The Impact of Maize Hybrids on Income, Poverty, and Inequality among Smallholder Farmers in Kenya

Mary K. Mathenge, Melinda Smale, and John Olwande

8/18/2012

Paper presented at the Pre-Conference Workshop on "Innovations in impact assessment of agricultural research: Theory and practice," Meetings of the International Association of Agricultural Economists (IAAE), Foz do Iguaçu, Brazil, 18th August 2012.

1. Introduction

For decades, Kenya has been depicted a maize “success story” in Sub-Saharan Africa, known for rates of hybrid maize adoption during the 1960s and 70s that paralleled those of the U.S. Corn Belt thirty years earlier (Gerhart 1975; Byerlee and Eicher 1997; Smale and Jayne 2010). Over the past few decades, however, a general perception of stagnating adoption and production (Hassan 1996; De Groote 2005) has been supported by FAO data and a rising maize import bill. Replacement of older hybrids by newer releases appears to have been slow (Hassan 1998; Smale, Olwande and De Groote, forthcoming), dampening yield potential on farms. For example, a hybrid released in 1986 and derived from this first hybrid still dominates the maize fields of Kenya, despite the dramatic increase in the number of hybrids and breadth of seed suppliers s seed markets liberalize (Swanckaert 2012).

A number of in-depth studies of maize seed adoption have been conducted in specific regions of Kenya (e.g., Ouma et al. 2002, Salasya et al. 2002, Wekesa et al. 2002). Based on large-scale surveys, analyses of seed adoption and maize productivity have been implemented by Kenya Agricultural Research Institute (KARI) with the International Maize Improvement Center (CIMMYT) in 1992, 2002, and 2010 (Hassan et al. 1998, De Groote et al. 2005, 2006). Using panel data collected in 1997, 2000, and 2004 by Tegemeo Institute and Michigan State University from a panel of households in maize-growing areas in Kenya, Suri (2011) applied a correlated random coefficient model to demonstrate that heterogeneous net returns to hybrid seed explain adoption rates. She compared distributions of hybrid and non-hybrid maize yields, illustrating that not only were mean yields higher, but the variance of yields was much lower among hybrid growers. Recently, Jones et al. (2012) tested the effect of hybrid use on the mean, variance, and skewness of maize yields with a stochastic production function applied to survey data collected by Tegemeo Institute during the 2006-7 cropping season. The authors found that hybrids enhance mean yields, and also reduce the exposure of smallholders to extremely low yields. Since Suri's analysis, overall adoption rates for hybrid seed have also risen to over 80%, in part as a reflection of seed-to-grain price ratios and the progress of seed liberalization (Smale, Olwande and De Groote, 2012).

To complement research about hybrid seed adoption, a number of studies have been conducted on maize seed industry and seed supply. For example, Ayieko and Tschirley (2006) assessed the structure of the seed system across a range of crops. The authors estimated that roughly two-thirds of maize seed planted was purchased from formal sources (primarily national parastatal organizations and private companies). Despite a highly diversified seed system, with the exception of maize and rice, most seed for other major crops in Kenya was farm-saved. Wangia et al. (2004) concluded that after the liberalization which took place from 1986 to 1995, private sector participation at all nodes in the marketing system increased substantially, resulting in an easy flow of supplies to many parts of Kenya. A seed sector study conducted by Nambiro et al. (2004) in the Trans Nzoia District found some impact of the liberalization of the seed industry on the distribution side, where private retailers had broken the previous monopoly of the Kenya Farmers' Association. However, according to the authors, at the impact of seed liberalization on maize production was minimal. At that time, they estimated that KSC provided 97% of the seed, dominated by one variety.

There is some more recent evidence that liberalization has led to entry of new seed companies in the maize market. In her thesis, Swanckaert (2012) reports that while KSC was the only maize seed company prior to 1992, currently there are 11 companies with varieties registered to their names. Currently, the plant variety registry of the Kenya Plant Health Inspectorate (KEPHIS) lists 164 varieties released from 1964 up to 2009, with 85 percent of these registered since 2000. The numbers of improved maize varieties and hybrids grown on farms has also increased tremendously. While Hassan (1998) found only 12 hybrids grown by farmers in 1992, Tegemeo data indicate that the number of hybrids on farms was 33 in 2004 to 50 in 2010. Nonetheless, Swanckaert (2012) concludes that although competition in the seed market has intensifiec, the impact of new seed companies on market concentration has been smaller than expected.

To our knowledge, a missing link in existing research on maize hybrids in Kenya has been a rigorous analysis of the impacts of seed adoption on farmer welfare. Mwabu et al. (2007) applied a bivariate probit model to explore the relationship of adoption of improved varieties to poverty of households in rural districts of Laikipia and Suba. They found a negative correlation between poverty and adoption of improved maize seed.

The objective of this study is to produce an initial assessment of the impact of maize hybrids on income, poverty, and inequality of Kenya smallholders from 2000 through 2010. We contribute to the body of knowledge about maize hybrids in Kenya by documenting these impacts with advanced econometric methods; we contribute to empirical examples of the application of instrumental variables, fixed effects models to estimate impacts of agricultural technology.

We use data collected by Tegemeo Institute through repeated surveys of smallholder farmers in the major maize-growing areas of Kenya. We measure income as current, total household income from farm and non-farm sources. Our indicator of inequality is Stark’s index of relative deprivation with respect to income. Unlike the Gini coefficient, Stark's index can be constructed for individual households. To estimate impacts on income and income inequality, we apply an instrumental variables regression with panel data and a fixed effects model to account for and test the endogeneity of hybrid seed use in impact equations caused by selection bias, simultaneity, or correlated errors. To compare poverty status among users and non-users, we apply Foster-Greer-Thorbecke indices. To address the potential endogeneity of hybrid seed use in poverty status in an econometric model, we then define poverty status as a binary outcome variable for each household, and apply an instrumented control function approach with the Mundlak-Chamberlin device (Correlated Random Errors, or CRE).

2. Data

Tegemeo Institute of Egerton University has implemented a five-year panel survey (1997, 2000, 2004, 2007, and 2010) in Kenya. The motivation for data collection has been to evaluate income, poverty, and developmental pathways, providing both routine policy advice and longer-term, in-depth analyses to Kenyan national decision-makers. Although crop-specific analyses have been viewed as secondary, Tegemeo, in partnership with Michigan State University, has emphasized policy research on the maize value chain, along with dairy and horticulture.

The sampling frame for data collection was prepared in consultation with the Kenya National Bureau of Statistics (KNBS) in 1997. The process is described by Argwings-Kodhek et al (1999). Census data were used to identify all non-urban divisions in the country, and these were assigned to one or more agro-ecological zones (AEZ) based on the 1990 Census, District Development Plans and the Farm Management Handbook. Within each AEZ, two or three divisions were chosen based on their importance (size of population). In each selected division, villages were selected by local officials using blind equal chance ballot. Households were selected within villages with systematic sampling from a list, and a random start. A total of 1,578 households were selected in 24 districts. The sample excluded large farms with over 50 acres and two pastoral areas. For the purposes of analysis, households were then grouped into 9 agro-regional zones, which represented a combination of agro-ecological zone, administrative and political boundaries: Northern Arid, Coastal Lowlands, Eastern Lowlands, Western Lowlands, Western Transitional, High-Potential Maize Zone, Western Highlands, Central Highlands, and Marginal Rain Shadow. The first survey was conducted in 1997, with a much more restricted survey instrument than those applied in later years. For that reason, 1997 is not included in this analysis. We use agro-ecological zone, rather than agro-regional zone, because it is most relevant when investigating maize production. Here, we employ a balanced panel of 1243 households.

Households in Turkana and Garissa districts (the Northern Arid zone) were not interviewed after 2000. Excluding these districts, the attrition rate for the panel was 13% in 2010 compared to the initial survey, conducted in 1997. Reasons for non-participation in subsequent surveys were recorded. Some of the main reasons for this attrition are related to death of household heads and spouses leading to dissolution of households, and relocation of households from the study areas.

3. Conceptual Approach

The conceptual framework for this analysis is the class of analytical approaches that is generally known as "treatment models," described in-depth in an extensive body of literature that addresses the statistically-based measurement of the social and economic impacts of public programs (e.g., Ravallion 1994, Angrist and Krueger 2001, de Janvry et al. 2010). The motivation for these approaches is understood as essentially one of missing data (Ravallion 1994). That is, we observe the values of outcome variables, such as indicators of income or poverty, for the group of households who are targeted by a program or policy, as well as those who are not. We do not observe the values of outcome variables for targeted households had they not been targeted.

In the case of hybrid seed use among smallholders in Kenya, we observe values of outcome variables for adopters and non-adopters, but not for adopters had they not adopted. Unlike households targeted by a program or policy, adopters in Kenya choose to grow hybrid seed, or "self-select" into the treatment group. We know from past research, and from our data, that adopting farmers are generally those who are wealthier in terms of various types of human capital and have more access to "soft" (information and financial services) and "hard" (roads, vehicles, and marketplaces) market infrastructure. Thus, any estimate of the impact of hybrid seed use on outcome variables that does not take this into account will exhibit a bias due to the underlying effects of these factors. This selection bias, attributable in our case to self-selection through seed choice rather than explicit targeting, reflects the fact that adopters are better off than non-adopters even before they adopt.

Experimental and quasi-experimental methods have been proposed to address selection bias. Experimental approaches include randomized treatments or randomized controlled trials. These approaches are not feasible in our case, where Kenyan farmers have grown maize hybrids since they were first released in 1964. Quasi-experimental approaches consist of instrumental variables regression, propensity score matching, difference-in-difference estimation, and regression discontinuity. Propensity score matching involves estimating the probability that a farmer plants hybrid seed as a function of a set of observed explanatory variables, and comparing outcome variables for adopters and non-adopters who have high likelihood of adoption. The implicit assumptions of this approach are that 1) only the factors that matter in adoption are those specified in the regression equation, and that 2) all relevant determinants are observable. We know that these assumptions are difficult to justify when we have a limited number of observed explanatory variables and we know that certain intrinsic, unobservable attributes influence the seed choices made by Kenyan farmers.

While feasible in our case, the matching approach is not well-suited to analyzing the impact of this project because specific socio-economic groups were not targeted for an intervention. In addition, the impact outcomes we measure are complex variables. Handa and Maluccio (2010) concluded that matching is more promising as an approach for evaluating easily measured outcomes, such as those related to child schooling and health, than it is for more complex outcomes, such as expenditures (or income).

Panel data methods are also designed to control for some unobserved heterogeneity that is correlated with observed variables, through fixed effects models or first-differencing (De Janvry et al. 2010; Ravallion 1994; Wooldridge 2002). Difference-in-difference models are the most common research design for policy analysis with panel data collecting in two phases, comparing the change in outcome variables between the sub-population that received a treatment (or self-selected into the group) and the sub-population that did not. In the case of a panel with only two years of observations, fixed effects estimation is equal to first-differencing. Generally, in applying difference-in-difference approaches, the treatment and control groups are separated geographically and sampling units chosen to ensure that mean outcomes are compared between otherwise comparable groups. This is not a feature of the sampling design that has generated Tegemeo's data, which supports handling self-selection bias through other means.

Of the feasible approaches to quantitative assessment of social and economic impacts, we conclude that instrumental variables combined with panel data methods is the best suited to our data-generating process and research hypotheses. The instrumental variable approach relies on econometric methods to separate the effects of belonging to a group (through targeting or choice) from those of other factors that influence impact. Identifying valid instrumental variables is the major challenge associated with this method. In our case, valid instrumental variables are those that determine whether or not a farmer uses hybrid seed in maize production, but only influence outcome variables through hybrid seed use.

In summary, we hypothesize that either because of observables (farm size, education, labor supply) or unobservables (e.g., intrinsic management ability, unmeasurable soil quality), or simultaneity (a feedback process), the decision by Kenyan smallholders to grow hybrid maize seed is endogenous in income. In the presence of endogeneity, which results from non-random selection of analytical units, the correlation of independent variables with error terms, or a chain of causality among independent and dependent variables, estimators generated by ordinary least squares are biased. They may still be consistent if correlations do not occur over time, which we doubt in the case of panel data. To test this hypothesis we need to use instrumental variables methods. Our approach estimates the local average treatment effect (LATE): the effect of self-selection into the group of farmers that uses maize hybrids on income, inequality, poverty and inequality, identified through instrumentation. Next we define the outcome variables, present the estimation procedure, and describe the explanatory variables.

4. Estimation Strategy

4.1. Impact outcomes

We considered three outcome variables, namely household income, income inequality, and poverty. Household income, expressed in current, nominal terms, is comprised of net crop income (gross value of crop production less input costs); net livestock income (gross value of livestock products plus sales of live animals less purchases of live animals plus input costs); salaries for household members; net business income for household members; income from informal labour employment for household members; and remittances, pension and share dividends received by all household members. For purposes of comparison, we have also considered real household income.

We estimate poverty outcomes from the alues of per capita monthly income, comparing them initially between users and non-users using summary statistics that do not account for potential endogeneity on hybrid seed use. Perhaps the three most widely used poverty indicators can all be derived from the Foster-Greer-Thorbecke (FGT) index. The FGT index combines a) the extent of poverty (*headcount*) b) the depth of poverty (*the gap*) and c) the severity of poverty, also interpretable as the coefficient of variation among the poor.

The formula for the FGT index is:

$$FGT_{\alpha} = (1/n)\sum_{i=1}^{h}(\frac{z-y_i}{z})^{\alpha} \tag{4}$$

The parameter z is the poverty line, n is the number of people in the reference area (village, district, nation), h is the number of poor people (those with incomes at or below z), y_i are individual incomes, and α is a weighting factor for social policy. If all persons below the poverty line are weighted in the same way, α is low. If it is high, those with the lowest incomes are weighted more heavily. The higher the FGT, the greater is the poverty in the reference area.

When $\alpha = 0$, the formula reduces to h/n, which is the headcount ratio, or fraction of the population below the poverty line. If $\alpha = 1$, the formula represents the average poverty gap. The poverty gap is the amount of income it would take to bring people in poverty up to the poverty line.

In another popular version of FGT, α=2, in which the index also can be rewritten in terms of the coefficient of variation, or inequality among those with incomes less than the poverty line. We call this the severity of poverty.

While we can use FGT indices to compare hybrid users and non-users, we needed an indicator of poverty status that is constructed per household to account for potential endogeneity in a regression context. [1] For this purpose, we employed a binary outcome variable measuring whether or not the household income fell below the poverty line. We applied the official poverty lines established by the Government of Kenya, for each survey year, in nominal KES: 1009 (2000), 1336 (2004), 1629 (2007), 2144 (2010). Since these are expressed in per capita terms per month, in order to compare them to nominal current income, we divided annual household income by the number of adult-equivalent members and 12 (the number of months/year).

Popular measures of inequality, such as the Gini or Theil indices, are also calculated over a distribution of households or individuals. People often compare themselves with others in their immediate reference group, such as a village, rather than with the whole society (Yitzhaki, 1979). Based on the observation and their analysis of the effects of migration on households in

[1] We initially considered use of predicted income values from the instrumental variables, fixed effects regression to estimate Foster-Greer-Thorbecke indices that account for endogeneity. However, the STATA routine we applied (XTIVREG2, see below) does not calculate fitted values. Similarly, we were not able to conduct non-parametric tests on equality of fitted outcome distributions when we failed to reject endogeneity.

Mexico, Stark and Taylor (1989) proposed an inequality index calculated at the individual or household level:

$$RD(Y_i) = AD(Y_i) * P(Y_i) \tag{5}$$

Relative deprivation $RD(Y_i)$ was calculated for each household i, taking the remaining households in the location as the reference group. $AD(Y_i)$ is the mean income of households in the village richer than a given household i and $P(Y_i)$ is the proportion of households in the village that are richer than a given household i in that same location (Stark and Taylor, 1989). To construct the index, households were ranked by Y from lowest to highest. As is the case with other indicators of inequality, relative deprivation is typically calculated with income data, although it can also be computed with other variables, such as land. Conforming to other outcome variables, we computed the index current nominal income. The higher the value of the index, the greater is the relative deprivation of the household.

4.2. Econometric models

Separate regressions were estimated for each outcome variable of interest described in 4.1. We applied panel data methods with fixed effects and the correlated random effects with control function approach to the data to estimate the outcome equations. For diagnostic purposes, we also estimated pooled, random effects, and first difference models. Fixed effects models help control for unobserved heterogeneity among farm households that is time-invariant and correlated with independent variables (Wooldridge 2002). Often this is considered to capture intrinsic features that are not easily measured (such as farm management capability). We add year effects to the estimation to control for other time-varying, unobservable effects.

The model of interest using the Tegemeo panel data then becomes:

$$\mathbf{Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma Z_{it} + \alpha_i + \varepsilon_{it}} \qquad \mathbf{i= 1,..,N\ t=1,..,T} \qquad (1)$$

Where Y is the outcome variable of interest, X is a vector of exogenous explanatory variables, Z is hybrid seed use, $\boldsymbol{\alpha_i}$ represents time-invariant unobserved factors that affect the outcome variable and $\boldsymbol{\varepsilon}$ is the random error term.

Model (1) can be estimated using the usual regression methods depending on the construction of the outcome variable. For a continuous outcome variable l, Model (1) can be estimated using the standard fixed effects model for panel data. As we have argued above, we have strong conceptual and empirical grounds to expect that our variable of interest, Z (hybrid seed use), is endogenous due to measurement error, simultaneity, or selection. In that case, estimating Model (1) would result in biased estimates, overstating the impacts of hybrid seed use on outcome variables.

To test the endogeneity of the binary variable for hybrid seed use, we applied a STATA module developed for instrumental variables analysis specifically with panel data (Schaffer 2010). The model is estimated via two-stage least squares, with a binary variable measuring hybrid seed use in the first stage. Angrist (2001) and Angrist and Krueger (2001) argue that even in the case of a dichotomous variable in the first of the two equations, two-stage least squares produces consistent estimators that are less sensitive to assumptions about functional form. They advocate this approach over use a nonlinear models such as probit or logit in two-stage least squares regression (2001: 80). Regression coefficients can be interpreted in terms of causality (sign and significance), but not in terms of the magnitude of their effects as would be the case in a structural model. Since all outcome variables are linear, we used the fixed effects, two-stage least squares method for regressions with a binary variable measuring hybrid seed use in the first stage.

Standard diagnostic statistics include tests of a) endogeneity of the adoption variable; b) the relevance of the instrument set, and c) model identification. We estimated all models with robust variance-covariance matrices.

Model diagnostics for instrumental variables regression include simple correlation coefficients among dependent variables and instruments, and evaluation of the F-test for excluded instruments in the first stage. Hansen's J is the relevant test for overidentifying restrictions with robust errors. Rejection of the null hypothesis that instruments are uncorrelated with the error term casts doubt on their validity. The Kleibergen-Paap statistic provides a test for the weakness (underidentification) of instruments. Rejection of the null hypothesis is supporting evidence that instruments are correlated with the endogenous regressor. Finally, the endogeneity test is defined as the difference between the Hansen-J statistics with and without the instruments that are hypothesized to identify the endogenous variable.

The poverty outcome variable is binary, which means that the FE2SLS method is inappropriate because it implies that, in the second stage, a nonlinear function of an endogenous variable is replaced with the same nonlinear function of fitted values from a first-stage estimation (Wooldridge 2002:236). The control function approach enables us to account for and test endogeneity bias in a non-linear model when both the suspected endogenous variable and the outcome variable are binary. As in a two-stage instrumental variables model, the control function approach requires an instrumental variable to be used in the first stage, reduced form estimation of poverty status. In the second stage, however, the structural model is estimated with the observed endogenous variable and the residual from the first stage as explanatory variables. The test of endogeneity is the statistical significance of the coefficient of the residual, with bootstrapped standard errors. The control function approach is described in early work by Smith and Blundell (1986).

To generate a fixed effect interpretation model in the context of a control function, we apply correlated random effects (CRE). As proposed by Mundlak (1978) and Chamberlain (1984), the CRE model controls for unobserved heterogeneity and its correlation with observed factors in a non-linear regression context. Application of the model requires that the means of time-varying variables are included in each stage of the regression.

4.3. Explanatory variables

Explanatory variables are defined and summary statistics reported in Table 1.

Table 1: Variable Definition and Summary Statistics

Variable	Construction	2000		2004		2007		2010		Pooled	
		Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
Dependent											
Income	Sum of net income (KES) from crops, livestock, salaries, remittance, business and informal labor activities	170,234	195,802	186,535	215,563	201,659	210,659	278,463	386,998	209,223	267,240
Income inequality	Stark's inequality index (with respect to all other households in the location)	121,972	69,785	134,660	73,783	143,538	79,733	203,303	122,359	150,868	94,240
Poverty	1=poor (income per adult-equivalent per month) < official poverty line; 0 else	0. 27	0.44	0.30	0.46	0.31	0.46	0.32	0.45	0.30	0.46
Explanatory											
Hybrid adoption	1=Grow hybrid seed in the year, 0 otherwise	0.68	0.47	0.61	0.49	0.73	0.44	0.82	0.38	0.71	0.45
De jure headship	1=Reported head of household is female, 0 otherwise	0.12	0.32	0.20	0.40	0.23	0.42	0.27	0.44	0.21	0.40
Education	Average educational attainment (years) of all adults in household	7.12	2.86	7.21	2.99	7.20	3.02	7.59	3.00	7.28	2.97
Young adults	Number of adults 15-24 years	2.07	1.62	1.96	1.63	1.91	1.60	1.81	1.53	1.94	1.60
Mature adults	Number of adults 25-64 years	2.35	1.26	2.26	1.30	2.10	1.26	2.06	1.34	2.19	1.29
Extension	Km to extension service	5.45	5.93	5.27	5.87	4.42	5.23	5.16	5.29	5.08	5.60
Tarmac	Km to tarmac road	7.67	7.87	7.59	7.81	7.61	7.48	7.11	7.23	7.50	7.60
Land size	Total land area (acres), last survey	6.04	9.71	6.06	8.48	6.12	8.98	5.82	8.81	6.01	9.00
Rain	Total mm rain in the growing season associated with the survey year	583.91	271.77	685.24	298.61	611.59	195.99	413.89	200.76	573.66	265.11
Seed access	Km to nearest seller of certified maize seed	5.59	7.46	3.85	7.38	3.41	4.42	4.13	5.70	4.25	2.70
Seed-to-grain price ratio	Predicted value of farmgate seed-to-grain price ratio	11.17	2.14	11.25	2.034	11.78	2.03	7.44	2.10	10.42	6.42

Source: Authors, based on Tegemeo survey data

Our explanatory variables include only those that vary over time, excluding dummies for agricultural potential, such as agroecological zone, and administration (district), which are highly correlated with other independent variables but are taken into account in the fixed effects model with year effects. To address the crucial importance of agroecological factors, we estimate separate regressions for high and low potential areas.

Other independent variables include labor quality and quantity, measured in terms of the average education of all adults in the household, and the numbers of young and mature adults in the household. While household size and the number of children, in particular, is often argued to be endogenous, we consider the number of working-age adults to be exogenous in the short-term. We include whether the recognized household head is female or male. In fact, most women household heads in the Tegemeo panel are widows, and there are strong reasons to believe that they differ from households headed by men, most of whom have spouses. Land size, lagged to the previous survey season, is included to represent endowments of physical capital. Unfortunately, while we do have soil quality data, these are measured in terms of categorical or binary variables that cannot be included in the estimation. Distances to extension services and the tarmac road are included as explanatory variables in both income and hybrid seed regressions.

Included in the hybrid seed category are all hybrids named to be hybrids by the farmer respondent or recognized as such by the enumerator who recorded the information. We did not include improved open-pollinated varieties, which represent a very small percentage of improved maize types grown by Kenyan farmers, as do recycled hybrids.

Theory suggests that the ratio of seed-to-grain prices, as well as the distance to the nearest seller of certified maize seed, are potentially good instruments for identifying the effect of hybrid seed use on household income (Heisey et al. 1998). We hypothesize that distance the nearest seller of certified maize seed is strongly related to hybrid seed use, but not necessarily to overall household income. Virtually all of households surveyed rely on a range of farm and non-farm income sources, in addition to maize. Seed price, calculated here as seed costs per kg at the farm gate, is observed only for farmers who purchase seed (most of which is hybrid maize), even though all farmers face prices. To address this challenge, we predicted the seed-to-grain price ratio using fixed effects for district, agro-ecological zone, year, and other household characteristics that are related to market participation. We preferred to use the natural logarithms of income because its distribution is less skewed that those of the level variables. The regression predicting the seed-to-grain price ratio and the distributions of incomes and its natural logarithm are included in the Appendix.

5. Results

5.1. Descriptive statistics

Use rates for hybrid seed are shown in Table 2, by year and AEZ. Tegemeo's panel data suggest that overall rates climb from about two-thirds of farmers in the early 2000s to over four-fifths a decade later. Rates have exceeded 80% in the highlands since the beginning of the period, and have varied most over the years in the lowlands.

Table 2: Percentage of households growing hybrid seed, by year

	2000	2004	2007	2010
Coastal lowland	28.4	1.3	37.8	37.7
Lowland	73.5	50.0	61.8	91.2
Lower midland 3-6	26.5	17.2	38.8	56.3
Lower midland 1-2	75.2	75.0	86.1	89.6
Upper midland 2-6	79.7	68.2	77.4	92.6
Upper midland 0-1	87.1	85.3	88.4	91.1
Lower highland	85.2	85.2	91.0	95.1
Upper highland	92.3	92.3	100.0	100.0
All zones	68.1	61.1	73.1	82.0

Source: Authors.

Comparisons of hybrid users and non-users on outcome indicators are presented in Tables 3 through 5. With respect to household income, hybrid maize growers have statistically higher income on average than farmers who grew local or improved open-pollinated varieties (Table 3).

Table 3: Household income by use of hybrid maize seed and year

Year	No hybrid maize seed		Hybrid maize seed	
	Mean	Std. Deviation	Mean	Std. Deviation
2000	108,175	112,909	199,419	218,840
2004	135,820	164,140	219,240	237,350
2007	127,596	126,258	229,442	229,039
2010	160,571	230,835	306,483	410,571

Source: Authors.
Note: T-tests show adopters have significantly higher income than non-adopters in each year at 1% significance. Household annual income reported in nominal KES.

As measured by Stark's index, average relative deprivation is significantly higher among non-hybrid growers than growers of maize hybrids, in each year (Table 4). This suggests that relative to other households in their sublocation, hybrid growers are generally ranked higher in terms of income.

Table 4: Income inequality (relative deprivation) by use of hybrid maize seed and year

Year	No hybrid maize seed		Hybrid maize seed	
	Mean	Std. Deviation	Mean	Std. Deviation
2000	133138.7	81240.28	98598.22	69954.22
2004	148034.2	94143.7	114019.2	64689.48
2007	152622.1	90100.65	117576.6	67259.29
2010	218899.6	155281.5	132879.6	97549.28

Source: Authors.
Note: T-tests show adopters have significantly lower relative deprivation than non-adopters in each year at 1% significance. Inequality measured by Stark's index of relative deprivation (defined in text) with respect to income in nominal KES.

The relationship of hybrid use to poverty status, as measured by the FGT indices described above, is summarized in Table 5. The proportion of households falling below the poverty line is around twice as high for farmers who do not grow hybrid maize in each year. The average depth of poverty is also significantly greater for poor households that do not grow hybrid maize, relative to hybrid maize growers in all years but 2010. Similarly, the severity of poverty, or the variation among the poor, is less for hybrid users than non-users in three out of the four survey years.

Table 5: Use of maize hybrids by poverty status and year

	Headcount			Poverty depth			Poverty severity		
	Grow maize hybrid			Grow maize hybrid			Grow maize hybrid		
Year	No	Yes		No	Yes		No	Yes	
2000	41.67	19.63	*	0.468	0.354	*	0.283	0.181	*
2004	42.14	21.96	*	0.450	0.386	*	0.267	0.206	*
2007	46.20	25.77	*	0.416	0.359	*	0.228	0.175	*
2010	54.70	25.43	*	0.417	0.389		0.229	0.199	*

Source: Authors.
Note: Statistical tests show adopters were less poor by any of the three indices in all years except 2010, when poverty depth and severity was not significantly different between adopters and non-adopters.

5.2. Regression models

Regression results for the fixed effects models of income and relative deprivation, by agroecological zone, are shown in Tables 6 and 7. For purposes of comparison, pooled, random effects, and fixed effects regressions are shown in the Appendix. We also estimated difference-in-difference models, as well as real income models without year effects. These generated consistent results. We report results of the fixed effects models with nominal income and year effects, which are methodologically superior. Findings are shown in Tables 6 and 7 by production potential and in the Appendix, by maize-growing potential.

The major hypotheses tested in these regressions are the impact of hybrid maize seed on income and the endogeneity of growing hybrid maize in income. Findings suggest that growing hybrid maize seed has a positive and significant influence on total household income and a negative and significant influence on relative deprivation of households in their location with respect to total household income. This is true regardless of whether the household is located in a more favorable or less favorable zone for agricultural production. Although Angrist and Krueger (2001) caution against interpretation of the magnitudes of estimated coefficients in instrumental variables regressions with a binary endogenous regressor, we may conclude that the relative magnitudes of effects on income and relative deprivation are larger in the zone with higher potential for agricultural production.

Households with a larger supply of adult labor earn more income, and adult labor also reduces relative deprivation significantly. As expected, female headship has a negative effect on total household income, and a positive effect on relative deprivation, other factors held constant, although the coefficients are not statistically significant. Consistent with earlier studies in Kenya, education has a strong positive effect on household income, but no significant effect on relative deprivation. Results on the year dummies indicate an increase in income over time, but also a general increase in relative deprivation, as a consequence of declining real income over the same period. Rainfall is not of the expected sign in the general model, although it displays a positive effect on income in the higher potential zone. Total farm size positively influences hybrid seed use.

Table 6. Impacts of hybrid maize on income, by agricultural potential

	(1) Income	(2) Income High Potential areas	(3) Income Low Potential areas
Grow hybrid	2.24**	3.63*	1.58**
	(0.602)	(1.578)	(0.599)
Female head	-0.01	-0.22*	0.20
	(0.078)	(0.110)	(0.130)
Education	0.02*	0.01	0.03+
	(0.010)	(0.016)	(0.018)
Children	0.03*	0.03+	0.04+
	(0.013)	(0.018)	(0.023)
Young adults	0.10**	0.09**	0.12**
	(0.016)	(0.024)	(0.026)
Mature adults	0.00	-0.00	-0.00
	(0.004)	(0.006)	(0.006)
Elderly	-0.00	0.00	-0.01
	(0.005)	(0.008)	(0.008)
Farm size	0.01*	0.00	0.01*
	(0.003)	(0.004)	(0.007)
Rainfall	-0.00*	0.00*	-0.00
	(0.000)	(0.000)	(0.000)
2004	0.28**	0.01	0.47**
	(0.061)	(0.061)	(0.107)
2007	0.15**	0.06	0.31**
	(0.050)	(0.075)	(0.095)
2010	0.12	0.25*	0.18
	(0.116)	(0.111)	(0.209)
Observations	4,885	3,342	1,543
Number of hhid	1,239	845	394

Source: Authors. Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1

Table 7. Impacts of hybrid maize on relative deprivation (income), by agricultural potential

	(1) Relative Deprivation	(2) High Potential Relative Deprivation	(3) High Potential Relative Deprivation
Grow hybrid	-118,708.21**	-163,700.11	-52,322.76+
	(42,335.282)	(108,306.322)	(28,744.235)
Female head	5,046.94	10,128.26	1,278.43
	(5,399.178)	(7,039.970)	(6,642.849)
Education	-548.78	-958.90	1,099.52
	(738.080)	(1,005.320)	(967.339)
Children	-1,428.81	-369.51	-4,453.09**
	(1,025.293)	(1,294.782)	(1,367.384)
Young adults	-4,651.97**	-3,538.23*	-6,690.59**
	(1,255.276)	(1,661.448)	(1,651.570)
Mature adults	543.32+	245.07	578.53
	(278.079)	(394.993)	(359.187)
Elderly	-148.21	-411.32	-17.55
	(414.053)	(637.174)	(444.114)
Farm size	-344.12	-404.81	109.07
	(304.598)	(350.962)	(478.830)
Rainfall	6.24	17.08	-34.71**
	(10.627)	(18.339)	(13.392)
2004	3,276.41	-4,787.76	20,336.06**
	(4,104.016)	(3,871.267)	(4,988.371)
2007	25,274.15**	20,024.38**	37,667.76**
	(3,387.459)	(4,957.336)	(4,560.632)
2010	97,487.26**	96,339.16**	86,641.13**
	(8,323.178)	(7,976.746)	(10,315.351)
Observations	4,711	3,242	1,469
R-squared	0.014	0.012	0.185
Number of hhid	1,226	838	388

Robust standard errors in parentheses
** p<0.01, * p<0.05, + p<0.1

The regression estimating the impact of growing hybrid maize on whether or not total household income falls below the national poverty line is shown in Table 8. Average partial effects, and bootstrapped standard errors are presented for the second-stage regression with hybrid seed use and residual from the first-stage regression included as explanatory variables.

The significance of the residual leads us to reject the null hypothesis that growing hybrid maize is exogenous in poverty outcomes. In the control function approach, this is our only test of endogeneity. At the same time, the negative coefficient on growing hybrid maize has strong statistical significance, indicating that adoption reduces the likelihood that household income falls below the national poverty line.

Other coefficients in this regression are of interest. First, education of adults in the household reduces the chances that the household's income will fall below the poverty line. This result is consistent with the broad development literature and the literature on rural Kenya. A more surprising result is the positive relationship of numbers of adults in the household to the probability of being poor. It may be that the size of the household offsets the income-generating effect among these households.

Table 9. Impact of hybrid maize on poverty, control function with CRE

	dy/dx	Bootstrap Std. Err.	z	P>z
Grow hybrid	-0.347	0.359	-3.340	0.001
Residual	0.200	0.372	1.850	0.064
Female head	0.003	0.110	0.100	0.921
Education	-0.015	0.019	-2.830	0.005
Young adults	0.028	0.019	5.040	0.000
Mature adults	0.014	0.019	2.620	0.009
Extension	-0.001	0.006	-0.670	0.500
Tarmac	-0.007	0.014	-1.770	0.077
Land	-0.001	0.009	-0.500	0.616
Rain	0.000	0.000	1.210	0.228
2004	0.001	0.067	0.070	0.943
2007	0.064	0.055	4.010	0.000
2010	0.120	0.097	4.260	0.000

Source: Authors
Coefficients of means of time-varying variables not presented
Bootstrap 50 iterations, n=4843.
Log likelihood= -2482.78
Prob > chi2=0.0000

The closer the household is to the tarmac road, the more likely to be poor. This finding probably reflects the structure of roads and settlement in Kenya. Higher rainfall is associated with more poverty, but the magnitude of this effect is very small. We also note that the impact of 2010 is strong in magnitude and significance.

Diagnostic statistics for fixed effects and CRE models are shown in Table 9. In all regressions, the statistical significance of the F-test statistics confirms the relationships between the instrumental variables and growing hybrid maize. Failure to reject the Hansen-J statistic in most cases also supports the validity of the instrument. The statistic in the regression for the low potential agroecological zone is significant at 7%—casting some doubt on the instruments in this case—although the number of households is much lower in this regression, which may explain weaker results. We reject the hypothesis that hybrid seed use is exogenous in all regressions at less than 1% statistical significance. Thus, through either self-selection bias, or simultaneity, or other underlying sources of correlated errors, growing hybrid maize seed is endogenous in total household income and relative deprivation with respect to total household income.

Table 9. Summary of diagnostic tests for instrumental variables, fixed effects models

Model	F -test	Endogeneity test	Hansen J Statistic
Income			
General	11.67 (0.000)	28.15 (0.000)	0.227 (0.6337)
High potential	3.63 (0.0267)	17.55 (0.000)	0.053 (0.8186)
Low Potential	7.97 (0.0004)	12.031 (0.0005)	0.407 (0.5234)
Inequality			
General	10.33 (0.000)	11.518 (0.0007)	0.013 (0.9092)
High Potential	3.16 (0.0424)	3.442 (0.0635)	1.381 (0.2399)
Low Potential	8.27 (0.0003)	7.178 (0.0074)	3.148 (0.076)
Poverty	2.57 (0.100)	0.200 (0.064)	-

6. Conclusions

Despite more than a half-century of improved maize production in Kenya, the impact of hybrid maize seed on the welfare of smallholders has not been estimated with rigorous quantitative methods. Our objective has been to begin that work. We applied an instrumental, fixed effects model and a CRE model with a control function to data collected from a balanced panel of 1243 households in four waves (2000, 2004, 2007, and 2010). The impact outcomes we considered were income, income inequality, and poverty status. Diagnostic statistics provide evidence of the endogeneity of hybrid use in income, inequality, and poverty outcomes, due either to self-selection bias or simultaneity.

Comparisons of sample statistics confirm that total household income is higher, relative deprivation with respect to income is lower, and poverty status is lower for hybrid growers as compared to non-hybrid growers. The poverty status of maize-growing farm families who do not grow hybrids is substantially higher, in terms of headcount ratios, the mean poverty gap, and mean severity of poverty.

Regression results confirm that growing hybrid maize seed is endogenous in the poverty and income outcomes we have considered, either through self-selection bias, simultaneity, or other underlying factors that cause errors to be correlated between seed choice and income. At the same time, tests of our major hypotheses concerning the positive impact of growing maize seed on income outcomes appear to be robust. In the years covered by Tegemeo survey data (2000 to 2010), data support the hypotheses that growing hybrid maize seed increases total household income, reduces income deprivation relative to other households in the location, and negatively influences the likelihood that household income falls below the national poverty line.

Further research might compare the findings derived from fixed effects estimation to those generated by other estimation procedures, such as propensity score matching, and application of control function approach with correlated random effects to estimate the impacts of the scale of hybrid seed use (kgs per hectare) on income and poverty. In addition, we are interested in the impacts of hybrid seed use on the diversification of income sources as Kenyan agriculture develops. Our objective in this body research is to assemble a robust set of econometric results that will be of use to decision-makers in the Kenyan maize seed industry.

References

Angrist, J.D. 2001. Estimation of limited dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice. *Journal of Business and Economic Statistics* 19 (1): 2-16**.**

Angrist, J. D. and A.B. Krueger. 2001. Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives* 15 (4): 69-85.

Argwings-Kodhek et al. 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.

Ayieko, M. W. and Tschirley, L.T. (2006). Enhancing access and utilization of quality seed for food security in Kenya. Working Paper No. 27/2006, Egerton University, Tegemeo Institute, Nairobi.

Baum, C.F., Schaffer, M.E., Stillman, S. 2010. ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression. http://ideas.repec.org/c/boc/bocode/s425401.html

Byerlee D. and C.K. Eicher. 1997. *Africa's Emerging Maize Revolution*. Boulder, CO: Lynne Rienner Publishers.

Chamberlain, G. 1984. Panel Data. In *Handbook of Econometrics, Vol. 2,* ed. Z. Griliches, and M.D. Intriligator. Amsterdam, North Holland: Elsevier.

De Groote, H., G. Owuor, C. Doss, J. Ouma, L. Muhammad, and K. Danda. 2005. The Maize Green Revolution in Kenya Revisited. Electronic *Journal of Agricultural and Development Economics* 2.1: 32-49. Available online at www.fao.org/es/esa/eJADE

Foster, J., J. Greer and E. Thorbecke. 1984. A class of decomposable poverty measures. *Econometrica*. 2, 81: 761–766.

Gerhart, J. 1975. *The Diffusion of Hybrid Maize in Western Kenya*. Mexico City: CIMMYT.
Handa, S. and J.A. Maluccio (2010). Matching the Gold Standard: Comparing experimental and non-experimental evaluation techniques for a geographically targeted program. *Economic Development and Cultural Change*, 58, 415-447.
Hassan, R. 1998. Maize Technology Development and Transfer. Published jointly by Oxon, UK: Centre for Agricultural Bioscience International (CABI); Mexico, D.F.: CIMMYT; and Nairobi, Kenya: KARI.
Heisey, P.W., M.L. Morris, D. Byerlee, and M. López-Pereira. 1998. Economics of hybrid maize adoption. In M.L. Morris (ed.), *Maize Seed Industries in Developing Countries,* Lynne Rienner and CIMMYT, Boulder and Mexico, DF
Kirimi, L., N. Sitko, T.S. Jayne, F. Karin, M. Muyanga, M. Sheahan, J. Flock, and G. Bor. 2011. A Farm Gate-to-Consumer Value Chain Analysis of Kenya's Maize Marketing System. Working Paper No. 44/2011, Egerton University, Tegemeo Institute, Nairobi.
Mundlak, Y. 1978. On the pooling of time series and cross section data. Econometrica 46: 69-85.
Mwabu, G., W. Mwangi and H. Nyangito.2007. Does Adoption of Improved Maize Varieties Reduce Poverty?Evidence from Laikipia and Suba Districts in Kenya. Productive Sector Division, Kenya Institute for Public Policy Research and Analysis. KIPPRA Discussion Paper No. 7, Nairobi, Kenya.
Schaffer, M.E. 2010. Xtivreg2: Stata module to perform extended IV/2SLS, GMM and AC/HAC/LIML and k-class regression for panel data models. http//ideas.repect.org/c/bod/bocode/s456501.html. Accessed May 5, 2012.
Singh, I., Squire, L., Strauss, J. eds., 1986. Agricultural household models: Extensions, applications, and policy. Baltimore: Johns Hopkins University Press.
Smale, M., and T.S. Jayne. 2010. "Seeds of Success" in Retrospect: Hybrid Maize in Eastern and Southern Africa. In *Successes in African Agriculture: Lessons for the Future,* ed. S. Haggblade and P.B.R. Hazell. Baltimore: Johns Hopkins University Press for the International Food Policy Research Institute.
Smale, M., J. Olwande, and H. De Groote. 2012. Demand for Maize Hybrids and Hybrid Change on Farms in Kenya. Mimeo. Michigan State University and Tegemeo Institute, East Lansing, MI and Nairobi, Kenya.
Smith, R.J., and R.W. Blundell. 1986. An exogeneity test for a simultaneous equation tobit model with an application to the labor supply. Econometrica 50(3): 679-685.
Stark, O. and E. Taylor. 1989. Relative Deprivation and International Migration. *Demography* 26: pp. 1-14.
Suri, T. 2011. Selection and Comparative Advantage in Technology Adoption. *Econometrica* 79.1: 159-209.
Wooldridge, J.W. 2002. Econometric Analysis of Cross Section and Panel Data. Massachusetts Institute of Technology (MIT) Press, Cambridge.
Yitzhaki, S. 1979. Relative Deprivation and the Gini Coefficient. *The Quarterly Journal of Economics* 93: 321-324.

Table A.1. OLS regression predicting farmgate seed-to-grain price ratio

		Coef.	Std. Err.	t	P>t
District					
	12	7.039	4.435	1.590	0.113
	13	-0.992	4.299	-0.230	0.818
	31	8.652	2.714	3.190	0.001
	32	12.463	2.712	4.590	0.000
	33	9.374	2.192	4.280	0.000
	34	9.379	2.332	4.020	0.000
	35	8.814	2.807	3.140	0.002
	41	3.575	2.328	1.540	0.125
	42	5.815	2.559	2.270	0.023
	43	6.209	2.732	2.270	0.023
	51	3.054	1.929	1.580	0.114
	52	2.637	1.934	1.360	0.173
	53	4.118	2.396	1.720	0.086
	61	3.316	2.010	1.650	0.099
	62	3.449	1.900	1.820	0.070
	71	1.076	1.865	0.580	0.564
	72	2.927	1.697	1.720	0.085
	73	2.582	1.766	1.460	0.144
	74	2.994	1.930	1.550	0.121
	75	2.325	1.587	1.470	0.143
	81	1.904	1.729	1.100	0.271
Year					
	2004	0.537	0.280	1.920	0.055
	2007	-3.708	0.309	-12.000	0.000
Agroecological zone					
	2	(omitted)			
	3	-4.385	1.880	-2.330	0.020
	4	-0.829	1.192	-0.700	0.487
	5	-0.742	1.094	-0.680	0.498
	6	-3.100	1.715	-1.810	0.071
	7	-0.273	0.637	-0.430	0.669
	8	(omitted)			
Km to fertilizer supplier		-0.069	0.039	-1.750	0.080
Crop share of income		-0.944	0.556	-1.700	0.090
Own transport equipment		-0.495	0.276	-1.800	0.073
Constant		14.167	2.427	5.840	0.000

Prob > F =0.0000

R-squared=0.2950

Figure A.1. Histogram of total household income, with normal curve

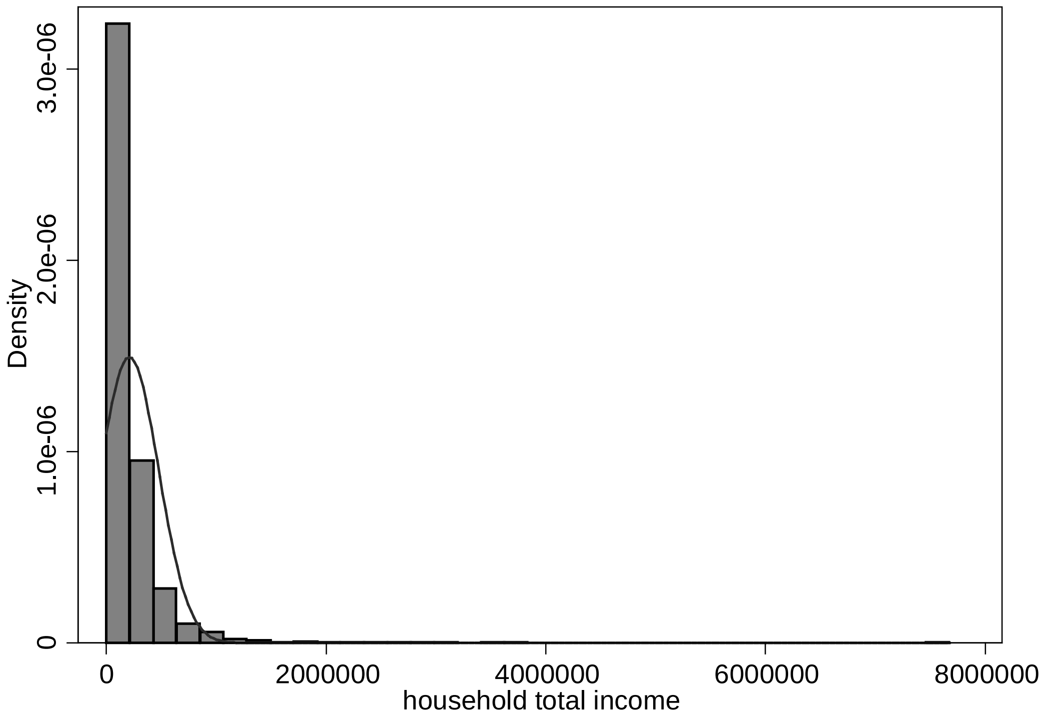

Figure A.2. Histogram of natural logarithm of total household income, with normal curve

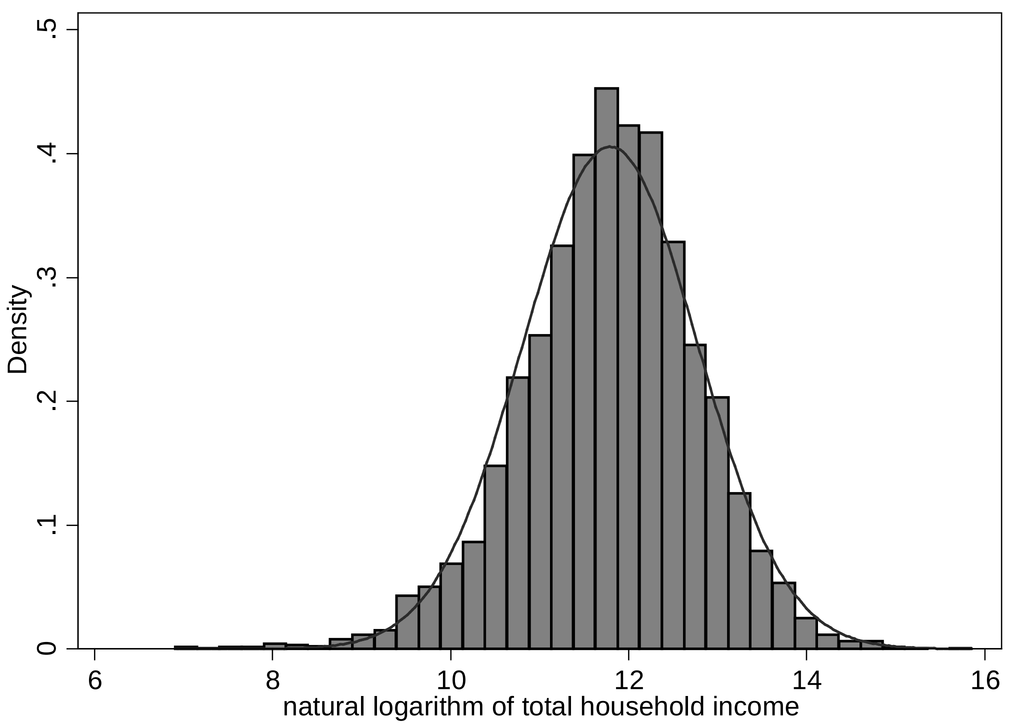

Table A.2. Preliminary pooled, random effects, fixed effects regressions

	(1) Pooled	(2) onlyre	(3) onlyfe	(4) PooledIV	(5) FEIV2
Grow hybrid	0.47**	0.30**	0.07*	0.85**	2.24**
	(0.026)	(0.028)	(0.032)	(0.221)	(0.602)
Female head	-0.32**	-0.28**	-0.12*	-0.27**	-0.01
	(0.029)	(0.035)	(0.053)	(0.037)	(0.078)
Education	0.09**	0.08**	0.02**	0.08**	0.02*
	(0.004)	(0.005)	(0.007)	(0.007)	(0.010)
Young adults	0.05**	0.05**	0.04**	0.05**	0.03*
	(0.007)	(0.007)	(0.008)	(0.007)	(0.013)
Mature adults	0.13**	0.12**	0.10**	0.13**	0.10**
	(0.009)	(0.009)	(0.010)	(0.010)	(0.016)
Extension	-0.01*	-0.00	0.00+	-0.00	0.00
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
Tarmac	0.00	-0.00	-0.00	0.00+	-0.00
	(0.002)	(0.002)	(0.004)	(0.002)	(0.005)
Land	0.02**	0.02**	0.00+	0.02**	0.01*
	(0.001)	(0.002)	(0.003)	(0.002)	(0.003)
Rain	-0.00**	-0.00**	-0.00	-0.00**	-0.00*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2004	0.20**	0.17**	0.11**	0.23**	0.28**
	(0.032)	(0.026)	(0.027)	(0.038)	(0.061)
2007	0.28**	0.28**	0.27**	0.26**	0.15**
	(0.030)	(0.025)	(0.025)	(0.033)	(0.050)
2010	0.39**	0.44**	0.51**	0.32**	0.12
	(0.032)	(0.028)	(0.030)	(0.053)	(0.116)
Constant	10.39**	10.58**	11.08**	10.23**	
	(0.053)	(0.062)	(0.079)	(0.104)	
Observations	4,895	4,895	4,895	4,878	4,885
R-squared	0.414		0.154	0.387	-1.011
Number of hhid		1,243	1,243		1,239

Robust standard errors in parentheses
** p<0.01, * p<0.05, + p<0.1

Table A.3. Impact of hybrid maize on income, high and low potential maize-growing areas

	(1) Income High potential maize	(1) Income Low potential maize
Grow hybrid	2.52**	2.21**
	(0.887)	(0.795)
Female head	-0.41**	0.09
	(0.130)	(0.097)
Education	0.03	0.02
	(0.018)	(0.012)
Children	0.04+	0.03+
	(0.019)	(0.017)
Young adults	0.10**	0.10**
	(0.027)	(0.019)
Mature adults	-0.00	0.00
	(0.005)	(0.005)
Elderly	-0.00	-0.00
	(0.010)	(0.006)
Farm size	-0.00	0.02**
	(0.003)	(0.006)
Rainfall	0.00	-0.00*
	(0.000)	(0.000)
2004	-0.03	0.29**
	(0.133)	(0.084)
2007	0.16*	0.15*
	(0.077)	(0.065)
2010	0.30**	0.11
	(0.113)	(0.163)
Observations	1,302	3,583
R-squared	-0.584	-1.151
Number of hhid	331	908

Robust standard errors in parentheses
** p<0.01, * p<0.05, + p<0.1

Source: Authors. Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1