The welfare effects of crop biodiversity as an adaptation to climate shocks in Kenya

Martina Bozzola[1*], Melinda Smale[2]

1 Corresponding Author. m.bozzola@qub.ac.uk. Lecturer, Economics of Food, Nutrition and Health. School of Biological Sciences, IGFS, Queen's University Belfast (UK) and the ZHAW Zurich University of Applied Sciences (CH).

2 Professor of International Development, Michigan State University (USA). msmale@msu.edu

Abstract.

This paper investigates the effects of crop biodiversity on farm income and production risk using a large panel dataset of rural households in Kenya. We consider three different metrics of in situ (on-farm) crop diversification (richness, evenness and concentration). We apply a partial moments-based model to test the effects of each strategy on welfare defined as expected crop income, variability (variance) and downside risk (skewness). Our comprehensive econometric approach differentiates climatic shocks, weather and climate change. The results suggest that the benefits from greater diversification in terms of enhanced land productivity and lower production costs could surpass the foregone benefit from greater efficiency associated with more concentrated production systems. Crop richness and evenness each reduce exposure to crop income risk, especially for more vulnerable farmers who produce below the expected revenue threshold. Farmers who rely on greater crop specialization, on the contrary, are more exposed to crop income risk.

Keywords: Crop diversification, Smallholder farmer, Vulnerability, Kenya

JEL Codes: D81, O13, Q12, Q18

1. Introduction

In Sub-Saharan Africa and other regions characterized by high exposure to changing climatic patterns and frequent weather shocks, smallholder farmers often possess very limited capacity for adaptation. They may use crop diversification as a risk management strategy rather than conceiving crop choices as a way to move from subsistence agriculture to a more integrated participation in local, regional or even global value chains. The need for smallholder farmers to rely on crop diversification strategies primarily as a means of coping with risk may hinder the gradual transition from subsistence-oriented towards more commercialized agriculture. Subsistence-oriented agriculture is characterized by numerous smallholder farmers who produce small quantities of several crops simultaneously, mainly to meet the consumption needs of their families. Commercial agriculture tends to be dominated by fewer larger farms with more specialized products destined for markets.

In Kenya, our country of study, most smallholders produce between these two extremes, with varying crop portfolios depending on the farming system. Agriculture is primarily rainfed, and the majority of smallholders' farmers have limited access to credit and limited or no access to crop insurance (Jensen and Barrett, 2017; Carter et al., 2017). Maize is the most important food staple and a ready source of cash for farm families in all the major smallholder farming systems of the country - typically dominating crop area per farm.

Crop biodiversity is composed of crop infra- and inter-specific diversity. Empirical analyses based on household survey data have shown that crop infra-specific diversity is positively correlated with productivity or profitability in various contexts such as Niger, Ethiopia, Pakistan, and China (Smale et al., 1998; Widawsky and Rozelle, 1998; Asfaw, Pallante and Pala, 2018). This literature has focused on expected crop yield and higher moments of the yield distribution as a proxy of farmers' welfare. Di Falco (2012) and DuVal, Mijatovic and Hodgkin, (2019) review a number of studies that have tested the relationship between various indicators of crop infra-specific diversity and yield variability in rice, wheat, barley and cereals, including yield variance and skewness (downside risk), using farm data at household and regional scales. Generally, these studies have shown that crop infra-specific diversity is associated with less exposure to production risk (Smale et al., 1998; Widawsky and Rozelle 1998; Di Falco and Perrins 2005; Di Falco and Chavas 2009; Di Falco, Bezabih and Yesuf, 2010).

A more recent study by Asfaw, Pallante and Palma (2018) examines the role of crop diversification on adaptation strategies in Niger, applying a two-stage model with seemingly-unrelated regression to estimate the determinants of diversity followed by a quantile model to measure effects on household welfare, including income. Conversely, in a study conducted with household data collected near Mount Kenya, McCord et al. (2015) concluded that household income and suitability of environmental conditions are related to the likelihood of smallholder crop diversification. Also in Kenya, Jaramillo et al. (2011) found that crop diversification might be a suitable adaptation option to cope with warming trends in coffee production. Ochieng et al. (2016) estimated the effects of climate variability and change in crop revenue on maize and tea revenues earned by smallholder farmers in Kenya, finding differences between the two crops; temperature affected crop revenues negatively in maize but positively in tea production, while rainfall had a negative effect on income from tea. An analysis by Wineman et al. (2017) explored the channels through which exposure to extreme weather in Kenya affects the well-being of smallholder farm households, based on longitudinal and spatial analysis of income- and calorie-based measures of welfare. The authors found that extreme weather generally affects household welfare via crop production, recommending the development of new varieties with enhanced tolerance of dry and moist extremes. Bozzola, Smale and Di Falco (2018) find that climate, climate risk and weather play an important role in maize intensification choice. The share of maize area planted to hybrid seeds contributes positively to expected crop income, without increasing exposure to income variability or downside risk.

In this analysis, our hypothesis is that crop diversification could have either a positive or a negative impact on expected crop income, but mitigates vulnerability to adverse factors such as drought spells, variable lengths of growing season and crop losses from pests and disease. Crop diversification may negatively affect crop income because more diversified farmers potentially forego some of the benefit associated with specialization. Previous studies that explored livelihood diversification strategies found that rural households in which each individual is engaged in many activities at low intensity (such as producing many subsistence crops or picking fruits) are often more vulnerable than households where an individual was able to specialize in one activity that is not necessarily related to crop production, such as charcoal burning (see e.g. Eriksen et al., 2005). Constraints to crop specialization are generally higher when labor, capital, insurance, input and output markets function poorly (Klasen et al., 2016). On the other hand, biodiversity may enhance crop income via cost reduction. Some research shows that by enhancing soil fertility and enabling farmers to better manage crop pests and plant diseases, crop diversification and crop rotation is likely to be associated with lower production costs over successive seasons (Tilman et al., 2005, Di Falco and Chavas, 2009). For example, farmers may need to rely less on application of mineral fertilizers to cope with soil degradation, and may be able to reduce their expenditures on plant protection products because of a lower incidence of harbored and carryover pests in a more diversified crop production system. In sum, some adaptation strategies may enhance short-term resilience but hinder better livelihood outcomes in the longer term.

Our analysis contributes to the literature reported above, and to the growing body of literature that explores the relationships among weather, climate and the economy with the aim of understanding the impacts of climate change on human welfare.

We apply three different indices of crop diversification, namely a simple count index (crop richness), the Shannon index of crop evenness and the Herfindahl-Hirschman index of crop concentration. Richness treats each crop equally, while the Shannon index considers the relative area abundance among crops planted on farms. The Herfindahl-Hirschman index is derived from the theory of industrial organization in economics. Applied here, the value of this index reflects the degree to which one or several crops dominate cultivated area per farm. For each of these indices, we test the impact of the underlying crop diversification strategy on crop income, and whether it serves as a partial insurance mechanism for growers by reducing income variability and skewness towards negative outcomes.

Here, we focus on the role of crop inter-specific diversity. The use of crop yields as a dependent variable in the econometric framework applied by earlier studies limits the researcher's capability to capture some of the income benefits of biodiversity. We use crop income, its variance and skewness to assess the welfare effects of biodiversity. We define crop income as the value of crop production minus the costs of purchased inputs and land preparation. We consider that farmers may rely on crop diversification to reduce the variance of crop income and reduce their exposure to downside risk in the presence of climate shocks.

We use a well-established approach for analyzing production risk in agriculture (Antle 1983). Previous applications of this method have posited a mean-variance setting (e.g., Schoengold, Ding and Headlee, 2015), mean-variance-skewness frameworks (e.g., Di Falco and

Chavas, 2006), and more recently, partial moments (e.g., Finger et al. 2018). The latter application provides a better overall perspective on farm-level risks and helps to identify farm-level mechanisms of risk management behavior. In order to distinguish whether these strategies are relevant for farmers who are the most vulnerable to weather shocks, we implement a semi-variance approach. This approach takes into account losses below a specified benchmark (Antle 2010).

We compiled a comprehensive and spatially detailed dataset that comprises socio-economic variables at farm and village level along with detailed weather, climate and soil quality variables at village level.[1] In terms of measurement techniques, we utilize the most advanced drought index available (SPEI). The Standardized Precipitation-Evapotranspiration Index (SPEI) is a multi-scalar drought index that accounts for the fact that the impact of rainfall on the growing cycle of a plant depends on the extent to which water can be retained by the soil. This data richness allows us to test the separate effects of weather shocks on smallholder decisions regarding crop diversification.

We address the potential endogeneity of crop choices in crop income. We apply our model to four waves of panel data collected over a 10-years period in the major agricultural regions of Kenya. We control for time-invariant heterogeneity by applying the Mundlak-Chamberlain procedure (Mundlak, 1978; Chamberlain, 1982) and performing an instrumental variables regression in a two-stage least squares design.

[1] In this article, *weather* indicates the rainfall and temperatures values registered during the main rainfall season of the corresponding data collection year, *climatic shocks* refer to the number of times during the previous decade that a village experienced a serious drought. *Climate* refers to climate normals, which are measured as village average over the period 1971-2010.

2. Theoretical basis

We use a stochastic production function framework to describe the welfare effects of crop diversification, captured by three different indices. We define welfare effects as the impact of different crop diversification strategies on expected crop income and on crop income risks. In defining risk, we assume that farmers maximize their utility with respect to a vector *d* which includes crop diversification decisions:

$$(1) \qquad \max_d E[U(\pi)] = h[E(\Pi(\mathrm{x,d})), Var(\Pi(\mathrm{x,d})),\ SVar^{-}(\Pi(\mathrm{x,d})), Skew(\Pi(\mathrm{x,d}))].$$

As specified in Equation 1, Π is the vector of realizations of crop income (π), which is a function of biodiversity, denoted by *d*, and a vector of other control variables denoted by x. The probability distributions of production and crop income vary with management decisions because management decisions, captured in our setting by vector *d*, influence, among other things, risk exposure (Chavas and Shi, 2015). Vector *x* includes variables that account for environmental conditions and other socio-economic factors influencing farm management choices and adaptation options. We will further describe specific control variables in Section 3.2. The *second* moment of crop income distribution is define as $Var(\Pi(\mathrm{x,d})) = E(\Pi - E(\Pi(\mathrm{x,d})))^2$ while the expression $SVar^{-}(\Pi(\mathrm{x,d})) = E\{\Pi(\mathrm{x,d}) - E(\Pi(\mathbf{x},\mathrm{d}))\}^2\ \forall\ \Pi(\mathrm{x,d}) < E(\Pi(\mathrm{x,d}))$ is a definition of semi-variance, which focuses on crop income realizations below the expected value, and $Skew\ (\Pi(\mathbf{x},\mathbf{d})) = E(\Pi - E(\Pi(\mathbf{x},\mathbf{d})))^3$, denotes the *third* moment of crop income distribution.

Through our theoretical framework, we recognize that variance does not capture the full extent of risk exposure because farmers are also averse to exposure to downside risk, or unfavorable risky events located in the lower tail of the payoff distribution (Antle, 1983, 2010;

Chavas and Shi, 2015). The variance of crop income may overlook this type of risk, as it does not distinguish between "downside risk" and "upside risk". To address this issue, some authors have relied on higher moments such as skewness (e.g., Antle, 1983; Bozzola, Smale and Di Falco, 2018) while other authors determined exposure to unfavorable events using "partial moments", which capture exposure to risk only below (lower partial moments) or above (higher partial moments) some reference point (see Antle, 2010; Finger at al., 2018). Notably, agricultural returns are characterized by extreme loss events and farmers are averse to downside risk, even more so in rural areas in developing countries with very limited coverage from crop insurance. In this context, downside risk exposure can hinder household welfare by creating major distress in term of destroying farm liquidity but also causing food insecurity. For this reasons we also test whether farmers take only into account losses below a specific benchmark when determining their crop diversification strategies.

Equation 1 implies that, in order to capture the full extent of risk exposure, we need to assess the impact of management choices, such as the adoption of different crop diversification strategies, on the distribution of expected crop income, its variance, negative semi-variance and skewness. The semi-variance and the variance differ when the distribution of crop income is not symmetrically distributed (Estrada, 2004). If the distribution of crop income is negatively skewed, representing downside risks, the negative semi-variance is larger than the variance. In this setting, the expected utility approximated with the negative semi-variances is lower than for an approximation with variances (Finger et al., 2018). Downside risk exposure can also be approximated by skewness. Reducing downside risk exposure through the skewness means that the skewness of crop income increases and becomes positive, holding both means and variance constant (Menezes et al. 1980; Di Falco and Veronesi, 2014).

In summary, farmers' utility increases with crop income, lower variance and semi-variance of crop income and if the skewness of crop income increases and becomes positive, meaning that the probability of crop failure decreases.

3. Econometric framework

We follow Antle's moment-based approach to specify the stochastic structure of the model (Antle 1983). We use this econometric approach to test how different forms of crop diversification affect crop income and its higher moments. Control covariates include climate, weather, and socio-economic characteristics of the household.

If crop income were exogenous to crop diversification strategies, the ordinary least squared (OLS) estimate coefficient associated with the crop diversification variables would represent the average treatment effect (ATE) of such variables on crop income. We could estimate a pooled OLS model for each crop diversification index (D^j) of interest:

$$y_i = \alpha + \beta w_r + \omega c_r + \mu x_i + \eta p_r + \varphi s_r + \Omega D_i^j + \xi Z_a + \varepsilon_i \,. \qquad (2)$$

Subscripts i indicate the i^{th} farm household, while the subscript r is used for village-level observations. ε_i is a household-specific error term with mean zero. The dependent variable y_i denotes the hyperbolic sine transformation of crop income for the i^{th} farm household.[2] The coefficients β, ω, μ, η, φ, ξ and Ω represent the vectors of parameter estimates for each

[2] All the non-binary farm-level variables include in this vector are expressed as hyperbolic sine transformations. That is, for each transformed variable x, $\text{arsinh}(x) = \ln(x + \sqrt{x^2+1})$. We do this in order to treat the zero values in the sample, which would result into a reduction of the sample size. Through this transformation, we ensure that all of the logarithms will exist. As a sensitivity analysis, we also obtained the results replacing the hyperbolic sine transformations with a logarithm transformation, adding the constant 1 to each variable before taking the natural logarithm i.e.: $\ln(x)=\ln[1 + (x)]$. Results with the latter transformations were robust and are available upon request.

associated vector of control variables (described below). The superscript j distinguishes three crop diversification indices. Equation 2 is initially estimated once for each crop diversity index: a count index of crop richness, the Shannon index, and the Herfindahl-Hirschman index (Table 1).

The estimation of Equation 2 provides a preliminary robustness check on the empirical results but it does not allow us to make a causal statement about the impact of different crop diversification strategies on crop income. Our main control variables of interest (the crop diversification indices) are potentially endogenous, due to omitted variable bias, unobserved heterogeneity, or reverse causality between crop income and crop diversification. We first test whether the crop diversity indices are endogenous variables using the Davidson-Mackinnon test. If we reject the null hypothesis of exogeneity, we treat the three indices as endogenous in the empirical estimations. We then implement a two stages least squared (2SLS) instrumental variables (IV) approach to mitigate statistical endogeneity concerns (Angrist and Krueger, 2001). The IV approach also has limitations. For example, it is impossible to claim that we control for all possible sources of unobserved heterogeneity across individuals.

We also conducted a multidimensional outlier detection analysis based on the 'bacon' algorithm, which identifies outliers based on the Mahalanobis distances (Weber, 2010; Billor, Hadi and Velleman, 2000). The algorithm allows the identification and removal of observations characterized by implausibly large or low entries of key variables.

We perform three first stage regressions, to determine the drivers of crop diversification choices, and then we use the first stage predictions, instead of the corresponding observed variables, as control variable in the second stage of the estimations. The second stage estimations capture the welfare impacts of crop diversification, as per the theoretical framework presented in

Section 2. We repeat the 2SLS procedure for each of the three crop diversification indexes under scrutiny.

The main challenge of our identification strategy is to find variables that are correlated with crop diversification choices but uncorrelated with the error term of the equations estimating the effects of different crop diversification indexes on expected crop income and production risk. We use as instrumental variables: the frequency of droughts (climatic shocks) over each decade preceding a data collection round and the proportion of households in the village that received credit. Incidences of past droughts is expected to influence smallholders' crop production choices (Di Falco and Veronesi, 2014; Bozzola, Smale and Di Falco, 2018). Already after the Sahel drought sequence of the 1980s, Binswanger and von Braun (1991) observed that in many parts of the region farmers responded to this decade of repeated droughts by changing their crop production choices, which in turn had implication in terms of crop mean yields.

We only refer to the decade preceding the crop year in each data collection round, but exclude the "current year" (time t). For this reason, we expect the SPEI drought index to be correlated with the crop diversification choices made by the households at time t, but not that it would have any direct effect on current household income, nor would it be correlated with the source of unobserved heterogeneity. Credit availability at village level correlates with the crop diversification choices of households because the sentiment of having liquidity constraints is likely to make the farmers more aware of the need to rely on other strategies, such as crop diversification, to cope with possible unexpected weather shocks (see Paxson, 1992).

We selected our instruments on the reasoning that they meet the exclusion restriction, that is, they only affect expected crop income and production risk through crop diversification choices. We address the issue of weak identification of these instruments using F statistics, the results of which are shown in Section 5. For identification purposes, these two instrumental variables can be excluded from the crop income and risk equations (our second stage regressions).

3.1 First stage regressions: determining the drivers of crop diversification choices

In the first stage regressions we represent crop diversification strategies undertaken by the representative farm household as:

$$D_{it}^{j} = \alpha + \beta w_{rt} + \omega c_r + \mu x_{\mathrm{it}} + \eta p_{rt} + \varphi s_{\mathrm{r}} + \varrho k_{rt} + \xi \mathrm{Z_a} + \varepsilon_{it}. \quad (3)$$

The superscript *j* distinguishes three crop diversification indices recorded for farm household *i* at time t. These are: i. a simple count index, denoting crop richness; ii. The Shannon index of crop evenness and iii. The Herfindahl-Hirschman index of crop concentration.

The subscript *r* is used for village-level observations. The coefficients β, ω, μ, η, φ, , ϱ and ξ represent the vectors of parameter estimates for each associated vector of variables. $\varepsilon_{\mathrm{it}}$ is a household-specific composite error term defined as it follows:

$$\varepsilon_{it} = \alpha_i + u_{it} \quad (4)$$

That is, ε_{it} is composed of a normally distributed random error term, $u_{it} \sim N(0, \sigma_u^2)$, and an unobserved, household-specific component that is time-invariant (α_i).

The panel structure of our dataset allows the use of a fixed effect estimator that permits the time-variant regressors to be correlated with the time-invariant component of the error term, while assuming that these regressors are uncorrelated with the idiosyncratic error. This estimation provides consistent parameters even if there is correlation between the independents variables and time invariant unobserved heterogeneity, such as altitude. The estimation of an instrumental variables model with fixed effect methodology allows us to test and control, at least to some extent, for potential endogeneity caused by a correlation between decisions regarding diversification and crop income and vulnerability outcomes. Fixed effect models rely on data transformation that removes the individual effect. However, we deem important to include in our analysis variables that are in their nature time-invariant regressors, such as climate normals.

One way to include time-invariant variables while addressing endogeneity is to estimate a random effects model while controlling for unobserved heterogeneity using the Mundlak-Chamberlain approach (referred to as the pseudo-fixed effects model). Following Mundlak (1978) and Chamberlain (1982) the right-hand side of our regression equation includes the mean value of the time varying explanatory variables. This approach relies on the assumption that unobserved effects are linearly correlated with the explanatory variables. Thus, the unobserved household specific time invariant component in Equation (4) can be specified as

$$\alpha_i = \zeta \bar{x} + v_i , \tag{5}$$

where $\bar{x}$ is the mean of the time-varying explanatory variables within each farm household (cluster mean), ζ is the corresponding vector coefficient, and v_i is a random error unrelated to the $\bar{x}'$s. The vector ζ will be equal to zero if the observed explanatory variables are uncorrelated with the random effects. The use of the Mundlak-Chamberlain device also addresses the problem of selection and endogeneity bias where these are due to time-invariant unobserved factors, such as household heterogeneity (Wooldridge, 2002). If we failed to control for these factors, we would not obtain consistent parameter estimates.

We also conducted a series of robustness checks that are available in supplementary appendices. First, we re-estimate equation 6A for each index with panel data fixed effects models. These control for unobservable household time invariant characteristics, but treat the diversification indices as exogenous. Finally, we repeat the empirical procedure for each index, estimating 2SLS instrumental variables with fixed-effects panel data models. In these, crop diversification indices are treated as endogenous but we cannot specifically control and obtain the estimated coefficients for the climate normals and other time invariant regressors.

3.2 Explanatory variables

Vectors w_{rt} and c_r include respectively weather and climate information. Weather fluctuations and climatic shifts are two different meteorological events and they have distinct implications on farming decisions (Seo, 2013). The inclusion of both weather and climate allows capturing the full extent of underlying adaptation decisions (Bezabih, et al., 2014, Iizumi and Ramankutti, 2015, Bozzola, Smale and Di Falco, 2018). For example, weather fluctuations are perceived as random while long term climatic shifts are perceived as non-random by the farmers (Seo, 2013). Ortiz-Bobea 2020 shows that long-run effects of climate change that allow for the full range of farmer adaptations should be more optimistic than short-run estimates that only account for limited within-year farmer adaptations to weather fluctuations. Our climate and weather vectors comprise both temperature and rainfall information. Recent articles warn against the common practice in the development literature of specifying shocks to only include rainfall in empirical analyses. Since temperature and precipitation are closely correlated, excluding temperature may lead to attributing to precipitation shocks an impact that could be due to temperature (Auffhammer et al., 2013; Dell, Jones and Oken, 2014; Letta, Montalbano and Tol, 2018). We avoid this possible source of omitted variables bias by including both temperature and precipitation, for both weather and climate, in the regressions.

Vector s_r includes soil quality information such as the capacity of the soil to store water. Variables comprised in these vectors allow us to account for environmental conditions. These are important covariates to be included in the analysis, because smallholder agricultural production in rainfed agriculture, like that found in Kenya, relies on environmental production conditions that are "exogenously" determined - largely outside the control of farm families (Sherlund, Barrett and Adesina, 2002).

Vector x includes socio-economic factors influencing farm management choices and adaptation options, such as human capital (labor supply and quality) and financial and physical capital (e.g. assets, access to credit, ownership of means of transport, farm size and land tenure). The development and resilience of farms depend on factors such as the availability of skilled labor (labor supply and quality), education, gender, financial and physical capital (e.g. assets, ownership of means of transport, farm size and land tenure). These factors also shape adaptation decision making and risk management decisions (Crick et al., 2018). Each of the socio-economic variables has been selected based on a careful review of the literature. For example, we control for the farm-level total nominal value (KES) of livestock assets. Previous studies suggested that livestock in the economies of semi-arid Africa may also be an insurance substitute to cope with drought and other shocks (e.g. see Fafcamps, Udri and Czukas, 1998). Hence, livestock ownership and crop diversification strategies may potentially be both substitute or complement buffer strategies in the absence of formal insurance. Vector p_{rt} includes the hyperbolic sine transformation of population density at village level (an indicator of intensification) and the proportion of households in the village that received credit. We also include a dummy variable for women's headship. Analyses by Asfaw, Pallante and Palma (2018) in Niger and Covarrubias (2015) in Uganda found that woman-headed households tend to diversify their crop portfolio.

Of special interest is vector k_{rt}, capturing how climatic shocks affects crop diversification decisions (Equation 3). This vector includes a climate risk proxy stemming from the SPEI3 index, determined for the last month of the main rainfall season. In establishing the reference month, we take into account the historical time and length of the rainy season in each village. For each data collection year, we look at the number of times during the previous decade that the value of the SPEI3 was lower than -1.28 in the last month of the main rainfall season, which varies across the country. This value is the conventional SPEI threshold indicating exposure to drought stress (see e.g. Asfaw Pallante and Palma (2018)). The SPEI3 drought index expresses the incidence of past droughts (climatic shocks) as determinants of crop diversification choices.

Finally, $\mathrm{Z_a}$ comprises agro-regional zones fixed effects. These dummy variables can capture exogenous variables that vary by agro-regional zone but have not been measured explicitly. We include them as we believe that farming systems and farm management decisions are influenced by agro-regional zone. For example, the way farmers adapt to climate change might differ considerably depending if the farm is located in a zone with bimodal or unimodal rainfall regime.

3.3. Second stage regressions: crop choices, expected income and risk exposure

The role of variables D_{it}^{j}, representing different metrics of crop diversification, enters the second stage of our estimation strategy via the predictions from Equation (3). Through this second step, we investigate how crop diversification affect households' expected crop income and risk exposure. As discussed, our hypothesis is that crop diversification could have both a positive or negative impact on expected crop income. We further test the hypothesis that in a volatile environmental context, specialization could aggravate smallholder vulnerability while crop diversity, in term of both crop richness and crop evenness, may serve as a coping strategy.

The estimated relationship between crop income, risk variables, crop diversification decisions, weather and climatic variables, and other covariates is given by:

$$y_{it} = \alpha + \beta w_{rt} + \omega c_r + \mu x_{\mathrm{i,t}} + \eta p_{rt} + \varphi s_{\mathrm{r}} + \eta \widehat{D_{it}^{J}} + \xi \mathrm{Z_a} + \varepsilon_{it} \tag{6a}$$

$$\hat{\varepsilon}_{it}^{2} = \alpha + \beta w_{rt} + \omega c_r + \mu x_{\mathrm{i,t}} + \eta p_{rt} + \varphi s_{\mathrm{r}} + \eta \widehat{D_{it}^{J}} + \xi \mathrm{Z_a} + \varepsilon_{it} \tag{6b}$$

$$\hat{\varepsilon}_{it}^{2} = \alpha + \beta w_{rt} + \omega c_r + \mu x_{\mathrm{i,t}} + \eta p_{rt} + \varphi s_{\mathrm{r}} + \eta \widehat{D_{it}^{J}} + \xi \mathrm{Z_a} + \varepsilon_{it} \quad \text{if } y_{it} < E(y_{it}) \tag{6c}$$

$$\hat{\varepsilon}_{it}^{3} = \alpha + \beta w_{rt} + \omega c_r + \mu x_{\mathrm{i,t}} + \eta p_{rt} + \varphi s_{\mathrm{r}} + \eta \widehat{D_{it}^{J}} + \xi \mathrm{Z_a} + \varepsilon_{it} \tag{6d}$$

The dependent variable y_{it} denotes the hyperbolic sine transformation of crop income for the i^{th} farm household at year t. The subscripts i, t and r are defined as in Equations 2 and 3. Similarly, all vectors are defined as in Equation 3 with the exception of p_{rt} where all variables are the same except the proportion of households that received credit, which is excluded from the vector.

The coefficients in Greek letters represent, as in Equation 3, the vectors of parameter estimates for each associated vector of variables and ε_{it} is the composite error terms for each equation. It has the same distribution properties discussed for Equation 4, and the unobserved household specific time invariant component is also specified as presented in Equation 5, following the Mundlak-Chamberlain approach.

4. Data and data sources

We compiled the dataset from three comprehensive data sources.

The first source is household survey data collected by Egerton University's Tegemeo Institute of Agricultural Policy and Development in partnership with Michigan State University in four rounds (2000, 2004, 2007, 2010). Argwings-Kodhek (1999) provides a detailed description of the sample design, implemented in 1997 in consultation with the Kenya National Bureau of Statistics (KNBS).[3] All non-urban divisions in the selected districts belong to one or more agro-regional zones based on agronomic information from secondary data. The panel dataset comprises eight agro-regional zones. Within each division, villages and households (in that order) were randomly selected. The original sample excluded large farms with over 50 acres and two pastoral areas. The final dataset used in this study is a balanced panel dataset contains detailed farm-level data from 1,243 agricultural households in 22 districts. Certain village-level covariates, such as population density and agro-regional zones, are included in these data and our analysis.

We obtained climate and weather variables using the Climatic Research Unit (CRU) TS3.21 dataset (Harris et al., 2014). We constructed our covariates of interest from monthly average temperature and monthly cumulative precipitation for 107 villages across Kenya from 1971 to 2010. We took into account the exact timing of the main rainy season. We consider local differences in the length and timing of these two seasons. We used these data to calculate the SPEI Index. The SPEI Index is a multi-scalar drought index that accounts for the impact of rainfall on plant growth in the context of the soil's capacity to retain water. This in turn depends on the characteristics of the soil and on the extent to which sunshine induces evaporation (Harari and La Ferrara, 2018). The index considers the joint effects of precipitation, potential evapotranspiration (PET) and temperature in determining droughts (Vicente-Serrano et al., 2010). Extending the widely-used Standardized Precipitation Index (SPI) (McKee, Doesken and Kleist, 1993), the SPEI index can be used for determining the onset, duration and magnitude of drought

[3] The first survey was implemented in the same year (1997), and covered both the 1996/97 and 1995/96 cropping seasons. We use in this article the subsequent surveys conducted in 2000, 2004, 2007 and 2010. Survey instruments are publicly available online at http://www.tegemeo.org/index.php/resources/data/528-survey-instruments-and-data-documentation.html (last accessed: 30.05.2020).

conditions with respect to normal local conditions. Increasingly, the SPEI index is considered an improved measure over similar indexes previously used because it provides a better measure of the effective amount of moisture received by the soil (Vicente-Serrano et al., 2010; Harari and La Ferrara, 2018; Bozzola, Smale and Di Falco, 2018; Asfaw, Pallante and Palma, 2018). We establish village-specific reference months for the SPEI3. In our analysis the SPEI3 is the 3 months Standardized Precipitation-Evapotranspiration Index for the last month of the main rainfall season (January, July or August, depending on the division and agro-regional zone to which each village belongs) and the two preceding months.

Third, we draw on soils data at the village scale from the Harmonized World Soil Database, a partnership between the Food and Agriculture Organization (FAO) and the European Soil Bureau Network (FAO, IIASA, ISRIC, ISSCAS, JRC, 2012).

We present definitions for each variable in Table 1 and their descriptive statistics in Table 2.

INSERT TABLE 1 HERE

INSERT TABLE 2 HERE

5. Results

Crop richness and crop evenness are both positively associated with crop income while crop specialization has the opposite effect on crop income, once we control for several confounding factors.

INSERT TABLE 3 HERE

Table 3 presents interesting associations between crop income and, respectively, crop richness (column 1), crop evenness (column 2) and crop concentration (column 3). However, those associations don't imply that crop abundance, evenness and concentration actually lead to lower or higher expected income unless the indices are exogenous. As we expected, the results of the Davidson-Mackinnon tests reject in all the cases the null hypothesis of exogeneity, confirming that the three indices should be treated as endogenous in the empirical estimations (Table 4).

INSERT TABLE 4 HERE

In order to attempt to make a causal statement about the impact of different crop diversification strategies on crop income and its higher moments, we turn to our 2SLS results.

First-stage regression results for each crop diversification index are reported in Table 5.

INSERT TABLE 5 HERE

We address the issue of under-identification and relevance of instrumental variables using the Angrist-Pischke statistics, which are reported at the bottom of the table. The F statistic is greater than 13 in all three first stage regressions, supporting the strength of the proposed instruments taken as a set (Staiger and Stock, 1997; Stock and Yogo, 2005). The choice of instruments seems appropriate and we turn to discussing our main regression results.

With respect to our primary hypotheses, as expected, persistent climatic shocks, captured by the frequency of droughts in the main growing season in the decade before each data collection round, appear to be an important driver of on-farm biodiversity, while these hinder crop specialization. Notably, the frequency of past droughts has a positive correlation with the richness and evenness in the field but a negative association with the index of crop specialization. This result means that frequent past droughts during the growing season, which are the main causes of crop failure in Kenya, make farmers more prone to increase crop richness and evenness.

The finding is consistent with Di Falco and Chavas (2008), who report that maintaining crop biodiversity may be desirable to farmers because it tends to provide the agroecosystem a wider range of productive responses to weather shocks. Weather and climate affect crop diversification decisions in a nonlinear way.

Credit is also an important determinant of crop diversification choices. Higher credit availability at village level negative correlates with crop richness and evenness but has a positive effect on crop specialization.

A larger land endowment is positively correlated with crop richness and crop evenness and negative with crop specialization. Again, our sample does not include large, specialized farms producing cash crops. The most frequently grown crop is maize, harvested dry, followed by maize harvested green, beans, bananas and *sukuma wiki* (kale). Beans and sukuma wiki are served with *ugali*, the staple maize dish.

The coefficients associated to livestock are positive and statistically significant in both the crop richness and crop evenness regressions, and it is not statistically significant in the crop concentration regression. These results suggest that even if farmers may rely on both livestock ownership and crop diversification strategies as buffer strategies to insulate their consumption from fluctuations in income in the absence of formal insurance, the two strategies may not be substitute to each other.

Population density has a negative association with crop richness and crop evenness. This can be due to the fact that farmers living in villages more densely populated have access to a broader range of crops for their consumption and food security at household level is less dependent by the capability of the single farm-family to diversify their food-intake. Economic principles suggest that the commercialization process that favors agricultural development spurs specialization among crop enterprises.

Tables 6, 7 and 8 report the results for the second stage regressions for the indexes of crop richness, crop evenness and crop specialization respectively.

INSERT TABLE 6 HERE

INSERT TABLE 7 HERE

INSERT TABLE 8 HERE

As anticipated, the impact on expected income and risk of crop richness and evenness follow similar patterns, while crop concentration has opposite income and risk effects.

In each table, column (1) reports the effect of crop diversification choices on crop income. Crop richness is positively and significantly correlated with expected crop income (Table 6). The sign of the coefficient of crop evenness is also positive and statistically significant (Table 7). Crop specialization, however, is negatively correlated with crop income, perhaps because of higher production costs (Table 8). This is also consistent with the fact that the crop inventory of our dataset reveals that when farmers specialize, they are allocating more of their land to maize and other food crops, which are produced mainly for home consumption. In this case, specialization is a subsistence rather than a marketing strategy.

Columns 2, 3 and 4 in Tables 6 to 8 show the effect of the three crop diversity indexes on the risk metrics. Results support the hypothesis that farmers rely on crop richness and evenness as strategies to cope with production risk. We find robust evidence that both crop richness and to a larger extent crop evenness decrease risk in terms of variance (column 2) and negative semi-variance (column 3) of the distribution of crop income, while crop concentration has an opposite effect on both dependent variables. This finding is consistent with Jones et al. (2012) and Bozzola, Smale and Di Falco (2018). The magnitudes of the impact of crop evenness (Table 7) on the variance is larger if restrict the sample to those crop income realizations below the expected value (i.e. when we use the lower partial moment as dependent variable). The proportional abundance of area planted among crops can therefore reduce vulnerability to crop income instability in more difficult years. A higher number of crops and more even area distribution among them on farms has the effect of reducing downside risk. By contrast, crop specialization increases risk, particularly so for those farmers more vulnerable to adverse weather conditions.

These results are coherent with those presented in column 4 of each table. Crop richness and evenness also decrease risk in term of the skewness of the distribution of crop income (decrease the risk of negative net incomes, with a positive sign) but crop specialization increases downside risk (increase the risk of negative net incomes, with a negative sign). These conclusions support the discussion in Binswanger and von Braun (1991) who suggest that in an environment characterized by risky markets for outputs and factors, and absent insurance markets, small farmers insure against risks to food security by prioritizing some self-sufficiency in food production (for example focusing on maize and beans production) rather than joining a commercialization process. Known as "safety-first" behavior in the literature about decision-making under risk, this, in turn, may lead to a failure to realize the maximum short-term gains from specialization.

6. Conclusions

This paper assesses the role of inter-specific crop diversity, expressed in three different metrics, in expected crop income and risk management on smallholder farms in Kenya.

First, we find that crop diversification decisions are strongly affected by past climate shocks, weather, and climatology, in addition to commonly cited, household-farm characteristics such as access to credit and off-farm earnings. Next, we find that a larger number of crops grown per farm (crop richness) has a positive effect on expected crop income. Third, both crop richness and greater proportional abundance of area allocated among crops (crop evenness) mitigate risk. Further, the risk reduction benefits of crop evenness are greater for those farm households with expected crop income below the reference value. Our results show that higher crop richness and especially crop evenness on farms significantly decrease the instability of net crop income, particularly among lower incomes. In other words, these are downside risk-decreasing strategies. while crop specialization increases risk.

On the other hand, crop specialization increases income risk among the smallholder farmers in our sample. Inspection of the data revealed that most of the specialization among these Kenyan smallholders involves allocating disproportionately large shares of farm area to maize, the food staple. While maize is a commercial crop in Kenya, we consider that most of this specialization still reflects “safety-first” behavior intended to ensure some food self-sufficiency in the face of risk. This is especially true if market prices rise later in the season when households have depleted their supplies and market purchases incur substantial transactions costs.

In this context, the benefits of crop diversification are twofold. Crop diversity (especially in terms of crop evenness) contributes to greater expected incomes among smallholder farmers in Kenya, but it also reduces risk when few alternative strategies for risk-reduction are available.

Higher, less risky incomes could result either from the benefits of adding other crops to a portfolio too often based on staple foods such as maize, or from a reduction of production costs related to pest or disease management, or soil fertility effects.

Testing which of these effects accounts for more of our findings is one future avenue of research. Another interesting avenue would be to distinguish between crop diversification decisions looking at the share of cash crops versus food crops included in the agricultural production. For example, we would expect farmers who have access to credit for cash crops to be more specialized in one farm enterprise or another.

Findings have significant policy implications where rural areas are characterized by market failures including credit constraints, information asymmetries, and commitment failures. These can cause weak insurance and risk-coping mechanisms (Fafchamps, 1992; Kurosaki and Fafchamps, 2002). Safety nets typically provide only limited support (Dercon and Krishnan, 2000; Dercon, 2004) while off-farm, non-covariant income is also limited in more remote rural areas. In Kenya, few options exist to diversify income or activities, especially for short term adjustments (Mathenge and Tschirley, 2015). For these reasons, farmers have incentives to diversify their crop choices as a strategy to buffer against risk. This conclusion calls for well-designed extension services to provide information on how to grow various crop combinations with different response to climate and offsetting probability distributions of costs and benefits. Possibly, the need to rely on crop diversification strategies would be lower if insurance schemes, at present mostly inexistent in rural Kenya, would be offered. In a context such as that of rural Kenya, characterized by poor crop insurance coverage, crop richness and evenness constitutes a possible substitute for financial insurance in hedging against the impact of risk exposure on welfare.

List of tables

Table 1: Variables Definitions

Variable	Description
Farm specific variables (Source: Tegemeo)	
Richness index	Count of crops planted by the farmer in the main rainfall season.
Evenness index	Shannon crop index, all seasons. Definition: $-\sum(a_{ji}/a_i) * ln(a_{ji}/a_i)$ where a_{ji} is the area of plot i planted with crop j and a_i is the total area of the plot *i*.
Specialization index	Herfindahl-Hirschman crop index, main seasons. Definition: $\sum(a_{ji}/a_i)^2$ where a_{ji} is the area of plot *i* planted with crop j and a_i is the total area of the plot *i*.
Crop income (KES)	Value of crop production (KES) minus input and land preparation costs (labor and seeds costs excluded). Input costs as reported by farmers; product price as reported for largest sale; district median used for missing data.
Distance fertilizer seller	Distance from household to the nearest fertilizer seller (km).
Transport dummy	Household holds a mean of transport (bicycle, motorcycle, car and/or track)
Livestock assets (KES)	Total nominal value (KES) of livestock assets.
Salaries & remittances	Share of salaries and remittance earnings in total household income.
Land (ac)	total household land area (ac).
Land title deed	=1 if land owned with no title deed, 0 otherwise.
Educated adults	No of adult women and men with secondary education.
Women's headship dummy	1=reported head of household female, 0 otherwise.
Mortality dummy	1=household experienced prime-age mortality since previous survey, 0 otherwise.
Village-specific climate characteristics (Source: CRU TS3.21)	
Temperature med (°C)*	Monthly average mean air temperature (°C) during the major rainfall season.
Rainfall (mm/mo)*	Cumulated rainfall (mm/mo) during the major rainfall season.
Temperature average climatologies*	Average air temperature (°C) 1971-2010 during the major rainfall season.
Rainfall climatologies (mm/mo)*	Cumulated rainfall (mm/mo) 1971-2010 during the major rainfall season.
Past droughts	Number of times in the last decade# the value of the SPEI3 was <-1.28 in the last month of the main rainfall season. We

	calculated the SPEI index manually, using the R routines developed by Vicente Serrano et al. (2010). To perform the calculation, we prepared a dataset of monthly precipitation and rainfall data at village level from the CRU TS3.21 dataset (Harris et al., 2014) for the period 1971-2012.
<u>Village-specific soil characteristics (Source: World Soil Database)</u>	
Ph top soil (-log(H+))	pH measured in a soil-water solution. It is a measure for the acidity/alkalinity of the soil.
Gravel top soil (%vol)	Volume % gravel (materials in a soil larger than 2mm) in the topsoil (i.e. 0-30 cm) (%vol).
AWC class (mm)	Available water storage capacity class of the soil unit, measured in mm/m.
<u>Village-specific socio economic variables & agro-regional zones</u>	
Population density	Village population density (cap/km²).
Credit village	Proportion of households in the village that received credit.
Agro-regional zone (% of farms)	HPMZ high potential maize zone (26.6); CHI central highlands (19.4), WLO western lowlands (12); WTR western transitional (11.7); ELO eastern lowlands (11.3%); WHI western highlands (10.3); CLO coastal lowlands (5.9%); MRS marginal rain shadow (2.7). Percentages indicate the frequency of farms in each agro-regional zone.

*We take into account the relevant cropping season: e.g. for villages in the Rift Valley, the reference period is March to August. For villages located in the Easter Lowlands the reference period is October to January. # Reference decades: 1989-1999 for 2000; 1993-2003 for 2004; 1996-2006 for 2007; 1999-2009 for 2010.

Table 2: Descriptive Statistics

Variables	Mean	Std. Dev.	Min	Max
Farm-specific variables				
Crop count richness index	5.420	2.778	1	23
Crop evenness Shannon index	1.097	0.500	0	2.725
Crop specialization Herfindhal index	0.444	0 .217	0.083	1
Crop income (KES)	87,911	142,264	0	3,883,123
Distance fertilizer seller (km)	4.267	6.422	0	78
Transport dummy	0.478	0.500	0	1
Livestock assets (KES)	81,367	217,682	0	8,679,900
Salaries & remittances	0.18	0.24	0	1
Land (ac)	5.80	8.72	0	157
Land title deed	0.36	0.48	0	1
Educated adults	1.61	1.43	0	13
Women's headship dummy	0.21	0.40	0	1
Mortality dummy	0.06	0.23	0	1
Village-specific climatic variables and soil characteristics				
Temperature med (°C)	20.27	3.62	13.3	28.68
Rainfall (mm/mo)	708.75	209.29	145	1,154
Temperature average climatologies (°C)	19.57	3.69	13.61	27.89
Rainfall climatologies (mm/mo)	708.95	186.32	184.58	946.44
Past droughts	1.59	1.14	0	5
Bimodal rainfall dummy	0.79	0.41	0	1
Ph top soil (-log(H+))	5.75	1.04	4.5	8.9
Gravel top soil (%vol)	1.25	4.09	0	21
AWC class (mm)	149.42	3.774	125	150
Village-specific socio economic variables				
Population density	363.48	214.88	16.43	1,245
Credit village	0.471	0.302	0	1

Notes: Number of farm households: 1,243 (balanced panel dataset).

Table 3: OLS Estimation results for crop diversification indices

	Dependent variable: hyperbolic sine transformations of crop income (y_{it})		
	[1]	**[2]**	**[3]**
Count index	0.0990*** [0.0058]		
Shannon index		0.5622*** [0.0343]	
Herfindahl-Hirschman index			-1.2117*** [0.0849]
Observations	4,960	4,960	4,960
Number of farm households	1,243	1,243	1,243
R-squared	0.453	0.454	0.450

Notes: OLS estimations. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. The complete results of these three OLS regressions are reported in Table A-1 in the supplementary appendices.

Table 4: Davidson-MacKinnon Test for endogeneity of crop diversification indices

	[1]	**[2]**	**[3]**
Count index	23.979*** [0.000]		
Shannon index		26.102*** [0.000]	
Herfindahl-Hirschman index			32.113*** [0.000]

Notes: Null hypothesis of exogeneity. Chi² (p-value) in brackets. *** p<0.01.

Table 5: Pseudo-Fixed effect IV estimations results – First Stage regressions

	[1] Richness	[2] Evenness	[3] Specialization
Past droughts	0.2495***	0.0558***	-0.0232***
	[0.0455]	[0.0079]	[0.0035]
Credit	-1.2969***	-0.2170***	0.0763***
	[0.2006]	[0.0357]	[0.0163]
Livestock assets	0.0701***	0.0084**	-0.0029
	[0.0191]	[0.0038]	[0.0018]
Salaries & remittances	-0.7709***	-0.1915***	0.0779***
	[0.2010]	[0.0365]	[0.0167]
Land	0.9252***	0.1317***	-0.0455***
	[0.0799]	[0.0140]	[0.0064]
Educated adults	0.0133	0.0018	-0.0001
	[0.0667]	[0.0113]	[0.0051]
Women's headship dummy	0.0738	0.0009	0.0045
	[0.1658]	[0.0270]	[0.0122]
Transport dummy	0.2032*	0.0367*	-0.0186**
	[0.1091]	[0.0193]	[0.0088]
Land title deed	0.1065	0.0348**	-0.0164**
	[0.0863]	[0.0150]	[0.0067]
Mortality dummy	-0.1067	-0.0388	0.0253*
	[0.1635]	[0.0279]	[0.0130]
Distance fertilizer seller	-0.0792	-0.0041	0.0012
	[0.0585]	[0.0098]	[0.0044]
Average temp.	-2.2688***	-0.1833**	0.0555
	[0.5398]	[0.0836]	[0.0342]
Average temp. squared (sq.)	0.0439***	0.0049**	-0.0019**
	[0.0124]	[0.0020]	[0.0008]
Rainfall	-0.0052***	-0.0006**	0.0002*
	[0.0016]	[0.0003]	[0.0001]
Rainfall sq.	0.000005***	0.0000005***	-0.0000002**
	[0.000001]	[0.0000]	[0.0000]
Average temp. climatologies	58.2636	10.4304**	-3.0117*
	[48.1049]	[4.7759]	[1.5365]
Average temp. climatologies sq.	-1.0942	-0.2045*	0.0614*
	[1.0534]	[0.1065]	[0.0346]
Rainfall climatologies	-0.1198***	-0.0235***	0.0090***
	[0.0150]	[0.0031]	[0.0014]
Rainfall climatologies sq.	0.0001***	0.00002***	-0.000006***
	[0.0000]	[0.0000]	[0.0000]
AWC class (mm)	0.0017	-0.0000	-0.0000

	[0.0121]	[0.0033]	[0.0016]
Ph top soil	2.7337	0.2147	0.0101
	[1.7488]	[0.2817]	[0.1224]
Ph top soil squared	-0.2681*	-0.0197	-0.0014
	[0.1572]	[0.0256]	[0.0112]
Gravel top soil	-0.0082	0.0014	-0.0003
	[0.0155]	[0.0033]	[0.0015]
Pop density	-2.3587***	-0.1412*	0.0476
	[0.4574]	[0.0741]	[0.0329]
Constant	57.3841	10.9236**	-2.6095
	[50.2235]	[5.0166]	[1.6341]
Agro-regional FE	Yes	Yes	Yes
Observations	4,960	4,960	4,960
Number of households	1,243	1,243	1,243
Angrist-Pischke first-stage F statistics of weak identification	F(4, 4912) =26.36	F(4, 4912) =33.59	F(4, 4912) =27.80
Angrist-Pischke first-stage Chi squared statistics of underidentification	AP Chi-sq (4) =106.47	AP Chi-sq (4) =135.69	AP Chi-sq (4) =112.28
R-squared	0.795	0.792	0.760

Notes: Pseudo-Fixed effect estimation. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Estimation results – Crop Richness Index: second stage regressions of Mean, Variance, Semi-Variance and Skewness of Crop Income

	[1] Crop income	[2] Variance	[3] Negative semi-variance	[4] Skewness
Count index	0.3096*** [0.0471]	-0.2333** [0.1060]	-0.1880 [0.1512]	1.6056** [0.7075]
Observations	4,960	4,960	2,352	4,960
Number of farm households	1,243	1,243	1,106	1,243

Notes: Pseudo-Fixed effect estimation. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. Full results of the second stage regressions are reported in Table A-2 in the supplementary appendices.

Table 7: Estimation results – Crop Evenness Index: second stage regressions of Mean, Variance, Semi-Variance and Skewness of Crop Income

	[1] Crop income	[2] Variance	[3] Negative semi-variance	[4] Skewness
Shannon index	2.0084*** [0.2387]	-0.9823* [0.5298]	-1.7670** [0.7622]	7.7899* [4.2511]
Observations	4,960	4,960	2,469	4,960
Number of farm households	1,243	1,243	1,110	1,243

Notes: Pseudo-Fixed effect estimation. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. Full results of the second stage regressions are reported in Table A-3 in the supplementary appendices.

Table 8: Estimation results – Crop Specialization Index: second stage regressions of Mean, Variance, Semi-Variance and Skewness of Crop Income

	[1] Crop Income	[2] Variance	[3] Negative semi-variance	[4] Skewness
Herfindahl-Hirschman index	-4.7716*** [0.6295]	2.6673** [1.3091]	5.6421** [2.5676]	-18.5475** [8.8782]
Observations	4,960	4,960	2,593	4,960
Number of farm households	1,243	1,243	1,120	1,243

Notes: Pseudo-Fixed effect estimation. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. Full results of the second stage regressions are reported in Table A-4 in the supplementary appendices.

References

Angrist, J. A. & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives, 15*, 69-85.

Antle, J. M. (1983). Testing the stochastic structure of production: A flexible moment-based approach. *Journal of Business and Economic Statistics, 1*, 192-201.

Antle, J. M. (2010). Asymmetry, partial moments and production risk. *American Journal of Agricultural Economics*, *92*(5), 1294-1309.

Argwings-Kodhek, G, Jayne, T.S., Nyambane, G., Awuor, T., & Yamano T. (1999). How Can Micro-level Household Information Make a Difference for Agricultural Policy Making? Selected Examples from the KAMPAP Survey of Smallholder Agriculture and Non-Farm Activities for Selected Districts in Kenya. Nairobi, Kenya: Tegemeo Institute, Egerton University.

Asfaw, S., Pallante G. & Palma A. (2018). Diversification Strategies and Adaptation Deficit: Evidence from Rural Communities in Niger. *World Development 101,* 19-234.

Auffhammer M., Hsiang, S.M., Schlenker, W. & Sobel A. (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy*, p. ret016.

Bezabih, M., Di Falco, S., & Mekonnen, A. (2014). (Rep.). Environment for Development Initiative.

Billor, N., Hadi, A. S. & Velleman, P.F. (2000). BACON: Blocked adaptive computationally efficient outlier nominators. *Computational Statistics & Data Analysis, 34*, 279-298.

Binswanger H. P. & von Braun J. (1991). Technological Change and Commercialization in Agriculture: The Effect on the Poor. *The World Bank Research Observer*, *6*(1), 57-80

Bozzola, M., Smale, M. & Di Falco, S. (2018). Maize Intensification among Smallholder Farmers in Kenya: Understanding the Impacts of Climate. Book Chapter in: Agricultural Adaptation to Climate Change in Africa: Food Security in a Changing Environment. RFF Press/Routledge, London and New York.

Carter, M., de Janvry, A., Sadoulet, E. & Sarris A. (2017). Index Insurance for Developing Country Agriculture: A Reassessment. *Annual Review of Resource Economics, 9*, 421-38

Chamberlain, G. (1982). Multivariate Regression Models for Panel Data. *Journal of Econometrics 1*, 5-46.

Chavas J-P & Shi G. (2015). An Economic Analysis of Risk, Management, and Agricultural Technology. *Journal of Agricultural and Resource Economics, 40*(1), 63–79.

Covarrubias, K.A. (2015). The role of crop diversity in household production and food security in Uganda: A gender-differentiated analysis. FOODSECURE WP 32, LEI Wageningen UR.

Crick, F., Eskander, S., Fankhauser, S., & Diop, M. (2018) How do African SMEs respond to climate risks? Evidence from Kenya and Senegal. *World Development, 108*, 157-168.

Dell, M., Jones, B.F. & Olken B.A. (2014). What do we learn from the weather? The new climate–economy literature. *Journal of Economic Literature*, *52* (3), 740-798.

Dercon, S. (2004). Insurance Against Poverty (Oxford: Oxford University Press, 2004).

Dercon, S. & Krishnan, P., (2000). Vulnerability, poverty and seasonality in Ethiopia, *Journal of Development Studies* 36, 25–53.

Di Falco, S. (2012). On the Value of Agricultural Biodiversity. *Annual Review of Resource Economics*, *4*(1), 207-223.

Di Falco, S. & Bezabih, M. & Yesuf, M. (2010). Seeds for livelihood: crop biodiversity and food production in Ethiopia. *Ecological Economics*, *69*(8), 1695-1702.

Di Falco S., & Chavas J. P. (2006). Crop genetic diversity, farm productivity and the management of environmental risk in rainfed agriculture. *European Review Agricultural Economics*, *33*(3), 289-314

Di Falco, S. & Chavas, J. P. (2008). Rainfall Shocks, Resilience, and the Effects of Crop Biodiversity on Agroecosystem Productivity. *Land Economics 84*(1), 83-96

Di Falco, S. & Chavas, J. P. (2009). On crop biodiversity, risk exposure and food security in the highlands of Ethiopia. *American Journal of Agricultural Economics 91*, 599-611.

Di Falco S., & Perrings C. (2005). Crop biodiversity, risk management and the implications of agricultural assistance. *Ecological Economics*, *55*(4), 459-466

Di Falco, S., & Veronesi, M. (2014). Managing Environmental Risk in Presence of Climate Change: The Role of Adaptation in the Nile Basin of Ethiopia. *Environmental and Resource Economics, 57*, 553-577.

DuVal, A., Mijatovic, D. & Hodgkin, T. (2019). The contribution of biodiversity for food and agriculture to the resilience of production systems - Thematic Study for The State of the World's Biodiversity for Food and Agriculture. FAO, Rome. 85 pp. Licence: CC BY-NC-SA 3.0 IGO.

Eriksen, S.H., Brown, K. & Kelly, P.M. (2005). The dynamics of vulnerability: locating coping strategies in Kenya and Tanzania. *Geographical Journal*, 171: 287-305.

Estrada, J. (2004). Mean-semi variance behaviour: an alternative behavioural model. *Journal of Emerging Market Finance*, *3*(3), 231-248

Fafchamps, M. (1992). Cash crop production, food price volatility and rural market integration in the third world. *American Journal of Agricultural Economics*, *74*, 90–99.

Fafchamps M., Udry C., & Czukas K. (1998). Drought and saving in West Africa: are livestock a buffer stock? *Journal of Development Economics*, *55*(2), 273-305

FAO/IIASA/ISRIC/ISS-CAS/JRC. (2012). Harmonized World Soil Database (version 1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Finger, R., Dalhaus, T., Allendorf, J., & Hirsch, S. (2018). Determinants of downside risk exposure of dairy farms. *European Review of Agricultural Economics*, *45*(4), 641–674.

Harari, M., & La Ferrara, E. (2018). Conflict, Climate and Cells: A disaggregated analysis. *The Review of Economics and Statistics*, *100*(4), 594-608.

Harris, I., Jones, P.D., Osborn, T.J., & Lister, D.H. (2014). Updated high-resolution grids of monthly climatic observations. *International Journal of Climatology*, *34*, 623-642.

Iizumi, T., & Ramankutty, N. (2015). How do weather and climate influence cropping area and intensity? *Global Food Security, 4*, 46-50, Jaramillo, J., Muchugu, E., Vega, F.E., Davis, A., Borgemeister, C., & Chabi-Olaye, A. (2011). Some like it hot: the influence and implications of climate change on coffee berry borer (Hypothenemus hampei) and coffee production in East Africa. *PLoS ONE*, *6*(9), e24528.

Jensen, N. & Barrett C. (2017). Agricultural Index Insurance for Development. *Applied Economic Perspectives and Policy*, *39*(2), 199-219

Jones, A., Dalton T., & Smale M. (2012). A Stochastic production function analysis of maize hybrids and yield variability in drought-prone areas of Kenya. Tegemeo Working Paper 49. Tegemeo Institute of Agricultural Policy and Development, Nairobi.

Klasen, S., Meyer, K. M., Dislich, C., Euler, M., Faust, H., Gatto, M., Hettig, E., Melati, D. N., Jaya, I.N.S., Otten, F., Pérez-Cruzado C., Steinebach, S., Tarigan S., & Wiegand, K. (2016) Economic and ecological trade-offs of agricultural specialization at different spatial scales. *Ecological Economics*, 122, 111-120,

Kurosaki, T. & Fafchamps, M., (2002). Insurance market efficiency and crop choices in Pakistan, *Journal of Development Economics*, *67*, 419-453.

Letta, M., Montalbano, P. & Tol, R.S.J., (2018). Temperature shocks, short-term growth and poverty thresholds: Evidence from rural Tanzania. *World Development*, *112*(C), 13-32.

Mathenge, M. K., Smale, M., & Olwande, J., (2014). The impacts of hybrid maize seed on the welfare of farming households in Kenya. *Food Policy*, *44*, 262-271.

Mathenge, M. K. & Tschirley, D. L. (2015). Off-farm Labor Market Decisions and Agricultural Shocks among Rural Households in Kenya. *Agricultural Economics 46*, 1–14.

McCord P.F., Cox, M., Schmitt-Harsh, M. & Evans, T., (2015). Crop diversification as a smallholder livelihood strategy within semi-arid agricultural systems near Mount Kenya. *Land Use Policy 42*, 738-750

McKee, T.B.N., Doesken J., & Kleist J., (1993). The relationship of drought frequency and duration to time scales. Eight Conf. On Applied Climatology. Anaheim, CA, Amer. Meteor. Soc. 179-184.

Menezes C., Geiss C., & Tressler J., (1980). Increasing downside risk. *American Economic Review 70*, 921-932

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica 46*, 69-85.

Ochieng, J., Kirimi, L. & Mathenge, M., (2016). Effects of climate variability and change on agricultural production: The case of small scale farmers in Kenya. *NJAS - Wageningen Journal of Life Sciences*, *77*, 71-78.

Ortiz-Bobea, A. (2020), The Role of Nonfarm Influences in Ricardian Estimates of Climate Change Impacts on US Agriculture. *American Journal of Agricultural Economics, 102:* 934-959.

Paxson, C., (1992). Using Weather Variability to Estimate the Response of Savings to Transitory Income in Thailand. *American Economic Review*, *82*(1), 15-33.

Sherlund, S. M., Barrett, C. B., & Adesina, A.A., (2002). Smallholder technical efficiency controlling for environmental production conditions. *Journal of Development Economics*, *69*, 85-101.

Schoengold, K., Ding, Y. & Headlee, R. (2015). The Impact of AD HOC Disaster and Crop Insurance Programs on the Use of Risk-reducing Conservation Tillage Practices. *American Journal of Agricultural Economic*s, *97*(3), 897-919.

Seo, S. N. (2013). An essay on the impact of climate change on US agriculture: weather fluctuations, climatic shifts, and adaptation strategies *Climatic Change,* 121:115-124

Staiger, D. & Stock J. H., (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65, 557-586.

Stock J., & Yogo M., (2005). Testing for Weak Instruments in Linear IV Regression. In: Andrews DWK Identification and Inference for Econometric Models. New York: Cambridge University Press; 2005. pp. 80-108.

Smale M., Hartell, J., Heisey, P. W., & Senauer, B., (1998). The Contribution of Genetic Resources and Diversity to Wheat Production in the Punjab of Pakistan. *American Journal Agricultural Economics 80*(3) 482-493.

Tilman, D., Polasky, S., & Lehman, C., (2005). Diversity, productivity and temporal stability in the economies of humans and nature. *Journal of Environmental Economics and Management*, 49(3), 405-426,

Vicente-Serrano S.M., Beguerìa S., López-Moreno J.I., Angulo M., & El Kenawy A., (2010). A global 0.5 gridded dataset (1901-2006) of a multiscalar drought index considering the joint effects of precipitation and temperature. *Journal of Hydrometeorology*, *11*(4), 1033-1043.

Weber, S. (2010). Bacon: An effective way to detect outliers in multivariate data using Stata (and Mata). *The Stata Journal 10*(3): 331-338.

Widawsky, D., & Rozelle, S., (1998). Varietal Diversity and Yield Variability in Chinese Rice Production, in M. Smale ed., Farmers, GeneBanks, and Crop Breeding, Boston: Kluwer.

Wineman A, Mason N.M., Ochieng J, & Kirimi L. (2017). Weather extremes and household welfare in rural Kenya. *Food Security 9*(2), 281-300.

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data (Cambridge, MA: MIT Press, 2002).

Acknowledgements

The farm household data used in this work was collected and made available by the Tegemeo Institute of Agricultural Policy and Development of Egerton University, Kenya. However, the specific findings and recommendations remain solely the authors' and do not necessarily reflect those of Tegemeo Institute.

We are grateful to Salvatore Di Falco for many useful discussions and thoughtful comments. We would also like to thank Luigi Biagini, Jean-Paul Chavas, Robert Finger, Nick Hanley, Mary K. Mathenge and Simone Severgnini for their comments to earlier versions of this article, Simone Fatichi and Xavier Vollenweider for assistance with the climate and soil quality data and Mark Schaffer, for sharing the user written stata command.The authors only are responsible for any omissions or deficiencies.

Appendix I

Table A-1: OLS Estimation for crop diversification indices (complete results)

	Dependent variable: hyperbolic sine transformations of crop income		
	[1]	[2]	[3]
Count index	0.0990***		
	[0.0058]		
Shannon index		0.5622***	
		[0.0343]	
Herfindahl-Hirschman index			-1.2117***
			[0.0849]
Livestock assets	0.0491***	0.0489***	0.0496***
	[0.0067]	[0.0067]	[0.0067]
Salaries & remittances	-0.7264***	-0.7047***	-0.7163***
	[0.0818]	[0.0812]	[0.0815]
Land	0.4496***	0.4748***	0.4884***
	[0.0235]	[0.0230]	[0.0228]
Educated adults	0.1361***	0.1472***	0.1512***
	[0.0223]	[0.0225]	[0.0226]
Women's headship dummy	-0.1575***	-0.1409***	-0.1475***
	[0.0368]	[0.0365]	[0.0365]
Transport dummy	0.1941***	0.1964***	0.1964***
	[0.0299]	[0.0299]	[0.0300]
Land title dee	-0.0679**	-0.0803***	-0.0823***
	[0.0278]	[0.0278]	[0.0278]
Mortality dummy	-0.1081*	-0.1006*	-0.0987*
	[0.0565]	[0.0558]	[0.0561]
Distance fertilizer seller	-0.0666***	-0.0686***	-0.0684***
	[0.0194]	[0.0194]	[0.0195]
Average temperature	1.3852***	1.2323***	1.1962***
	[0.1676]	[0.1700]	[0.1711]
Average temp. squared	-0.0260***	-0.0243***	-0.0240***
	[0.0041]	[0.0041]	[0.0042]
Rainfall	0.0071***	0.0068***	0.0067***
	[0.0006]	[0.0006]	[0.0006]
Rainfall squared	-0.000005***	-0.000005***	-0.000004***
	[0.0000]	[0.0000]	[0.0000]
Average temp. climatologies	-1.1752***	-1.0660***	-1.0083***
	[0.1751]	[0.1779]	[0.1791]
Average temp. climatologies squared	0.0222***	0.0215***	0.0205***
	[0.0044]	[0.0045]	[0.0046]
Rainfall climatologies	-0.0053***	-0.0059***	-0.0060***
	[0.0016]	[0.0016]	[0.0017]
Rainfall climatologies squared	0.000005***	0.000005***	0.000005***

	[0.0000]	[0.0000]	[0.0000]
AWC class (mm)	0.0186**	0.0191**	0.0194**
	[0.0093]	[0.0093]	[0.0093]
Ph top soil	-1.3844***	-1.1524***	-1.1715***
	[0.1972]	[0.2020]	[0.2057]
Ph top soil squared	0.1041***	0.0829***	0.0833***
	[0.0176]	[0.0181]	[0.0184]
Gravel top soil	-0.0306***	-0.0302***	-0.0298***
	[0.0042]	[0.0042]	[0.0042]
Population density	0.1702***	0.1727***	0.1900***
	[0.0413]	[0.0420]	[0.0422]
Agro-regional FE	Yes	Yes	Yes
Constant	6.3367***	6.2650***	7.2481***
	[1.6517]	[1.6542]	[1.6650]
Observations	4,960	4,960	4,960
R-squared	0.453	0.454	0.450

Notes: Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A-2: Count index Crop Richness Second Stage Regressions Pseudo-Fixed effect IV estimations (complete results)

	[1] Crop income	**[2] Variance**	**[3] Negative semi-variance**	**[4] Skewness**
Count index	0.3096***	-0.2333**	-0.1880	1.6056**
	[0.0471]	[0.1060]	[0.1512]	[0.7075]
Livestock assets	0.0281***	-0.0039	-0.0203	0.0026
	[0.0095]	[0.0146]	[0.0342]	[0.0895]
Salaries & remittances	-0.6353***	0.4387	1.5511*	-4.5271
	[0.1148]	[0.3741]	[0.7957]	[3.0165]
Land	0.0423	0.3321**	0.2684	-1.8027**
	[0.0589]	[0.1318]	[0.2154]	[0.9036]
Educated adults	0.0598**	-0.0144	-0.0494	-0.0057
	[0.0297]	[0.0741]	[0.1534]	[0.5724]
Women's headship dummy	0.0093	-0.1573	-0.3080	1.9898
	[0.0793]	[0.2794]	[0.5313]	[2.4777]
Transport dummy	0.0419	-0.0133	0.0672	0.2425
	[0.0489]	[0.1223]	[0.2009]	[1.0112]
Land title deed	-0.0613*	0.0251	0.1332	-0.4096
	[0.0360]	[0.0646]	[0.1310]	[0.4345]
Mortality dummy	0.0082	-0.1799	-0.4883	1.6235
	[0.0683]	[0.1517]	[0.3199]	[1.0712]
Distance fertilizer seller	-0.0552**	0.0569	0.0313	-0.0993
	[0.0245]	[0.0559]	[0.1082]	[0.4400]
Average temperature	1.5324***	-0.5729	-0.4514	5.2866*

	[0.2220]	[0.4450]	[0.6986]	[2.8743]
Average temp. squared	-0.0335***	0.0157	0.0127	-0.1098*
	[0.0050]	[0.0098]	[0.0153]	[0.0627]
Rainfall	0.0085***	-0.0019	-0.0021	0.0103
	[0.0008]	[0.0015]	[0.0024]	[0.0098]
Rainfall squared	-0.000006***	0.0000	0.0000	-0.0000
	[0.0000006]	[0.0000]	[0.0000]	[0.0000]
Average temp. climatologies	-14.9785	29.8370***	15.2799	-27.9565
	[11.0784]	[7.8212]	[23.5941]	[32.8589]
Average temp. climatologies sq.	0.2930	-0.6205***	-0.2894	0.3481
	[0.2472]	[0.1888]	[0.5610]	[0.7931]
Rainfall climatologies	0.0277***	-0.0224	-0.0169	0.1326*
	[0.0074]	[0.0139]	[0.0208]	[0.0754]
Rainfall climatologies sq.	-0.00002***	0.00002*	0.0000	-0.0001
	[0.000005]	[0.00001]	[0.0000]	[0.0001]
AWC class (mm)	0.0175	-0.0611	-0.1709	0.3311
	[0.0113]	[0.0571]	[0.1544]	[0.3791]
Ph top soil	-2.3035***	0.0709	0.4846	-5.7633
	[0.6047]	[0.9208]	[2.0341]	[4.3063]
Ph top soil squared	0.1875***	-0.0096	-0.0538	0.5580
	[0.0553]	[0.0843]	[0.1854]	[0.4010]
Gravel top soil	-0.0272***	-0.0219	-0.0395	0.0769
	[0.0070]	[0.0173]	[0.0325]	[0.0970]
Population density	1.8790***	-2.1749***	-3.3388***	9.3552**
	[0.1878]	[0.5154]	[1.2163]	[3.7156]
Agro-Ecological Region FE	yes	yes	yes	yes
Constant	-8.663	36.484***	34.172	-71.140
	[12.121]	[12.491]	[35.804]	[65.277]
Observations	4,960	4,960	2,352	4,960
Number of farm households	1,243	1,243	1,106	1,243

Notes: Robust standard errors in brackets *** p<0.01, ** p<0.05, * p<0.1

Table A-3: Shannon index Crop Evenness Second Stage Regressions Pseudo-Fixed effect IV estimations (complete results)

	[1] Crop income	[2] Variance	[3] Negative semi-variance	[4] Skewness
Shannon Index (Evenness)	2.0084***	-0.9823*	-1.7670**	7.7899*
	[0.2387]	[0.5298]	[0.7622]	[4.2511]
Livestock assets	0.0332***	-0.0069	-0.0074	0.0637
	[0.0100]	[0.0147]	[0.0321]	[0.0994]
Salaries & remittances	-0.4909***	0.4039	0.5757	-4.4271
	[0.1174]	[0.3865]	[0.5230]	[3.2783]
Land	0.0627	0.1962*	0.3261**	-1.3764*
	[0.0493]	[0.1061]	[0.1617]	[0.8038]
Educated adults	0.0628**	-0.0335	-0.1211	0.0213
	[0.0301]	[0.0756]	[0.1120]	[0.6041]
Women's headship dummy	0.0275	-0.1773	-0.1146	2.3599
	[0.0805]	[0.2912]	[0.3717]	[2.7034]
Transport dummy	0.0326	-0.0368	-0.0913	0.2398
	[0.0510]	[0.1236]	[0.1779]	[1.0826]
Land title deed	-0.0931**	0.0478	0.1798	-0.6259
	[0.0374]	[0.0639]	[0.1206]	[0.4568]
Mortality dummy	0.0552	-0.1848	-0.2897	1.8658
	[0.0696]	[0.1555]	[0.1979]	[1.1827]
Distance fertilizer seller	-0.0720***	0.0492	0.0559	-0.2481
	[0.0249]	[0.0570]	[0.0759]	[0.4875]
Average temperature	1.2333***	-0.2488	-0.1515	3.0539
	[0.2020]	[0.3188]	[0.5371]	[2.1008]
Average temp. squared	-0.0296***	0.0083	0.0068	-0.0794
	[0.0049]	[0.0080]	[0.0136]	[0.0528]
Rainfall	0.0085***	-0.0010	-0.0026	0.0031
	[0.0008]	[0.0014]	[0.0021]	[0.0098]
Rainfall squared	-0.000005***	0.0000	0.0000	-0.0000
	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Average temp. climatologies	-39.2374***	10.7134	58.5796**	-35.9383
	[9.4934]	[7.6957]	[26.6909]	[139.3041]
Average temp. climatologies sq.	0.8470***	-0.2256	-1.2602**	0.6012
	[0.2147]	[0.1872]	[0.6421]	[3.1506]
Rainfall climatologies	0.0341***	-0.0131	-0.0044	0.1280
	[0.0088]	[0.0143]	[0.0271]	[0.0804]
Rainfall climatologies sq.	-0.00002***	0.0000	0.0000	-0.0001
	[0.0000]	[0.0000]	[0.0000]	[0.0001]
AWC class (mm)	0.0164	-0.0644	-0.1508	0.3312

	[0.0139]	[0.0608]	[0.1887]	[0.4138]
Ph top soil	-0.8225	0.6416	-0.8477	-2.6139
	[0.6989]	[0.9385]	[2.3949]	[6.8430]
Ph top soil squared	0.0485	-0.0547	0.0830	0.2485
	[0.0638]	[0.0853]	[0.2182]	[0.6285]
Gravel top soil	-0.0286***	-0.0199	-0.0291	0.0787
	[0.0084]	[0.0182]	[0.0390]	[0.1045]
Population density	1.2313***	-1.3483***	-2.5969***	6.9111**
	[0.1624]	[0.4164]	[0.8066]	[3.1230]
Agro-Ecological Region FE	yes	yes	yes	yes
Constant	-36.2024***	17.8063	77.4022*	-81.2070
	[10.7041]	[12.4709]	[40.8915]	[160.4480]
Observations	4,960	4,960	2,469	4,960
Number of farm households	1,243	1,243	1,110	1,243

Notes: Robust standard errors in brackets *** p<0.01, ** p<0.05, * p<0.1

Table A-4: Herfindahl-Hirschman (HH) index Crop Specialization Second Stage Regressions Pseudo-Fixed effect IV estimations (complete results)

	[1] Crop income	**[2] Variance**	**[3] Negative semi-variance**	**[4] Skewness**
HH Index (Specialization)	**-4.7716*****	**2.6673****	**5.6421****	**-18.5475****
	[0.6295]	**[1.3091]**	**[2.5676]**	**[8.8782]**
Livestock assets	0.0358***	-0.0019	-0.0166	0.0593
	[0.0105]	[0.0154]	[0.0294]	[0.1035]
Salaries & remittances	-0.5004***	0.4206	0.4761	-4.2701
	[0.1222]	[0.3914]	[0.4976]	[3.3468]
Land	0.1113**	0.1511	0.1966	-1.3077*
	[0.0487]	[0.0971]	[0.1304]	[0.6877]
Educated adults	0.0641**	-0.0211	-0.1179	-0.0147
	[0.0312]	[0.0741]	[0.1077]	[0.5927]
Women's headship dummy	0.0521	-0.1828	-0.1755	2.5172
	[0.0835]	[0.2954]	[0.3389]	[2.7664]
Transport dummy	0.0180	-0.0255	0.0455	0.1596
	[0.0538]	[0.1272]	[0.1816]	[1.1058]
Land title deed	-0.1037***	0.0521	0.1587	-0.6981
	[0.0392]	[0.0676]	[0.1162]	[0.4832]
Mortality dummy	0.0972	-0.2121	-0.2197	2.0216*

	[0.0751]	[0.1613]	[0.2005]	[1.2119]
Distance fertilizer seller	-0.0745***	0.0515	0.0846	-0.2988
	[0.0263]	[0.0591]	[0.0776]	[0.5044]
Average temperature	1.1443***	-0.2791	-0.2316	2.6442
	[0.1990]	[0.3079]	[0.4815]	[2.0653]
Average temp. squared	-0.0289***	0.0091	0.0102	-0.0787
	[0.0049]	[0.0080]	[0.0123]	[0.0540]
Rainfall	0.0082***	-0.0007	-0.0001	0.0012
	[0.0008]	[0.0014]	[0.0019]	[0.0095]
Rainfall squared	-0.0000***	0.0000	-0.0000	0.0000
	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Average temp. climatologies	-9.9647	5.0341	-0.6344	-22.8794
	[16.5844]	[5.1903]	[21.1712]	[39.0007]
Average temp. climatologies sq.	0.2163	-0.1109	0.0595	0.4147
	[0.3771]	[0.1337]	[0.5222]	[0.9652]
Rainfall climatologies	0.0348***	-0.0071	-0.0144	0.1144
	[0.0092]	[0.0138]	[0.0273]	[0.0765]
Rainfall climatologies sq.	-0.0000***	0.0000	0.0000	-0.0001
	[0.0000]	[0.0000]	[0.0000]	[0.0001]
AWC class (mm)	0.0168	-0.0661	-0.1719	0.3328
	[0.0142]	[0.0617]	[0.1909]	[0.4257]
Ph top soil	-1.2539	1.0006	2.5961	-0.4872
	[0.9733]	[0.9578]	[2.3759]	[4.6327]
Ph top soil squared	0.0883	-0.0836	-0.2306	0.0466
	[0.0897]	[0.0867]	[0.2159]	[0.4197]
Gravel top soil	-0.0231***	-0.0200	-0.0452	0.0845
	[0.0088]	[0.0182]	[0.0386]	[0.1056]
Population density	1.1649***	-1.1846***	-2.5065***	6.8435**
	[0.1658]	[0.4112]	[0.7661]	[3.2094]
Agro-Ecological Region FE	-0.4240	0.3342	-1.4341	2.4563
Constant	-2.1998	10.6029	13.3853	-56.6244
Observations	4,960	4,960	2,593	4,960
Number of hhid	1,243	1,243	1,120	1,243

Notes: Robust standard errors in brackets *** p<0.01, ** p<0.05, * p<0.1

Appendix II Robustness check using alternative estimation procedures

In this Appendix, we present the results of additional sensitivity analysis.

Table A-5: Results Fixed Effects estimations (crop diversification indices treated as exogenous) - crop income regressions (complete results)

	Dependent variable: hyperbolic sine transformations of crop income		
	[1]	**[2]**	**[3]**
Count index	0.0797***		
	[0.0067]		
Shannon index		0.4798***	
		[0.0370]	
Herfindahl-Hirschman index			-1.0337***
			[0.0893]
Livestock assets	0.0422***	0.0438***	0.0446***
	[0.0086]	[0.0086]	[0.0085]
Salaries & remittances	-0.7924***	-0.7626***	-0.7727***
	[0.1205]	[0.1215]	[0.1219]
Land	0.2652***	0.2759***	0.2928***
	[0.0353]	[0.0348]	[0.0346]
Educated adults	0.0533**	0.0538**	0.0538**
	[0.0263]	[0.0263]	[0.0263]
Women's headship dummy	0.0328	0.0378	0.0431
	[0.0766]	[0.0772]	[0.0776]
Transport dummy	0.0856**	0.0846*	0.0830*
	[0.0435]	[0.0436]	[0.0437]
Land title dee	-0.0500	-0.0574*	-0.0585*
	[0.0311]	[0.0308]	[0.0308]
Mortality dummy	-0.0223	-0.0120	-0.0047
	[0.0605]	[0.0600]	[0.0603]
Distance fertilizer seller	-0.0714***	-0.0759***	-0.0765***
	[0.0216]	[0.0215]	[0.0217]
Average temperature	0.9840***	0.8989***	0.8679***
	[0.1490]	[0.1513]	[0.1526]
Average temp. squared	-0.0240***	-0.0229***	-0.0225***
	[0.0037]	[0.0038]	[0.0038]
Rainfall	0.0062***	0.0062***	0.0061***
	[0.0006]	[0.0006]	[0.0006]
Rainfall squared	-0.000004***	-0.000004***	-0.000004***

	[0.0000]	[0.0000]	[0.0000]
Population density	1.6052***	1.4438***	1.4358***
	[0.1630]	[0.1566]	[0.1561]
Constant	-12.1116***	-10.0020***	-8.5126***
	[1.8358]	[1.8009]	[1.7873]
Observations	4,960	4,960	4,960
Number of farm households	1,243	1,243	1,243

Notes: Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A-6: Estimation results – Crop Richness Index: Fixed Effects estimations (crop diversification index treated as exogenous) of Mean, Variance, Semi-Variance and Skewness of Crop Income

	[1] Crop income	[2] Variance	[3] Negative semi-variance	[4] Skewness
Count index	0.0797*** [0.0067]	-0.682*** [0.0214]	-0.0818** [0.0359]	0.4986** [0.2052]
Observations	4,960	4,960	2,540	4,960
Number of farm households	1,243	1,243	874	1,243

Notes: Fixed effect estimation. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. Full results of the crop income regression [1] are reported in column 1 of Table A-5. Full results of the variance, semi-variance and skewness regressions are available from the authors upon request.

Table A-7: Estimation results – Crop Evenness Index: Fixed Effects estimations (crop diversification index treated as exogenous) of Mean, Variance, Semi-Variance and Skewness of Crop Income

	[1] Crop income	[2] Variance	[3] Negative semi-variance	[4] Skewness
Shannon index	0.4798*** [0.0370]	-0.2734*** [0.0935]	-0.3469*** [0.1257]	2.0641** [0.8696]
Observations	4,960	4,960	2,581	4,960
Number of farm households	1,243	1,243	892	1,243

Notes: Fixed effect estimation. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. Full results of the crop income regression [1] are reported in column 2 of Table A-5. Full results of the variance, semi-variance and skewness regressions are available from the authors upon request.

Table A-8: Estimation results – Crop Specialization Index: Fixed Effects estimations (crop diversification index treated as exogenous) of Mean, Variance, Semi-Variance and Skewness of Crop Income

	[1] Crop Income	[2] Variance	[3] Negative semi-variance	[4] Skewness
Herfindahl-Hirschman index	-1.0337***	0.5549**	0.7722**	-4.3260**
	[0.0893]	[0.2269]	[0.3215]	[2.0365]
Observations	4,960	4,960	2,571	4,960
Number of farm households	1,243	1,243	897	1,243

Notes: Fixed effect estimation. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. Full results for the crop income regression [1] are reported in column 3 of Table A-5. Full results of the variance, semi-variance and skewness regressions are available from the authors upon request.

Table A-9: Fixed effect IV estimations results – First Stage regressions

	[1] Richness	[2] Evenness	[3] Specialization
Past droughts	0.2503***	0.0560***	-0.0232***
	[0.0456]	[0.0082]	[0.0036]
Credit	-1.2987***	-0.2179***	0.0767***
	[0.1931]	[0.0368]	[0.0171]
Agro-regional FE	Yes	Yes	Yes
Observations	4,960	4,960	4,960
Number of households	1,243	1,243	1,243

Notes: Fixed Effects IV regression estimations. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. Full results are available from the authors upon request.

Table A-10: Second Stage Results Fixed Effects IV estimations - crop income regressions

	Dependent variable: hyperbolic sine transformations of crop income		
	[1]	[2]	[3]
Count index	0.2903***		
	[0.0519]		
Shannon index		1.6160***	
		[0.2684]	
Herfindahl-Hirschman index			-4.2978***
			[0.7236]
Livestock assets	0.0288***	0.0355***	0.0361***
	[0.0100]	[0.0098]	[0.0106]
Salaries & remittances	-0.6445***	-0.5594***	-0.5338***
	[0.1201]	[0.1260]	[0.1338]
Land	0.0631	0.1201**	0.1384**
	[0.0642]	[0.0546]	[0.0557]
Educated adults	0.0595**	0.0605**	0.0628**
	[0.0294]	[0.0292]	[0.0310]
Women's headship dummy	0.0125	0.0316	0.0516
	[0.0832]	[0.0833]	[0.0889]
Transport	0.0437	0.0447	0.0248
	[0.0502]	[0.0506]	[0.0556]
Land title deed	-0.0600	-0.0839**	-0.0974**
	[0.0368]	[0.0368]	[0.0402]
Mortality dummy	0.0055	0.0372	0.0842
	[0.0680]	[0.0674]	[0.0762]
Distance fertilizer seller (km)	-0.0563**	-0.0729***	-0.0743***
	[0.0244]	[0.0236]	[0.0259]
Average temperature	1.4922***	1.1526***	1.1084***

	[0.2268]	[0.1901]	[0.1981]
Average temp. squared	-0.0328***	-0.0280***	-0.0280***
	[0.0050]	[0.0046]	[0.0049]
Rainfall	0.0083***	0.0079***	0.0080***
	[0.0009]	[0.0009]	[0.0009]
Rainfall squared	-0.000006***	-0.000005***	-0.000006***
	[0.0000]	[0.0000]	[0.0000]
Population density	1.8551***	1.2853***	1.1991***
	[0.1979]	[0.1644]	[0.1713]
Constant	-21.4534***	-13.3746***	-8.3064***
	[3.3745]	[2.2971]	[2.1283]
Observations	4,960	4,960	4,960
Number of farm households	1,243	1,243	1,243

Notes: Fixed Effects (within) regression estimations. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1.