EVANS SCHOOL OF PUBLIC AFFAIRS
UNIVERSITY of WASHINGTON
Evans School Policy Analysis and Research (EPAR)

Crop Diversity Amongst Smallholder Farmers in Tanzania: What are the Important Factors?

Alexander Chew, Danielle Fitts
Marieka Klawitter, Mary Kay Gugerty, & C. Leigh Anderson

EPAR Brief No. 237

Prepared for the Agricultural Policy Team of the Bill & Melinda Gates Foundation

Professor Leigh Anderson, Principal Investigator
Associate Professor Mary Kay Gugerty, Principal Investigator

June 9, 2013

I. Introduction and Purpose

Local crop diversity and crop cultivation patterns among smallholder farmers have implications for two important elements of the design of agricultural interventions in developing countries. First, crop cultivation patterns may aid in targeting by helping to identify geographic areas where improved seed and other productivity enhancing technologies will be most easily applicable. For example, promoting a new improved rice variety may not provide much benefit in areas not agro-ecologically suited to rice cultivation or where farmers do not have a strong preference for rice. Second, these patterns may help to identify potential unintended consequences of crop interventions focused on a single crop (e.g. maize). For example, research has shown that some farmers cultivate sorghum and millet as a risk mitigation strategy, as these crops are more drought-tolerant than other major cereal crops and produce even in the face of severe water constraints. Drought tolerant maize allows for maize to grow in lower-rainfall environments, but is generally still less tolerant than sorghums and millets. If drought tolerant maize *replaces* sorghum-millet cultivation on marginal lands, farmers cultivating drought tolerant maize may see substantial economic benefits in higher rain years, but may also face catastrophic crop failure in years of severe drought. Increased cultivation of maize relative to other crops may also impact maize prices and affect nutritional outcomes. Understanding the level of substitutability between crops, and the factors that affect crop-cultivation patterns over time is critical to evaluating whether single-crop focused interventions are linked to the level of crop diversity and local and national crop cultivation patterns.

The purpose of this paper is to assess the distribution of crop diversity and crop cultivation patterns, and factors that can lead to changes in these patterns among smallholder farmers in Tanzania with a focus on three main questions.

- What are the regional patterns of crop cultivation and the distribution of crop diversity among smallholder farmers in Tanzania? Have these patterns changed between the years of the panel?
- What are the factors that affect crop diversity and changes in crop diversity?
- What is the level of substitutability between crops grown by smallholder farmers?

All analysis is based on the Tanzania National Panel Survey (TNPS) datasets from 2008 and 2010. (See Appendix A for a description of the data source). The paper is structured as follows. **Section I** provides a description of regional patterns of crop cultivation and crop diversity between the two years of the panel. **Section II** presents background on the theoretical factors affecting crop choice, and presents our findings on the results of a multivariate analysis on the factors contributing to crop diversity. Finally, **Section 3** provides a preliminary analysis of the level of substitutability between cereal crop of importance in Tanzania (maize, rice and sorghum/millet) and also between these cereal crops and non-cereal crops.

NOTE: The findings and conclusions contained within this material are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

II. Regional Patterns in Crop Cultivation and Crop Diversity

Measures of Crop Diversity

The definition of agricultural diversity often depends on the field of expertise of the analyst and the goal of the study.

> *"To a geneticist, diversity may be the number of alleles or the polymorphism found in the genome. An agronomist will focus on variability in traits such as yield, flowering time, plant height, or seed color. An ecologist or conservationist may concentrate on population structure, coevolution, patchy agricultural environments, biogeography, rarity, and abundance of different varieties or traits in relation to area. An anthropologist will look at folk nomenclature and taxonomy, the number of farmers who grow landraces, the number of varieties that are recognized, the link between knowledge and behavior in selection and management, and the exchange and market systems that move diversity between famers and regions" (Brush, 2004, p. 54).*

Ecologists and geneticists use indexes to measure crop diversity in a variety of different ways. Generally, indexes are constructed such that they emphasize richness or evenness. Richness is "a measure of the number of distinct species, varieties, or morphotypes that are found in a population" and does not take into account the number of individuals of each species type present (Brush, 2004, p. 62). Evenness, on the other hand, "shows whether diversity is provided by the presence of a few rare individuals that are distinct or by comparable proportions of different types in a population" (Brush, 2004, p. 62) and "measures how similar the frequencies of the different variants are, with low evenness indicating dominance by one or a few types" (D. I. Jarvis et al., 2008a). In contrast, farmers assess diversity based on the number of varieties in a field or garden as well as the shape, structure, color, pattern, and yield of the crops. While biologists often use scientific methods for measuring diversity, farmers are limited to basing their measures on variation that can be seen with the human eye (Cromwell & Van Oosterhout, 2000).

For this study, we derived three different measures of crop diversity in order to provide estimations of both richness and evenness. A summary of these variables can be found in **Table 1**. Two of the indices used, the Shannon Index and Herfindahl indices, have been termed a "heterogeneity indices" because they accounts for both richness and evenness (M. Smale, Meng, Brennan, & Hu, 2003). These types of indices have been incorporated into similar regression models by other scholars (Benin, Smale, Pender, Gebremedhin, & Ehui, 2004). Measuring both richness and evenness allows for the analysis of the number of distinct crop types as well as the distribution of different crop types across Tanzania (Brush, 2004). It should be noted, however, that "no index is perfect. Any index compresses a lot of information into a scalar and is bound to be ambiguous" (M. Smale et al., 2003, p. 4).

Table 1: Summary of Indices

Index	Measure	Description
Crop Count	$C = \sum_{h=1}^{N} z_h$	*N* = total number of crops cultivated by all households Z_h = number of crops cultivated by *h* household
Shannon Index	$H = \sum_{i=1}^{R} p_i \ln p_i$	*R* = total area devoted to cultivating crops P_i = proportion of cultivated area devoted to *i* crop
Herfindahl (Dominance) Index	$D = 1 - \left(\sum_{i=1}^{R} p_i^2\right)$	*R* = total area devoted to cultivating crops P_i = proportion of cultivated area devoted to *i* crop

- *Crop Count:* Crop count is a measure of crop richness and was calculated by simply summing the number of crops grown by each household. Permanent cash crops and fruits were not included in the sample because we assumed decisions to plant these crops would be different than decisions to plant annual crops. The resulting data contained 37 different crop types (see Appendix C for list of crops and a more detailed description of the variable creation process and some potential data issues).

- *Shannon Index*: The Shannon Index, originally used in information theory, is derived by calculating the proportion of cultivated area of a specific crop relative to the total cultivated area, multiplying the proportion by the natural logarithm, and then summing that product across species. Unlike crop count, the Shannon Index is an important method for calculating drop diversity while controlling for crop scarcity (see Appendix C for a more detailed description of the variable creation process and discussion of some potential data issues). **An increase in the household Shannon index would represent an increase in household crop diversity.**

- *Herfindahl Index:* The Herfindahl Index, also known as the Simpson Index, has been used for measuring the market power of a single firm in an industry in order to assess the degree of competition or the presence of monopolies. The Herfindahl Index is constructed by calculating the proportion of cultivated area devoted to each specific crop type relative to the total cultivated area, squaring that number, summing those proportions across all proportional cultivated areas, and taking the reciprocal (Gross, Mullin, & Riechert, n.d.). **For the purposes of this paper, the complement of the Herfindahl Index was used, known as a Dominance Index**, to conform to the interpretations used for the crop count and Shannon Index measures (see Appendix B for a more detailed description of the variable creation process and discussion of some potential data issues). **An increase in the household dominance indicator would represent an increase in household crop diversity.**

For many farmers in Tanzania, a substantial proportion of household land area was devoted to cereal crops. To show regional variation in cereal crop production, we also derived measures of the proportion of household landholding size devoted to specific cereal crops. In our analysis we focus on the three most commonly grown cereal crops in Tanzania: **maize, rice and sorghum/millet.**

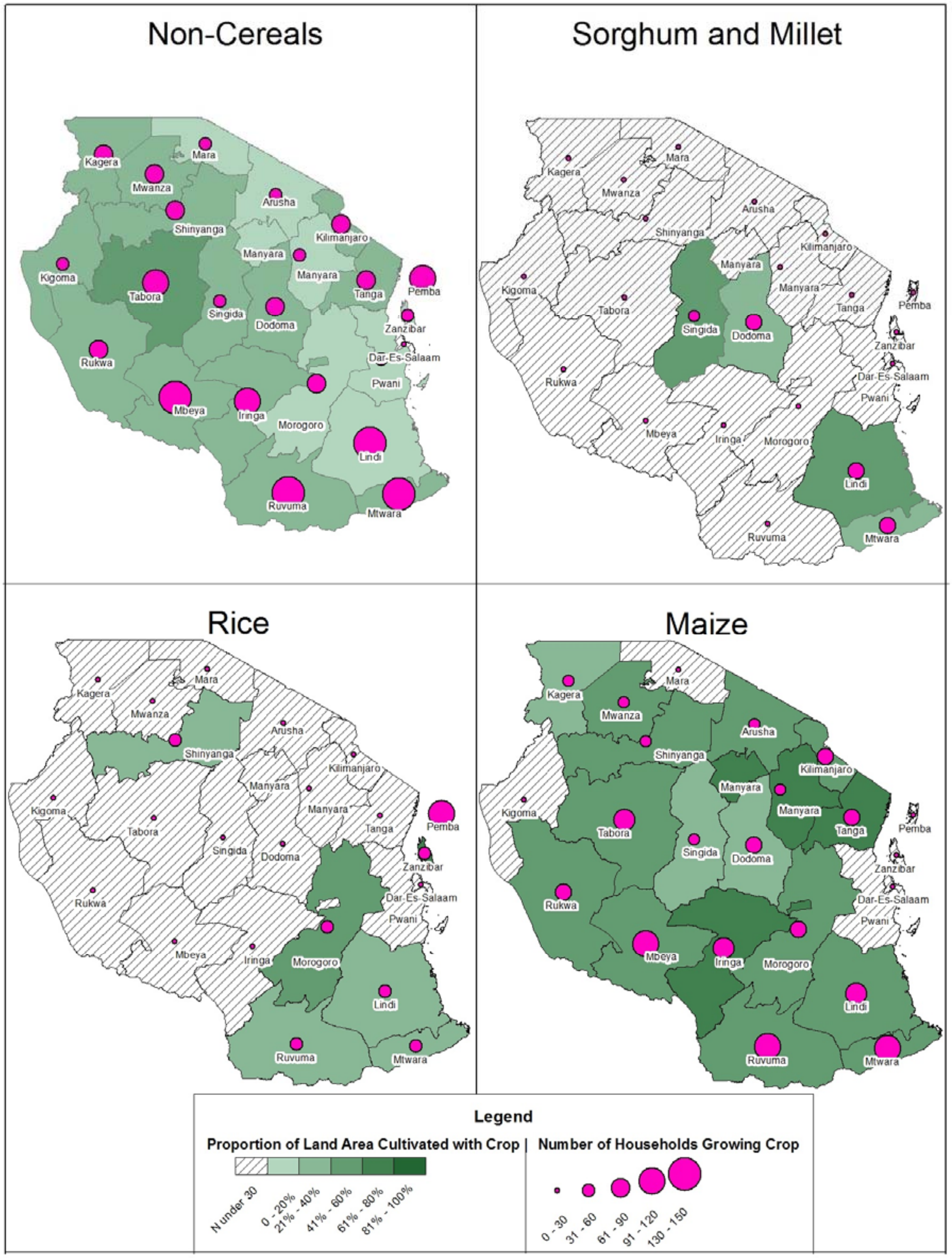
Regional Patterns in Diversity and Cereal Crop Cultivation

Cereal crop cultivation

Crop cultivation of and area devoted to specific crops varies by region. Maize is the most prevalent crop in Tanzania although it is cultivated on a smaller proportion of land in the central region. A greater number of farmers in southern regions reported growing maize. On the other hand, rice is one of the least frequently grown cereal crops, reported by less than 30 farmers in most regions. However, rice was grown on a significant portion of the land area in south regions and Zanzibar although there were a relatively few number of farmers growing it. Like rice, sorghum and millet cultivation were reported by relatively few farmers. Farmers dedicating large portions of land to those crops were located in the central regions, interestingly the same regions that have less maize concentration, and the south. Crops that are not cereals are grown by a large number of farmers in every region in Tanzania but those crops take up a larger proportion of land area in western regions.

Figure 1 shows the regional breakdown of the average proportion of cultivated crop area devoted to maize, sorghum and millet, rice and all non-cereal crops. The green color gradations are bins representing ranges of crop area percentages. The pink dots represent the sample sizes used to calculate the average proportions of area devoted to each crops type, with larger size indicating a larger sample size, and a greater number of regional households cultivating the crop. **We do not report means for sample sizes under 30.**

Figure 1: Land Area Cultivated by Crop, 2008



Crop diversity

Similarly, the change in diversity indices varies by region. The change in crop count suggests that farmers grew fewer crops, on average, in 2010 than they did in 2008, although the change was relatively minor. In 2008, farmers in the northwestern region of Tanzania grew a greater number of crop varieties than other farmers, and the 2010 crop count shows that farmers across the nation grew about the same number of crops. This implies that the crop count changed most significantly for farmers in the northwest while those in the south remained relatively stable and those in the east actually grew more crop varieties. **Figure 2** shows the average household crop count by region in both 2008 and 2010. Color gradations show different ranges of average household crop count by region.

Increases in the Shannon and Dominance indices would indicate increases in the average level of household crop diversity. The Shannon and Dominance Indices tell essentially the same story (**Figures 3 and 4**). Diversity is greater in the central and southern regions and less in the eastern regions. In 2010, the regions became more similar to each other and diversity, overall, decreased. Color gradations represent different ranges of the average value of the household diversity indices by region. Higher values indicate greater diversity.

The most apparent regional patterns in crop diversity, crop cultivation, and changes in crop diversity are:

- Pemba and Zanzibar have much less crop diversity than the rest of Tanzania
- Crop diversity decreased by a small amount in 2010 and there was less variation between regional crop diversity
- The eastern regions tended to be the least diverse regions in 2008 (and those regions devoted less are to non-cereals) but their diversity improved in 2010
- Diversity indices in southern and western regions were least likely to change between 2008 and 2010
- Central regions experienced a decrease in crop diversity in 2010 in every index measure

Figure 2: Average Crop Count by Region, 2008 and 2010

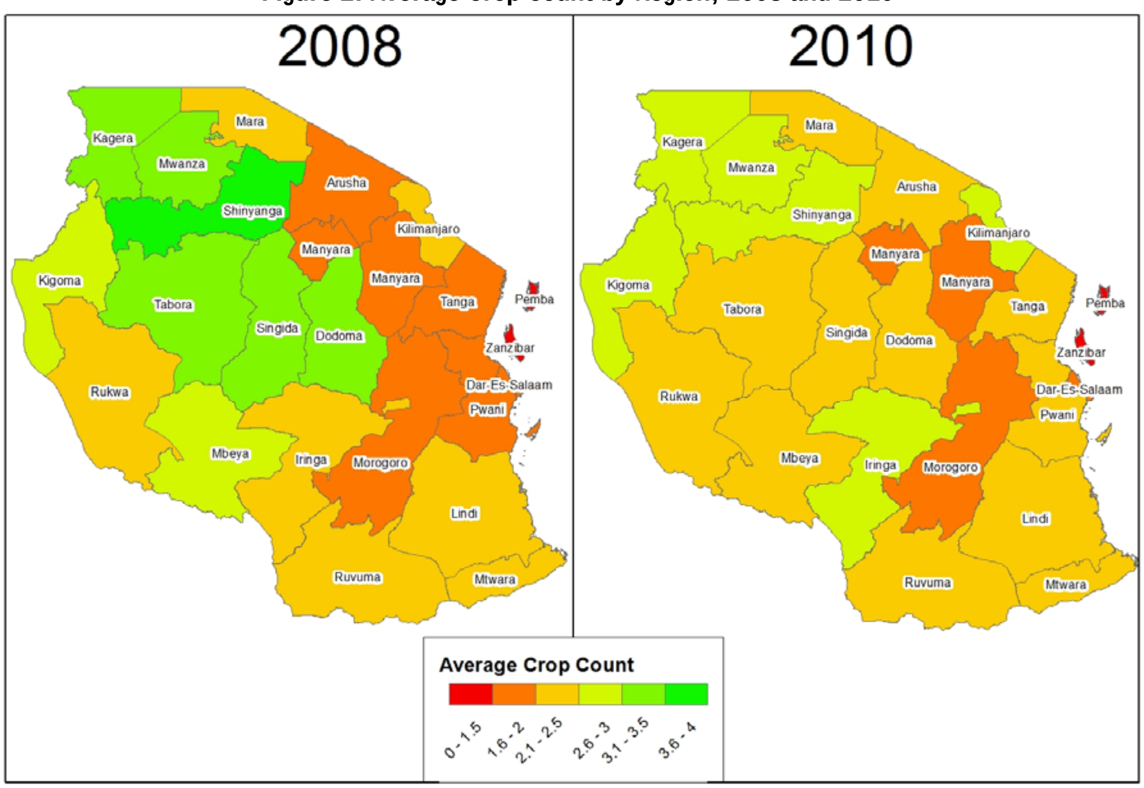
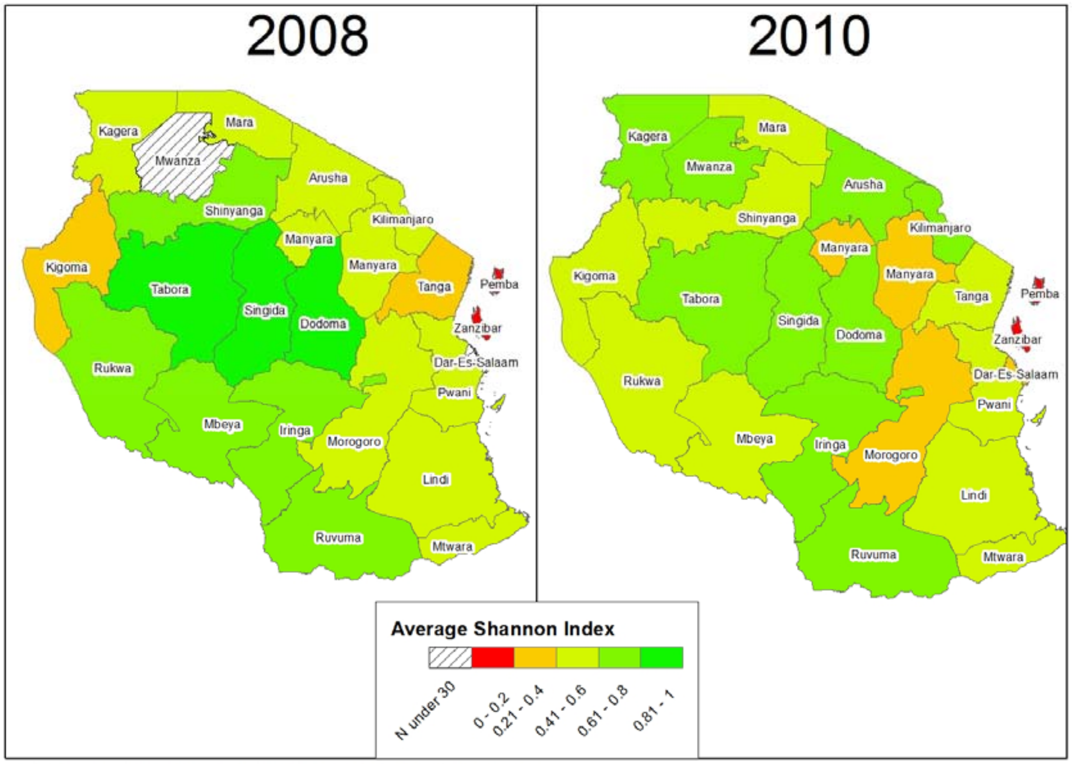
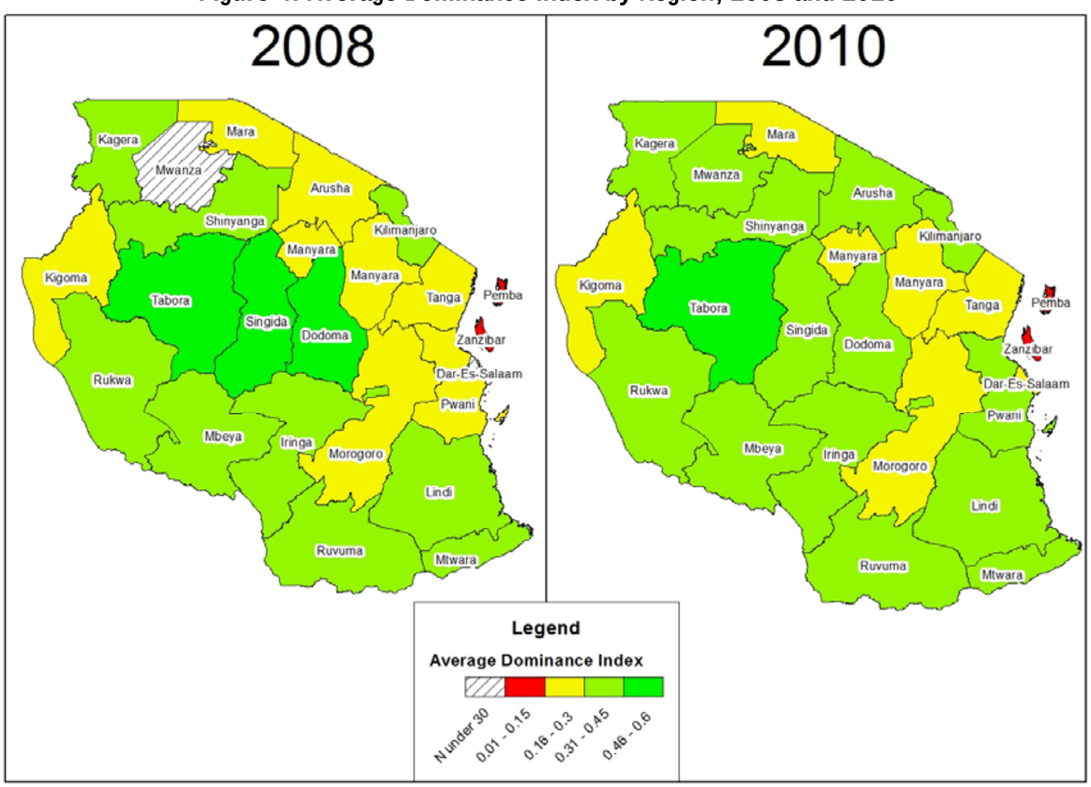


Figure 3: Average Shannon Index by Region, 2008 and 2010



Diversity increases as the Shannon Index increases

Figure 4: Average Dominance Index by Region, 2008 and 2010



Diversity increases as the Dominance Index approaches 1

III. Multivariate Analysis of Factors Affecting Crop Diversity and Changes in Crop Diversity Theoretical Factors Affecting Crop Cultivation Patterns and Diversity

Literature Review on the Factors Contributing to Crop Diversity

Factors affecting farmers' decisions can be broken into three categories: individual farmer characteristics, farmers' preferences, and community-level characteristics. Previous scholarly articles have estimated the impact of the impact of a number of individual factors on crop diversity. It should be noted, however, that many of these studies measure crop diversity based on genetic variation and are concerned with specific geographic areas. The extent to which these previous studies' findings are applicable to this case are not known. However, in aggregate these findings can help to determine the general contributing factors to crop diversity and can be used to inform model specification for this study. These findings are summarized in **Table 2**.

Table 2:
Summary of Previous Research

Variable	Impact on Diversity
Land area planted	+
Area devoted to staple crops	- (proportion of land allocation) + (absolute number of crops)
Extension Services and Financial Assistance	+
Integration with markets	-
Access to transportation, communication, education	+ (number of varieties)
Age of head of household	+
Gender	No consensus
Income	-
Access to resources	+
Membership in Groups	+
Education	None

Individual Farmer Characteristics

- **Age** – One study found that younger farmers were more likely to specialize in specific crop farming, suggesting a positive relationship between age and level of diversity (Van Dusen & Taylor, n.d.). Similarly, another study found that older household heads were more likely to allocate a greater proportion of their farm to small grains (Cromwell & Van Oosterhout, 2000). Another study found that age had no association with diversity (Rana, Garforth, Sthapit, & Jarvis, 2007).
- **Gender** – The division of labor within farming households can impact the level of diversity within the farm. One study proposes that the gender of the decision-maker could have dramatic impacts on crop diversity:

> "Women are usually the seed selectors for the range of criteria required domestically by households, such as taste, color, smell, cooking time, etc. Where a division of labor exists, women are often responsible for staple or subsistence crops and men for cash crops. Women's concern with the household economy provides a balance to the market-oriented pressures that emphasize higher yield and uniformity. In many households, women manage components of the farming system containing high levels of biodiversity, such as home gardens, and make extensive use of gathered species and tree products. Since women often prepare family meals, this influences the variety of crops which they select for the home garden" (Cromwell, Cooper, & Mulvany, n.d., pp. 7–8).

Conversely, other studies have found that the gender of the decision-maker was not a significant factor in crop diversity decisions and determined that gender was more likely related to access to resources (Cromwell & Van Oosterhout, 2000; Rana et al., 2007).

- **Income and Resources** – In general, greater crop diversity is associated with areas of greater poverty and is most concentrated in the least developed regions of countries (Cromwell et al., n.d.). Similarly, poorer families are more likely to allocate a greater proportion of their farm area to small grains (Cromwell & Van Oosterhout, 2000). However, this same study found that access to resources such as land and livestock also play a vital role in crop diversity. The authors found that resource-poor households do not have the ability to grow more than a few crop varieties and families with access to a large number of on and off-farm resources allocate a greater proportion of farm area to small grains (Cromwell & Van Oosterhout, 2000). Similarly, households with more local authority in a community are more likely to grow more crop varieties than those with less authority (Cromwell & Van Oosterhout, 2000).
- **Education** – Several studies found that level of education had no association with diversity (Cromwell & Van Oosterhout, 2000; Rana et al., 2007).
- **Membership Groups** – Participation in membership groups was found to promote the introduction of new diversity in a farming system (Rana et al., 2007). Other studies have found that membership in certain ethnic groups, social status with the community, and gender relations can impact a farmers' capacity to manage crop diversity within a community as well as the farmers' goals and preferences for farming (Cromwell et al., n.d.).

Farmer Preferences

- **Risk** – risk averse farmers to will choose to plant a higher level of crop diversity than risk-neutral farmers in order to reduce uncertainty (Di Falco & Perrings, 2005).
- **End use** – Crops can be used for human consumption, sale, animal feed, building material, and medicines as well as for other uses. The goals of the farmer for crop use will contribute to the amount of diversity within crop varieties (Cromwell, 2001).
- **Culture** – Crop diversity is surely affected by cultural identity although its effects have not been well studied. Cultural identity as observed through gift giving, ritual, prestige, and cuisine can all impact crop diversity decisions (Brush, 2004; Cromwell, 2001).
- **Diversity** - For some farmers, "maintaining diversity is an active and purposeful part of farm management and not an unthinking or reflexive act by farmers who lack alternatives" (Brush, 2004, p. 257).
- **Agromorphological characteristics** – Farmers may decide to plant crops based on preferred characteristics such as "early flowering, height, denseness or inflorescence, or a particular color, shape, or taste" (D. Jarvis & Hodgkin, 2000, p. 270).
- **Farm management practices** – These include "land preparation, planting, thinning and weeding, fertilizer application, pest control, irrigation, harvesting, and post-harvest processing" (D. Jarvis & Hodgkin, 2000, p. 271).
- **Size of plant population** – The number of seeds planted will impact overall diversity within the planted area and can impact genetic variation of the crop population over time (D. Jarvis & Hodgkin, 2000).
- **Seed source** – Each year, farmers decide whether to plant his or her own seeds, acquire seeds from another farmer, or purchase seeds in a marketplace. These decisions will impact the overall crop diversity as well as genetic variation (D. Jarvis & Hodgkin, 2000).
- **Area Planted** - In general, larger planted areas are associated with greater diversity (Cromwell & Van Oosterhout, 2000; Joshi, Upadhyay, Baniya, & Gauchan, 2013a; Rana et al., 2007).
- **Area Devoted to "Staples"** – In one study, staples were found to have higher measures of richness (number of species) and evenness (relative abundance of species) than other crop types (D. I. Jarvis et al., 2008b). Connected to this concept is the finding that families who values small grains highly were more likely to grow more crop varieties than families who do not (Cromwell & Van Oosterhout, 2000). Although not measured in this study, many previous researchers have found that farmers devoting large areas of land to a single crop are more likely to grow diverse varieties of that crop. In addition, farmers growing a diversity of a single crop are more likely to grow a diversity of a second crop (Joshi, Upadhyay, Baniya, & Gauchan, 2013b). Another study addressed this concept by measuring the "maize-mindedness",

the degree to which resources are centered around the production of maize, of farmers in Zimbabwe. The authors found that farmers who allocate a smaller proportion of land to maize allocate a greater proportion to small grains. They concluded that "maize-mindedness" has a negative effect on crop diversity by reducing the amount of land dedicated to small grains and by reducing the number of small grain varieties planted. However, the authors also found that although the proportion of land allocated to small grains by poorer households is greater, the absolute number of varieties grown by richer households is relatively larger (Cromwell & Van Oosterhout, 2000).

Community-Level Characteristics

- **Environmental Conditions and Planting location** – Diversity is impacted by the expected likelihood of uncontrollable factors such as weather, pest infestations, or disease outbreaks that will affect yield (Di Falco & Perrings, 2005). The environment of the site (e.g. prone to flooding) will impact the crop choice as well as the genetic makeup of the specific varieties. The optimum yield for each crop variety will depend on the soil quality, availability of water, and disease risk. Mitigating the risk of variation in crop yield between seasons will depend on the incidence of drought and disease in a given area (Cromwell, 2001).
- **Economic Conditions -** Farmers make decisions about allocating land to certain crops are based on relative prices and financial incentives (Di Falco & Perrings, 2005).
- **Extension Services and Financial Assistance** – As mentioned above, risk averse farmers tend to plant a greater diversity of crops. However, financial assistance for farmers has been found to stabilize farmers' revenues and reduce risk for farmers (Di Falco & Perrings, 2005). Access to extension services was found to promote the introduction of new diversity in a farming system and to be positively associated with greater crop diversity (Cromwell & Van Oosterhout, 2000; Rana et al., 2007).
- **Community Makeup** – Access to crop diversity, the existence of cultural diversity within a community, and level of exposure to external influences such as modernization and consumerism can impact the capacity to manage and the willingness to grow diversity crops within a community (Cromwell et al., n.d.). Farmers in communities that are less integrated into a market network are more reliant on crop diversity to increase stability and improve productivity (Cromwell, 2001) and greater integration with markets reduces diversity on farms (Van Dusen & Taylor, n.d.). In contrast, another study found that access to transportation, communication, and educational systems led to decreased genetic diversity within crops but a greater number of varieties grown per farm (Melinda Smale, Bellon, & Aguirre Gomez, 2001).

Model Specification

Dependent Variables

To identify factors associated with changes in crop diversity, we performed a multivariate regression analysis to explain variation in each of our crop diversity indicators (Herfindahal Index, Shannon Index, and the household crop count) in 2008. We also carried out regression using the *change* in these diversity indicators between 2008and 2010 as the dependent variable. We calculated the change as the difference between the diversity indicators between the two years (e.g. (Crop Count in 2010)-(Crop Count in 2008).

Independent Variables

We included the following explanatory variables in all of our models.

All Models

a. **Inorganic fertilizer** (Dummy variable indicating a household used inorganic fertilizer during 2008)
b. **Improved variety seed** (Dummy variable indicated that a household used at least on improved seed variety in 2008)
c. **Age and age squared** (Age and squared age of the household head in years)
d. **Distance to nearest market center** (Measured in kilometers)
e. **Education level** (Years of schooling of the household head)

f. **Gender of household head** (Dummy for whether a household was female-headed)
g. **Consumption** (Daily household consumption variable in USD)
h. **Extension** (Dummy for whether a household received agricultural extension during 2008)
i. **Household members** (Number of adults and children in the household)
j. **Farmer coop** (Dummy for the presence of a farmer coop in the village)
k. **Crop area cultivated in 2008** (Sum of areas devoted to all crops in 2008)
l. **Region fixed effects** (Vector of dummy variables for Tanzania administrative regions)

The following variables were *ONLY* in the models where the dependent variable was the *change* in diversity indicators between 2008 and 2010.

m. **Change in crop area**(Difference in total area devoted to all crops between 2008 and 2010)
n. **Household Split**[1] (Dummy variable indicating that households from the 2008 sample had split into multiple household in 2010).

Model specification

We ran the following model types for dependent variables (**Table 3**). These models were all run using Stata. Regression observations were weighted to account for over selection in some regions using the TNPS household weights.

Table 3:
Model Specifications

Dependent Variable	**Model Type**
Household Crop Count (2008)	Poisson
Household Herfindahl Index (2008)	Ordinary Least Squares
Household Shannon Index (2008)	Ordinary Least Squares
Change in the Household Crop Count (2008-2010)	Ordinary Least Squares, Multinomial Logit
Change in the Household Herfindahl Index (2008-2010)	Ordinary Least Squares, Multinomial Logit
Change in the Household Shannon Index (2008-2010)	Ordinary Least Squares, Multinomial Logit

See Appendix D for descriptive statistics and histograms for our independent and dependent variables.

Model Results: Ordinary Least Squares

The results of our OLS models are shown in **Table 4.** Cultivated area was significantly and positively correlated with crop diversity for all of the 2008 crop diversity regressions. An increase in in cultivated area between 2008 and 2010 was also associated with a positive increase in crop diversity. Other variables were less consistent across models. Household use of an improved seed variety was positively and significantly correlated with household diversity indicators in 2008. However, the sign on this variable changed to negative in all of the change in diversity regressions and was not significant. The same was true of the variable indicating whether a household used inorganic fertilizer. Age of household head was marginally significant in some models, as was whether a household had received agricultural extension in 2008. **The regional dummy variables were not reported in the table, but some of them were significantly correlated with household crop diversity. However, the signs could be either positive or negative.**

Table 4:

[1] In cases where households did split into multiple households, we followed the observation associated with household head in 2008.

OLS Model Results							
	Crop Count Regressions			**Dominance Index Regressions**		**Shannon Index Regressions**	
	Crop Count, 2008	Crop Count, 2008 (Poisson)	Change Crop Count, 2008-2010	Dominance Index, 2008	Change Dominance Index, 2008-2010	Shannon Index, 2008	Change Shannon Index, 2008-2010
Improved Seed (Dum)	0.320***	0.118***	-0.179	0.0411*	-0.0122	0.0784**	-0.0198
	-3.93	-3.87	(-1.77)	-2.4	(-0.57)	-2.66	(-0.54)
Inorganic Fertilizer (Dum)	0.353**	0.132**	-0.0762	0.0437*	0.0119	0.0962*	0.0134
	-2.93	-2.82	(-0.61)	-2.01	-0.52	-2.42	-0.32
Age of household head(yrs)	0.0237*	0.00990*	-0.00677	-0.00076	0.0031	-0.000282	0.00427
	-1.98	-2.04	(-0.44)	(-0.31)	-0.84	(-0.07)	-0.7
Age of household head squared (yrs)	-0.000214	-0.000089	0.0000709	0.00000662	-0.000028	0.00000262	-0.0000365
	(-1.92)	(-1.95)	-0.49	-0.28	(-0.78)	-0.06	(-0.61)
Distance to the nearest market center (Km)	0.00101	0.000349	0.000383	0.000134	-0.000217	0.000291	-0.00027
	-1.28	-1.11	-0.46	-0.78	(-1.06)	-1	(-0.79)
Education level of household head	-0.0178	-0.00702	-0.00621	-0.00425	-0.00151	-0.00726	-0.00203
	(-1.51)	(-1.47)	(-0.50)	(-1.78)	(-0.54)	(-1.81)	(-0.44)
Female headed household (Dum)	0.0254	0.00474	-0.115	0.0315	-0.0263	0.0477	-0.0483
	-0.36	-0.17	(-1.23)	-1.95	(-1.28)	-1.75	(-1.36)
Daily consumption (USD)	-0.0346	-0.014	0.0498	-0.00455	0.00142	-0.00755	0.00153
	(-0.87)	(-0.80)	-1.02	(-0.50)	-0.12	(-0.50)	-0.08
Agricultural extension (Dum)	0.169*	0.0674*	-0.0144	0.0308	0.00561	0.0630*	-0.00468
	-2.13	-2.29	(-0.14)	-1.81	-0.27	-2.14	(-0.13)
Number of household members	-0.00428	-0.00124	0.0125	0.00285	-0.00377	0.00367	-0.00516
	(-0.31)	(-0.24)	-0.65	-0.97	(-0.99)	-0.73	(-0.77)
Presence of farmer coop (Dum)	-0.00563	-0.00379	0.0635	0.000297	-0.00332	-0.00179	-0.00361

	(-0.10)	(-0.18)	-1	-0.03	(-0.21)	(-0.09)	(-0.14)
Cultivated crop area (acres)	0.0492***	0.0165***	0.00548	0.0132***	-0.00272	0.0250***	-0.0041
	-5.84	-6.04	-0.46	-6.75	(-1.26)	-7.14	(-1.07)
Change in cultivated crop area between 2008 and 2010 (acres)			0.0432***		0.0116***		0.0214***
			-3.84		-5.18		-4.98
Household split			-0.0828		-0.00771		-0.0186
			(-0.88)		(-0.34)		(-0.50)
Constant	1.757***	0.616***	-0.241	0.339***	-0.0657	0.517***	-0.114
	-5.57	-4.84	(-0.60)	-4.92	(-0.68)	-4.34	(-0.71)
N	1587	1587	1372	1587	1372	1587	1372
R-sq	0.257		0.111	0.212	0.089	0.244	0.103
* p<0.05, ** p<0.01, *** p<0.001 Standard Errors in parentheses							

Model Results: Multinomial Logit

We might think that different factors are associated with increase and decreases in household crop diversity. Based on the change in household diversity variables in our previous OLS models, we created three categorical variables that took a 1 if the household had reduced household crop diversity between 2008 and 2010, a 2 if the household had seen no change, and 3 if the household had increased crop diversity. These categorical variables were the dependent variables in our multinomial logistic regression models. The explanatory variables were the same as in our OLS models. Multinomial logistic regression models allow for us to test the influence of the explanatory variables in comparison to a base outcome. In our case, we tested the influence of our explanatory variables on the predicted probability that household would have reduced or increased crop diversity relative to no change.

The results of our multinomial logistic regression models are shown in **Table 5.** The Negative columns in the table show the coefficients for households that reduced crop diversity, in comparison to no change. The Positive columns in the tables show the coefficients of household that increased crop diversity, relative to no change.

Strangely, these results show that an increase in household crop area cultivated led to an increase in crop diversity indices, for negative and positive change households, in comparison to households that had no change. There was one exception to this outcome; the Crop Count index shows that a change in crop area has opposite effects on the probability of positive and negative change in crop count compared to no change. Regional dummies are not shown in the table, but some of them were significant, with signs going in both directions.

Table 5: Multinomial Logit Results

	Change Crop Count, 2008-2010		Change Dominance Index, 2008-2010		Change Shannon Index, 2008-2010	
Direction of Change	Negative	Positive	Negative	Positive	Negative	Positive

Improved seed (Dum)	0.0892	-0.418	0.39	0.193	0.357	0.224
	-0.45	(-1.93)	-1.5	-0.73	-1.37	-0.86
Inorganic Fertilizer (Dum)	0.245	-0.0224	0.642*	0.347	0.655*	0.33
	-1.12	(-0.10)	-2.01	-1.05	-2.05	-0.98
Household split	-0.185	-0.482*	0.0356	-0.0946	0.0448	-0.107
	(-0.88)	(-2.10)	-0.12	(-0.34)	-0.15	(-0.39)
Age of household head(yrs)	0.00353	-0.00441	-0.0264	0.00789	-0.0211	0.00298
	-0.11	(-0.16)	(-0.71)	-0.21	(-0.57)	-0.08
Age of household head squared (yrs)	-0.0000192	0.0000249	0.000236	-0.0000958	0.000184	-0.000048
	(-0.06)	-0.09	-0.69	(-0.29)	-0.54	(-0.14)
Distance to the nearest market center (Km)	0.0012	0.000422	0.000767	0.0000521	0.00112	-0.000272
	-0.73	-0.24	-0.32	-0.02	-0.47	(-0.12)
Education level of household head	-0.0249	-0.0472	-0.0502	-0.0608	-0.0542	-0.0573
	(-1.06)	(-1.70)	(-1.40)	(-1.78)	(-1.56)	(-1.63)
Female headed household (Dum)	0.209	0.179	-0.0817	-0.144	-0.133	-0.0972
	-1.24	-0.78	(-0.33)	(-0.59)	(-0.52)	(-0.41)
Daily consumption (USD)	-0.1	0.0227	-0.214	-0.0688	-0.21	-0.0711
	(-0.99)	-0.21	(-1.86)	(-0.61)	(-1.81)	(-0.63)
Agricultural extension (Dum)	0.0777	0.132	0.316	0.32	0.305	0.331
	-0.39	-0.68	-1.19	-1.21	-1.16	-1.23
Number of household members	0.0251	0.0667	0.0532	0.0566	0.0489	0.0612
	-0.68	-1.77	-0.96	-1.16	-0.89	-1.24
Presence of farmer coop (Dum)	-0.0865	0.0488	0.153	0.249	0.154	0.247
	(-0.59)	-0.37	-0.52	-0.83	-0.53	-0.83
Cultivated crop area (acres)	-0.0125	0.0124	0.159**	0.146*	0.159**	0.146*
	(-0.51)	-0.48	-2.8	-2.49	-2.8	-2.48
Change in cultivated crop area between 2008 and 2010 (acres)	-0.0461	0.0631***	0.107**	0.188***	0.107**	0.189***
	(-1.61)	-3.61	-2.77	-4.15	-2.76	-4.14
Constant	-0.144	-0.745	2.222*	1.311	2.179*	1.34
	(-0.16)	(-0.95)	-2.27	-1.29	-2.24	-1.32
N	1372		1372		1372	

* p<0.05, ** p<0.01, *** p<0.001
Standard Errors in parentheses
Note: Reference group for all models is zero change.

IV. Analysis of Substitutability of Frequently Grown Crops in Tanzania

If two crops were substitutes, we would expect a strong negative correlation between the percentages of land area devoted to each crop. We would also expect a strong correlation between *changes* in the land area devoted to each crop between the two years of the panel dataset. In our analysis, we focused on testing the level of substitutability between cereal crops (Maize, rice and Sorghum/Millet) and between cereal crop and all other crops. We used two methods to assess the level of substitutability.

- **Seemingly Unrelated Regressions (SUR):** SUR models allow for the simultaneous estimation of multiple OLS equations, whose errors are assumed to be correlated. We specified six SUR regression models, each containing two OLS regression equations, where the dependent variable was the percentage of household landholding size devoted to a one of four crops (maize, rice, sorghum millet, and all non-cereal crops. In all cases, the explanatory variables were same as those presented in **Section II.** SUR models allow for the calculation of the level of correlation between the residuals of the two specified OLS equations. The level correlation of the residuals can be interpreted as the extent to which changes in one crop area are associated with changes in the other crop area, while controlling for household and regional farmer characteristics. We also specified an additional six SUR equations where the dependent variables were the changes in percentage of household land area devoted to each crop between 2008 and 2010. The correlations of the residuals for the SUR models are shown in **Table 6**. The level of significance was calculated using the Breusch-Pagan test of independence. Although the SUR models allow for us to control for household and regional factors that might affect the percentage of crop area devoted to each crop, STATA does not allow for the use of the **household sampling weights** specified by the TNPS in SUR models. Therefore, the correlations in the table are unweighted.

- **Simple Pairwise Correlations:** We also tested the level of substitutability of crops using simple pairwise correlations. We ran six pairwise correlations between cereal crops, and between cereal crops and non-cereal crops. These simple correlations do take into account the **household sampling weights** for the TNPS, but they do not control for household and regional factors that might affect household crop areas. **Table 6** shows the results of these simple correlations.

Table 6: Correlations between crop areas, and crop area changes between cereal and non-cereal crops.

	Unweighted		Sample Weighted	
Crops	Correlation of Residuals for SUR (%Crop area 2008)	Correlation of Residuals for SUR(Change in %crop area 2008-2010)	Pairwise Correlations (% Crop area 2008)	Pairwise Correlations (Change in % crop area 2008-2010)
Maize and Sorhum/Millet	-0.4709* (N=249)	-0.6365* (N=249)	-0.3567* (N=329)	-0.4704* (N=329)
Maize and Rice	-0.6802* (N=240)	-0.7162* (N=240)	-0.5460* (N=315)	-0.4758* (N=315)
Sorghum/Millet and Rice	-0.4518 *(N=55)	-0.2139 (N=55)	-0.2959* (N=77)	0.0184* (N=77)
Maize and Non-staple crop	-0.5791* (N=1080)	-0.5410* (N=1080)	-0.5719* (N=1382)	-0.2899* (N=1382)
Rice and Non-staple crop	-0.4522* (N=358)	-0.5090* (N=358)	-0.4795* (N=498)	-0.2598* (N=498)
Sorghum/Millet and Non-staple crop	-0.3800* (N=290)	-0.3339* (N=290)	-0.4382* (N=387)	-0.3156* (N=387)
* Significant at 5%				

The correlations presented in Table 6 suggest that there is generally a high level of substitutability, both between different cereal crops, and between cereal crops and non-cereal crops. Maize area appeared to be highly negatively correlated with the area devoted to other cereal and non-cereal crops. The highest negative correlations in crop area were between maize and rice. Sorghum and millet area appeared to be slightly less highly correlated with area devoted to other cereal and non-cereal crops. A possible explanation is that sorghums and millets are grown on marginal lands not suitable for other crops.

While these results indicate that there may be high level of substitutability between crops, the samples sizes for many of the correlations are relatively small, because only a small number of households grew some of the crop combinations in question. Because correlations were only calculated for household that currently grew both of the crops in question, they may also not reveal whether farmers that do not currently grow a certain cereal crop might be likely to start growing that cereal crop, thereby reducing area devoted to other currently grown crops.

V. Conclusion

In the analysis we have taken some preliminary steps in assessing regional crop cultivation patterns, and crop diversity among smallholder farmers in Tanzania. However, further research is needed to assess whether agricultural interventions are likely to affect regional and national crop cultivation patterns. Key takeaways from this analysis include:

- There are regional patterns in crop cultivation and the distribution of crop diversity and both have shown some change between 2008-2010.
- Regional and household level factors affect both the level of crop diversity and changes in crop diversity.
- Preliminary evidence suggests that there is a high level of correlation between the cultivated area devoted to cereal crops, and to cereal crops and non-cereal crops for farmers growing combinations of cereals and non-cereal crops.

Future directions of research might include:

- An assessment of the factors that might cause farmers to start growing a previously un-cultivated crop variety.
- An assessment of the level of substitutability between disaggregated non-cereal crops.
- Inclusion of a variable to control for crop price might help to improve the explanatory power of all of our models.
- Incorporation of future TNPS panels might allow for confirmation or disproving of the findings presented in this brief.

Limitations of the analysis include:

- The quality of the data used in the calculation of the crop areas is questionable (Issues described in Appendix C). Our findings should therefore be taken with a grain of salt.
- Our analysis does not offer strong insights as to the conditions under which agricultural interventions are likely to affect crop diversity or regional crop cultivation patterns.

.

Appendix A: Data Source

This report uses data collected from the Tanzania National Panel Survey (TNPS) conducted by the Tanzania National Bureau of Statistics (NBS). The first wave was conducted from October 2008 to October 2009 and the second wave took place October 2010 to September 2011. The TNPS is part of the Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA) which supports governments in seven Sub-Saharan African countries to collect panel data focusing on agriculture and rural development.

The TNPS is a national stratified panel survey. The principal strata were the seven administrative zones shown, with Zanzibar added as an eighth. These zones were then divided into rural and urban, creating 16 total strata. Within strata, clusters were chosen randomly, and represent census enumeration areas in urban areas and entire villages in rural areas. The sample gives slightly greater weight to urban areas (due to the higher levels of variability in these areas) and to Zanzibar (in order to allow for separate Zanzibar-specific estimates). The resulting data can produce nationally representative estimates of poverty, agricultural production, and other key indicators. Representative estimates of key variables can also be generated for the eight zones shown above. Sample size limitations preclude reliable statistics at the regional or district level.

The total sample size for the first wave of the survey included 3,265 households (2,063 households in rural areas and 1,202 in urban areas). The second wave of surveys revisited all households originally surveyed and information about all original household members was recorded. If members re-located they were interviewed at their new location or in their new household. The second survey has a total sample size of 3,924 households, representing 3,169 round-one households, a re-interview rate of over 97% (National Bureau of Statistics).

Appendix B: Calculation of Household Crop Count

Our household crop count variable is the sum of annual crop varieties grown Tanzanian households during the 2008 or 2010 growing season. The TNPS dataset provides information on four types of crop cultivation: Long-rainy season (LRS) cultivation, Short-rainy season(SRS) cultivation, permanent crop cultivation and fruit tree cultivation. Crops listed in the LRS and SRS sections of the survey are generally annual crops, while crops listed in the permanent and fruits survey sections have longer growing periods. Under the assumption that land use decisions for annual crops take place on a different scale than for longer-term crops, we limit our analysis to crops grown during the LRS and the SRS. In some cases, crops identified in the survey as permanent or fruit crops were incorrectly listed in the LRS or SRS survey sections. In those cases, these crops observations were dropped from the analysis.

For the year 2008, this left a list of 37 crops cultivated during the agricultural season (**Table 1**).

Crops Cultivated During the 2008 Growing Season					
Maize	Paddy	Field peas	Tobacco	Cowpeas	Pumpkin
Sorghum	Bulrush Millet	Sesame	Cabbage	Chickpeas	Eggplant
Finger Millet	Wheat	Soyabeans	Spinach	Okra	Cauliflower
Sweet potatoes	Irish potatoes	Amaranth	Chilies	Carrot	Fiwi
Yams	Cocyams	Cucumber	Groundnut	Tomatoes	Cotton
Onions	Beans	Sunflower	Pyrethrum	Watermelon	Bambara nuts
Green gram					

Appendix C: Calculation of Crop Areas

The Herfindahl index, the Shannon Index and the calculation of crop area devoted to different crops were all calculated using the farmer-reported proportion of household landholding size devoted to each crop. The quality of this data is questionable for a number of reasons. First, the farmers were only asked to provide very rough estimates of crop-area planted, 100% or the entire plot area, 75% of the, 50% of plot area and 25% of plot area. This means that our estimates of crop area are only very rough, especially in cases where many crops were grown.

Furthermore, inconsistent reporting of crops area, particularly with intercropped plots, complicated calculation of crop areas. In many cases, intercropped plots, were reported to have an area planted that exceeded 100%. In these cases, the reported area estimates were summed (e.g. 50%+50%+= 50%=150%), and each crop area was divided by the sum of the proportions, to yield a normalized estimate of crop area (e.g. 33%).

These factors may both introduce error into our crop diversity indices and crop area estimates.

Appendix D: Descriptive Statistics for the Change in Diversity Models and Multinomial Logits

Dependent Variables	**Mean**	**Std. Erro**	**95%**	**C.I.**	**N**
Change Household Crop count	-0.16407	0.039115	-0.24109	-0.08705	1372
Change Household Shannon Index	-0.02193	0.016381	-0.05418	0.010327	1372
Change Household Herfindahl Index	-0.00857	0.009621	-0.02752	0.010371	1372
Independent Variables					1372
Improved Variety Seed	0.210749	0.015504	0.180221	0.241276	1372
Inorganic Fertilizer	0.168952	0.018939	0.13166	0.206243	1372
Age	47.58701	0.536123	46.53135	48.64267	1372
Age Squared	2507.387	54.8346	2399.414	2615.36	1372
Distance to Nearest Market	78.62114	3.514837	71.70021	85.54206	1372
Gender of Household Head	0.234158	0.013868	0.206852	0.261464	1372
Education Level	4.641714	0.113136	4.418942	4.864486	1372
Consumption	1.224043	0.029072	1.166798	1.281288	1372
Extension	0.23822	0.016261	0.206201	0.270239	1372
Household Members	5.39527	0.095469	5.207286	5.583254	1372
Farmer Coop	0.426388	0.036856	0.353816	0.498961	1372
Crop Area Cultivated in 2008	5.673355	0.609897	4.472432	6.874277	1372
Change in Crop Area(2008-2010)	-0.37047	0.429939	-1.21708	0.476132	1372
Household Split	0.179266	0.011833	0.155966	0.202566	1372

Descriptive Statistics for 2008 *Only* Models

Dependent Variables	Mean	Std. Erro	95%	C.I.	N
Household Crop count	2.53146	0.04206	2.448655	2.614264	1587
Household Shannon Index	0.59792	0.016191	0.566044	0.629796	1587
Household Herfindahl Index	0.635619	0.009258	0.617392	0.653846	1587
Independent Variables					1587
Improved Variety Seed	0.211176	0.014722	0.182192	0.24016	1587
Inorganic Fertilizer	0.161527	0.017802	0.126481	0.196573	1587
Age	47.34667	0.512238	46.33822	48.35513	1587
Age Squared	2489.132	52.88272	2385.021	2593.244	1587
Distance to Nearest Market	77.97791	3.323102	71.43564	84.52018	1587
Gender of Household Head	0.246178	0.013303	0.219988	0.272368	
Education Level					1587
Consumption	1.236806	0.027684	1.182303	1.291309	1587
Extension	0.236288	0.015395	0.205981	0.266596	1587
Household Members	5.358392	0.08717	5.186778	5.530005	1587
Farmer Coop	0.442218	0.037172	0.369036	0.515401	1587
Crop Area Cultivated in 2008	5.37743	0.534437	4.325272	6.429588	1587

Bibliography

Benin, S., Smale, M., Pender, J., Gebremedhin, B., & Ehui, S. (2004). The economic determinants of cereal crop diversity on farms in the Ethiopian highlands. *Agricultural Economics*, *31*(2-3), 197–208. doi:10.1016/j.agecon.2004.09.007

Brush, S. B. (2004). *Farmers' Bounty: Locating Crop Diversity in the Contemporary World* (p. 327). Yale University Press. Retrieved from http://books.google.com/books?hl=en&lr=&id=8XVSlnsAB9YC&pgis=1

Cromwell, E. (2001). Incentive measures at the farmers' level to enhance use and conservation of agricultural biodiversity. *Incentive Measures for Sustainable Use and Conservation of Agrobiodiversity*. Retrieved from http://www.agbioinfo.com/literatura/agricultura/lusakaco.pdf#page=32

Cromwell, E., Cooper, D., & Mulvany, P. (n.d.). Agriculture, Biodiversity and Livelihoods: Issues and Entry Points for Development Agencies. Retrieved from http://www.eldis.org/vfile/upload/1/document/0708/DOC7300.pdf

Cromwell, E., & Van Oosterhout, S. (2000). On-farm conservation of crop diversity: policy and institutional lessons from Zimbabwe. In S. B. Brush (Ed.), *Genes in the Field: On-Farm Conservation of Crop Diversity* (pp. 217–238). Ottawa: IDRC. Retrieved from http://books.google.com/books?hl=en&lr=&id=c1f09mKWxNMC&pgis=1

Di Falco, S., & Perrings, C. (2005). Crop biodiversity, risk management and the implications of agricultural assistance. *Ecological Economics*, *55*(4), 459–466. Retrieved from http://dx.doi.org/10.1016/j.ecolecon.2004.12.005

Gross, L. J., Mullin, B. C., & Riechert, S. E. (n.d.). Alternative Routes to Quantitative Literacy for the Life Sciences. Retrieved June 11, 2013, from http://www.tiem.utk.edu/~gross/bioed/

Jarvis, D., & Hodgkin, T. (2000). Farmer decision making and genetic diversity: linking multidisciplinary research to implementation on-farm. In S. B. Brush (Ed.), *Genes in the Field: On-Farm Conservation of Crop Diversity* (p. 288). IDRC. Retrieved from http://books.google.com/books?hl=en&lr=&id=c1f09mKWxNMC&pgis=1

Jarvis, D. I., Brown, A. H. D., Cuong, P. H., Collado-Panduro, L., Latournerie-Moreno, L., Gyawali, S., Tanto, T., et al. (2008a). A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(14), 5326–31. doi:10.1073/pnas.0800607105

Jarvis, D. I., Brown, A. H. D., Cuong, P. H., Collado-Panduro, L., Latournerie-Moreno, L., Gyawali, S., Tanto, T., et al. (2008b). A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(14), 5326–31. Retrieved from http://www.pnas.org/content/105/14/5326.short

Joshi, B. K., Upadhyay, M. P., Baniya, B. K., & Gauchan, D. (2013a). Area And Diversity Of A Crop Reflects The Diversity Of Others. *Journal of Agriculture and Environment*, *13*, 5–8. Retrieved from http://nepjol.info/index.php/AEJ/article/view/7581

Joshi, B. K., Upadhyay, M. P., Baniya, B. K., & Gauchan, D. (2013b). Area And Diversity Of A Crop Reflects The Diversity Of Others. *Journal of Agriculture and Environment*, *13*, 5–8. Retrieved from http://nepjol.info/index.php/AEJ/article/view/7581

Rana, R. B., Garforth, C., Sthapit, B., & Jarvis, D. (2007). Influence of socio-economic and cultural factors in rice varietal diversity management on-farm in Nepal. *Agriculture and Human Values*, *24*(4), 461–472. Retrieved from http://link.springer.com/10.1007/s10460-007-9082-0

Smale, M., Meng, E., Brennan, J. P., & Hu, R. (2003). Determinants of spatial diversity in modern wheat: examples from Australia and China. *Agricultural Economics*, *28*(1), 13–26. doi:10.1111/j.1574-0862.2003.tb00131.x

Smale, Melinda, Bellon, M. R., & Aguirre Gomez, J. A. (2001). Maize Diversity, Variety Attributes, and Farmers' Choices in Southeastern Guanajuato, Mexico. *Economic Development and Cultural Change*, *50*(1), 201–225. Retrieved from http://www.jstor.org/stable/10.1086/340010

Van Dusen, M. E., & Taylor, J. E. (n.d.). Missing markets and crop diversity: evidence from Mexico. *Environment and Development Economics*, *10*(04), 513–531. Retrieved from http://journals.cambridge.org/abstract_S1355770X05002317