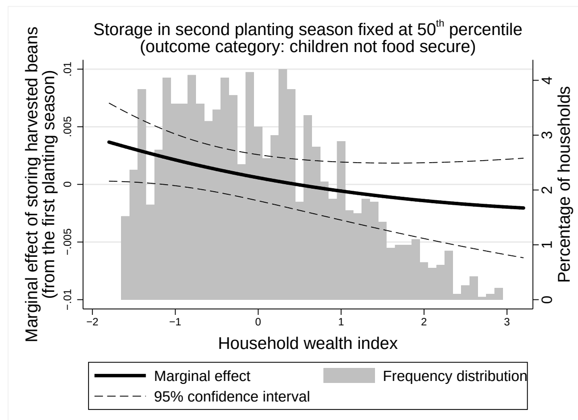
(a) Low food variation for adults



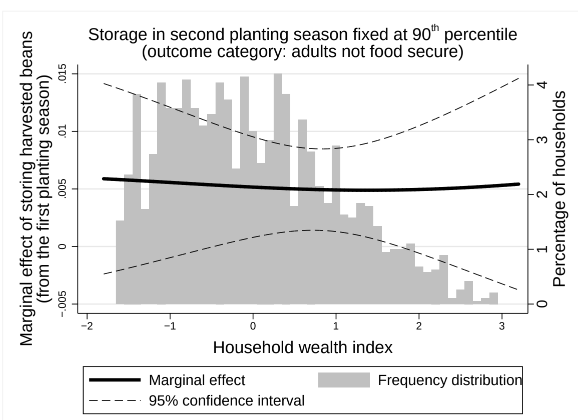
(b) Low food variation for children



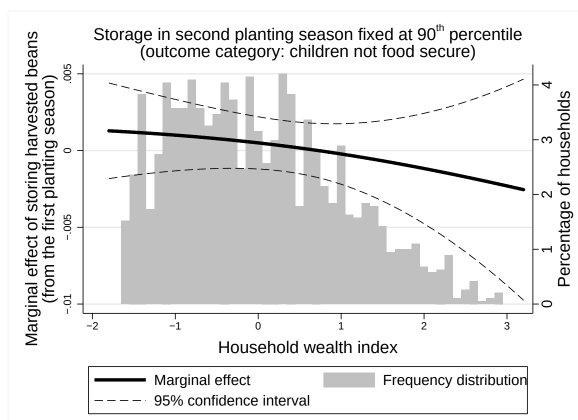
(c) Low food variation for adults



(d) Low food variation for children



(e) Low food variation for adults



(f) Low food variation for children

Figure 3: Changes in average marginal effects of storage decision on food variation across different wealth levels and different storage amounts in another season, estimated with multinomial logit regression.

5 Conclusions

Farmers face a decision after harvest about what to do with their produce, especially the staple crop that they consume regularly. They could keep the harvest for own consumption or sell it in the market and use the income to buy food. In our study, we examine whether this decision affects their level of food security. In addition, as we have the data for the harvest from two previous planting seasons, we examine whether the effect, if any, of storing harvest for own consumption is the same for the two seasons. Instead of examining the seasonality of food security by taking measures of food security at different points in the year, we look at how the staple crop harvest with a different time lag affects the food security of a household. Using the household production and food security data from more than 1,800 households that harvested beans in 2015, we find that the size of the most recent harvest of the staple crop significantly reduces food insecurity, but the size of the previous season's harvest has almost no significant impact on current food insecurity.

By including a full factorial of the three variables of wealth and the harvest storage decisions from both planting seasons, we examine the marginal effects of saving harvested beans for own consumption on food insecurity across different wealth levels and different storage amounts from another season. Results show that storing more beans for own consumption does not have a significant impact on reducing food insecurity, but it does help to improve the sense of food security among the poorer farmers, as they are less likely to worry about food in the household running out. This presents an interesting policy question about whether saving more of a staple crop from own production should be recommended if it does not lead to any tangible improvement in food security. However, if we assume that being worried does negatively affect the livelihood of a household and that storing more of a harvest does not lead to any actual adverse effect on food security, it does seem advisable to increase storage of a harvested staple crop, especially among the less wealthy households.

A potential downside to saving more of a harvested crop is that we do find a somewhat negative impact on different types of food insecurity. Increasing the amount stored seems to lead to a lower variation in the food being consumed among the adults of the households. Therefore, one should make sure that when increasing the storage of a staple crop for own consumption, one does not neglect to consume a variety of food as well. A potential extension to this research topic is to examine whether harvest storage has an impact on diet diversity. As mentioned earlier, the food insecurity measures in this study are based on a list of questions answered by the households, ranging from mild food insecurity to severe food insecurity. In the case of a household running out of food, we do find a different impact from storing beans on the food actually running out and being worried that the food would run out. Therefore, it would be interesting to see whether diet diversity is affected even though based on the questions regarding food variation, it does seem like there is a negative effect from storing more beans on the variety in food consumed.

References

Abdullah, M. and Wheeler, E. F. (1985). Seasonal variations, and the intra-household distribution of food in a Bangladeshi village. *The American Journal of Clinical Nutrition*, 41(6):1305–1313.

Allison, P. D. (2014). Measures of fit for logistic regression. In *SAS Global Forum 2014*, Washington, DC.

Babatunde, R. O. and Qaim, M. (2010). Impact of off-farm income on food security and nutrition in Nigeria. *Food Policy*, 35(4):303–311.

Battese, G. E. (1997). A note on the estimation of Cobb-Douglas production functions when some explanatory variables have zero values. *Journal of Agricultural Economics*, 48(1-3):250–252.

Birol, E., Boy, E., Lividini, K., Moursi, M., de Brauw, A., Gilligan, D., Maruyama, E., Gonzalez, C., Perez, S., Reyes, B., and Khor, L. Y. (2015). Impact evaluation of agricultural and health interventions to alleviate iron deficiency in Guatemala: Updated study design and results from the community listing exercise. In *Guatemala High-Iron Beans Project Workshop*, Guatemala City, Guatemala.

Bouis, H. E. and Welch, R. M. (2010). Biofortification – sustainable agricultural strategy for reducing micronutrient malnutrition in the global south. *Crop Science*, 50(Suppl1):S–20–S–32.

Branca, F., Pastore, G., Demissie, T., and Ferro-Luzzi, A. (1993). The nutritional impact of seasonality in children and adults of rural Ethiopia. *European Journal of Clinical Nutrition*, 47(12):840–850.

Carletto, C., Corral, P., and Guelfi, A. (2017). Agricultural commercialization and nutrition revisited: Empirical evidence from three African countries. *Food Policy*, 67(Supplement C):106–118.

Chinn, D. L. (1976). The marketed surplus of a subsistence crop: Paddy rice in Taiwan. *American Journal of Agricultural Economics*, 58(3):583–587.

Feleke, S. T., Kilmer, R. L., and Gladwin, C. H. (2005). Determinants of food security in Southern Ethiopia at the household level. *Agricultural Economics*, 33(3):351–363.

Garrett, J. L. and Ruel, M. T. (1999). Are determinants of rural and urban food security and nutritional status different? some insights from Mozambique. *World Development*, 27(11):1955–1975.

Greene, W. H. (2003). *Econometric Analysis*. Pearson Education, New Jersey, 5th edition.

Grootaert, C. and Narayan, D. (2004). Local institutions, poverty and household welfare in Bolivia. *World Development*, 32(7):1179–1198.

Henry, C., Sharma, M., Lapenu, C., and Zeller, M. (2003). *Microfinance Poverty Assessment Tool*. Technical Tools Series No. 5. The World Bank and Consultative Group to Assist the Poor, Washington, DC.

Hillbruner, C. and Egan, R. (2008). Seasonality, household food security, and nutritional status in Dinajpur, Bangladesh. *Food and Nutrition Bulletin*, 29(3):221–231.

Hosmer, D. W. and Lemeshow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics - Theory and Methods*, 9(10):1043–1069.

Hotz, C., Loechl, C., de Brauw, A., Eozenou, P., Gilligan, D., Moursi, M., Munhaua, B., van Jaarsveld, P., Carriquiry, A., and Meenakshi, J. V. (2012a). A large-scale intervention to introduce orange sweet potato in rural Mozambique increases vitamin A intakes among children and women. *British Journal of Nutrition*, 108(1):163–176.

Hotz, C., Loechl, C., Lubowa, A., Tumwine, J. K., Ndeezi, G., Nandutu Masawi, A., Baingana, R., Carriquiry, A., de Brauw, A., Meenakshi, J. V., and Gilligan, D. O. (2012b). Introduction of Beta-Carotene-rich orange sweet potato in rural Uganda resulted in increased vitamin A intakes among children and women and improved vitamin A status among children. *The Journal of Nutrition*, 142(10):1871–1880.

Kennedy, E. and Cogill, B. (1988). The commercialization of agriculture and household-level food security: The case of Southwestern Kenya. *World Development*, 16(9):1075–1081.

Long, J. S. and Freese, J. (2014). *Regression Models for Categorical Dependent Variables using Stata, 3rd Edition*. Stata Press.

Mallick, D. and Rafi, M. (2010). Are female-headed households more food insecure? evidence from Bangladesh. *World Development*, 38(4):593–605.

Medani, A. I. (1975). Elasticity of the marketable surplus of a subsistence crop at various stages of development. *Economic Development and Cultural Change*, 23(3):421–429.

Meenakshi, J., Johnson, N. L., Manyong, V. M., DeGroote, H., Javelosa, J., Yanggen, D. R., Naher, F., Gonzalez, C., Garcia, J., and Meng, E. (2010). How cost-effective is biofortification in combating micronutrient malnutrition? an ex ante assessment. *World Development*, 38(1):64–75.

Mujica-Coopman, M. F., Brito, A., Lopez de Romana, D., Rios-Castillo, I., Cori, H., and Olivares, M. (2015). Prevalence of anemia in Latin America and the Caribbean. *Food and Nutrition Bulletin*, 36(2 Suppl):S119–S128.

Muriithi, B. W. and Matz, J. A. (2015). Welfare effects of vegetable commercialization: Evidence from smallholder producers in Kenya. *Food Policy*, 50(Supplement C):80–91.

Norton, E. C., Wang, H., and Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *Stata Journal*, 4(2):154–167.

Pereira, L. M., Cuneo, C. N., and Twine, W. C. (2014). Food and cash: understanding the role of the retail sector in rural food security in South Africa. *Food Security*, 6(3):339–357.

Renkow, M. (1990). Household inventories and marketed surplus in semisubsistence agriculture. *American Journal of Agricultural Economics*, 72(3):664–675.

Saltzman, A., Birol, E., Bouis, H. E., Boy, E., De Moura, F. F., Islam, Y., and Pfeiffer, W. H. (2013). Biofortification: Progress toward a more nourishing future. *Global Food Security*, 2(1):9–17.

Scoones, I. (1998). Sustainable rural livelihoods: A framework for analysis. Working paper no. 72, Institute of Development Studies, Brighton, United Kingdom.

Tarasuk, V. S. (2001). Household food insecurity with hunger is associated with women's food intakes, health and household circumstances. *The Journal of Nutrition*, 131(10):2670–2676.

Toquero, Z., Duff, B., Anden-Lacsina, T., and Hayami, Y. (1975). Marketable surplus functions for a subsistence crop: Rice in the Philippines. *American Journal of Agricultural Economics*, 57(4):705–709.

Van den Broeck, G. and Maertens, M. (2016). Horticultural exports and food security in developing countries. *Global Food Security*, 10(Supplement C):11–20.

von Braun, J. (1995). Agricultural commercialization: impacts on income and nutrition and implications for policy. *Food Policy*, 20(3):187–202.

Weesie, J. (2002). On Pearsons X^2 for categorical response variables. In *Belgian / Dutch / German Stata Users Meeting*, Maastricht, The Netherlands.

Wetterberg, A. (2007). Crisis, connections, and class: How social ties affect household welfare. *World Development*, 35(4):585–606.

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal*, 6(1):58–82.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata Journal*, 12(2):308–331.

Williams, R. (2016). Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology*, 40(1):7–20.

Windmeijer, F. A. G. (1990). The asymptotic distribution of the sum of weighted squared residuals in binary choice models. *Statistica Neerlandica*, 44(2):69–78.

Wossen, T., Di Falco, S., Berger, T., and McClain, W. (2016). You are not alone: social capital and risk exposure in rural Ethiopia. *Food Security*, 8(4):799–813.

Zeller, M., Sharma, M., Henry, C., and Lapenu, C. (2006). An operational method for assessing the poverty outreach performance of development policies and projects: Results of case studies in Africa, Asia, and Latin America. *World Development*, 34(3):446–464.