Selection and Comparative Advantage in Technology Adoption

Tavneet Suri[1]

Preliminary

Please Do Not Circulate

October, 2005

Abstract

This paper examines a well known empirical puzzle in the literature on technology adoption: there is a strong pattern of non-increasing adoption of new technologies over time, even though these technologies are known to have high returns. I study the use of hybrid maize and fertilizers in Kenya, where there are persistent cross sectional differences in aggregate adoption rates. I approach the question from an extremely different perspective. The yield returns to adopting these technologies may vary across a given sample of farmers, such that high average returns to the technologies overstate the returns for the marginal farmer. I therefore analyze adoption decisions within a model that allows for household specific heterogeneity in returns and hence selection via both absolute and comparative advantage differences across households. Despite high average returns, I find low marginal returns to the use of these technologies, pointing to the adoption decisions being rational, given the costs and constraints households face.

---

[1] Department of Economics, Yale University. Email: tavneet.suri@yale.edu. I am extremely grateful to Michael Boozer for countless hours of discussion on this project. I would also like to thank the rest of my committee, Gustav Ranis, Paul Schultz and Christopher Udry, as well as Robert Evenson, Fabian Lange and seminar audiences at Yale and NEUDC for comments.

A special thanks to Thom Jayne and Margaret Beaver at Michigan State University for help with the 1997-2002 rounds of the data. I am sincerely grateful to the Tegemeo Institute and Thom Jayne for including me in the 2004 survey and especially to Tegemeo for their hospitality during the field work in Kenya.

I would like to thank James Nyoro, Director of Tegemeo, as well as Tegemeo Research Fellows Miltone Ayieko, Joshua Ariga, Paul Gamba and Milu Muyanga, Senior Research Assistant Frances Karin, and Research Assistants Bridget Ochieng, Mary Bundi, Raphael Gitau, Sam Mburu, Mercy Mutua and Daniel Kariuki and all the field teams for their impressive work through all the rounds of the survey, especially 2004. An additional thanks to Margaret Beaver and Daniel Kariuki for spending immense energy ensuring the data was squeaky clean.

The data come from the Tegemeo Agricultural Monitoring and Policy Analysis (TAMPA) Project, between Tegemeo Institute at Egerton University, Kenya and Michigan State University, funded by USAID, April 1996 - June 2005. Any rainfall data used come from the Climate Prediction Center and are part of the USAID/FEWS project - thanks to Tim Love for his repeated help with this data.

1 Introduction

Food security is currently at the forefront of the policy agenda across many Sub-Saharan economies. Governments and policy makers are trying to understand the low yields of agricultural staples across various parts of Africa and looking for ways to enhance food production and incomes. Agricultural yields have fallen in recent decades across many Sub-Saharan countries, despite the widespread availability of technologies that increase yields. For example, Table 1 shows the falling yields of staple crops in Kenya, compared with increasing yields in India and Mexico. These are worrisome numbers given that improved agricultural technologies, like hybrid seed and fertilizers, have been available and marketed since the 1960's in Kenya. Promoting high yielding agricultural technologies is one way to improve yields. This paper examines the role of agricultural technologies in Kenya. What agricultural technologies are households using and governments promoting to increase or at least sustain crop yields? Are these technologies actually increasing yields? If so, how well are they adopted/used? And, if they are not widely used, why not? Understanding how households decide what technologies to adopt and why is crucial for economies where risk and incomplete factor markets are potentially important. this paper looks at the use/adoption of the main agricultural technologies in Kenya, hybrid maize and fertilizers.

It is clear from field trials at experiment stations across Kenya that hybrid maize and fertilizers can increase yields of maize dramatically (see Kenya Agricultural Research Institute, KARI (1993)). The Fertilizer Use Recommendations Project (FURP) conducted in Kenya in the early 1990's also shows high returns to these technologies in the field trials, even though the actual yields that farmers get from hybrid seed and increased fertilizer use tend to be lower than those on experiment stations. However, aggregate adoption rates of hybrid maize and fertilizer remain far below 100% across Kenya, with a large variation in individual adoption status in my sample from year to year. This has posed an empirical puzzle: why are there patterns of stagnating aggregate adoption rates when returns to the technologies are high? What could explain this puzzle? The literature has put forth a variety of explanations that cover models of learning, credit constraints and lack of commitment devices (see Duflo, Kremer and Robinson (2003)).

I examine the role of heterogeneity in returns to adoption. The yield returns to adopting a technology may vary across any given sample of farmers, with those who would profit having already adopted. This would imply that high average returns often overstate the returns for the farmer at the margin of adopting. Knowledge of average returns is not enough to understand the adoption processes across a sample of farmers. I therefore study the adoption of agricultural technologies in Kenya within a framework of heterogeneous returns where a farmer's expected returns are allowed to be correlated with his decision to adopt the technology. Technology adoption is modeled as a selection process where yields may be heterogeneous and individual specific benefits to the technology may drive adoption decisions in each period. This is an extremely general framework for the analysis of adoption decisions.

I use micro panel data for most of rural Kenya to first document the adoption patterns of households over the period 1996 to 2004, for both hybrid maize and fertilizers. Next, as a benchmark, I estimate models with homogeneous returns. I then present empirical tests of the validity of such a model, including some suggestive evidence for selection and heterogeneity in returns. In general, models of homogeneous returns are rejected. I then describe my preferred model of heterogeneous returns and selection and outline how to estimate it. Once I have estimates of the parameters of the heterogeneous returns model, I can derive (under certain assumptions) the distribution of heterogeneity. I then analyze this distribution further to understand better the policy implications of my results. Estimates of the parameters of the heterogeneous returns model can also be interpreted in a framework that allows me to distinguish absolute and comparative advantage in the use of technology. This involves relating my model to a standard cross sectional selection model to clarify what are the underlying assumptions of my framework. This allows me to compare my estimates of comparative advantage to the standard average treatment effects, treatment on the treated, marginal treatment and the local average treatment effects.

I find strong evidence of heterogeneity and selection of farmers into technology adoption. There is evidence of farmers responding to their expected benefits to select into high yielding varieties. What I define as comparative advantage plays an important role in determining the adoption of technologies by farmers. This points to the need for investment in developing technologies with improved returns for farmers who are currently below the margin of adoption. More importantly, I find that even though these agricultural technologies have high average returns, the marginal farmer has low returns and switches easily in and out of adoption when subjected to yield shocks. This illustrates how adoption decisions are rational and not based on some irrationality in household decision making as has been suggested in some of the recent work on this topic.

The rest of this paper is structured as follows. In Section 2, I outline some of the relevant literature, focusing on empirical studies. Section 3 describes the data and some of the history of maize cultivation in Kenya. Section 4 outlines panel data models of the adoption of technologies for both the homogeneous and heterogeneous returns cases. Section 5 discusses both the estimation and tests of the homogeneous returns model. In addition to presenting the benchmark results, I discuss the evidence from tests of the household fixed effects model. I also outline some suggestive evidence for selection and heterogeneity evidence in returns. Section 6 describes estimation of the heterogeneous returns model and the associated distribution of returns. In Section 7, I present a simple selection model along with estimators of various features of the distribution of returns that are common in the current empirical literature. This selection model allows me to interpret my estimates in terms of comparative and absolute advantage. Finally, Section 8 concludes with some of the implications for policy, as well as a discussion of why a framework such as mine is needed for a country like Kenya, which has widely varying agronomic conditions.

2 Literature Review

I briefly summarize some of the empirical literature on technology adoption[2] in developing countries, with a focus on the studies that have been conducted in Sub-Saharan Africa. I split this review into four main areas, first discussing studies that relate to heterogeneity, then looking at research on learning externalities and credit constraints and ending by discussing the recent experimental research.

The seminal empirical paper on new agricultural technology adoption is Griliches (1957)[3]. He looks at the heterogeneity across local conditions in the adoption speeds of hybrid corn in the Midwestern US. He emphasizes the role of economic factors like expected profits and scale in determining the variation in diffusion rates. He also notes how the speed of adoption across geographical space depends on the suppliers of the technology and when the seed was adapted to local conditions. Other researchers have looked at heterogeneity along other dimensions and what sorts of heterogeneity drive the decisions of households to adopt. This covers quite a range of papers, from Schultz (1963, 1964) and for Ethiopia Weir and Knight (2000), both of whom emphasize the role of education, to the various CIMMYT (The International Wheat and Maize Improvement Center) studies[4] that collect data on what underlies adoption decisions across different parts of Kenya. For example, Makokha et al (2001) look at the determinants of fertilizer and manure use in Kiambu district in Kenya, focusing on measuring soil quality and showing how farmers' perceptions of soil quality are reasonably accurate. The main (self reported) constraints to using fertilizers were high labor costs, high prices of the inputs, unavailability of demanded packages and untimely delivery. Ouma et al (2002) look at the adoption of improved seed and fertilizers in Embu District in Kenya. They find that gender, agroclimatic zone, manure use, hiring of labor and extension services are all significant in determining adoption. Similarly, Wekesa et al (2003) look at the adoption of the Coast Composite, Pwani 1 and Pwani 4 hybrids and fertilizers in the Coastal Lowlands of Kenya where the non availability and high cost of seed, unfavorable climatic conditions, perceptions of sufficient soil fertility, and lack of money/credit are cited as reasons for low use.

Much of the recent literature on technology adoption has focused on the learning externality, described well by Besley and Case (1993), and which relates to the literature on social interactions. Foster and Rosenzweig (1995) look at the adoption of HYV's in post Green Revolution India, allowing for learning by doing, learning from others, and costly experimentation. They

[2]The literature on technology adoption is too vast to review here, excellent reviews can be found in Sunding and Zilberman (2001), Rogers (1995), Sanders, Shapiro and Ramaswamy (1996) and Feder, Just and Zilberman (1985). There is also a large theoretical literature, for example, Banerjee (1992), Just and Zilberman (1983) and Besley and Case (1993). Hall (2004) reviews well the social, economic and institutional determinants of diffusion rates in other fields. I do not include studies that focus on livestock or land management practices (Mugo et al (2000) and Ovuka and Ekbom (1999)), those that look at agricultural extension (Evenson and Mwabu (1998)), or those that focus on property rights (Place and Swallow (2000)).

[3]Also see David (2003).

[4]See Doss (2003) and De Groote et al (2002) for a review of all the CIMMYT micro surveys in Kenya. Also, see http://www.cimmyt.org/research/economics/map/impact%5Fstudies/impstudea%5Flist/ for all the impact studies done by CIMMYT in East Africa.

find that farmers with more experienced neighbors are more profitable than those without. Munshi (2003) looks at the same question and finds that the impacts are heterogenous: wheat growers respond strongly to their neighbors' experiences while rice farmers experiment. He finds greater variations in yields in rice growing areas and notes that rice HYV's, unlike those for wheat, tend to be sensitive to soil characteristics and managerial inputs, which are difficult to observe. Conley and Udry (2003) study the adoption of fertilizers in the small-scale pineapple industry in Ghana. They have a unique dataset, as they directly collect information on farmers' sources of information. They find evidence of learning, not only from own experiences, but also within information neighborhoods. Bandiera and Rasul (2003) look at decisions to plant sunflower in the Zambezia province of Mozambique. They find that adoption decisions are correlated within networks of family and friends and that this effect is stronger for disadvantaged farmers. Moser and Barrett (2003) look at a high yielding low external input rice production method in Madagascar, analyzing decisions to adopt, expand and disadopt. They find seasonal liquidity constraints and learning effects from extension agents and other households to be important.

A hypothesis that is often raised in the literature is that credit constraints explain adoption patterns. For example, Croppenstedt, Demeke and Meschi (2003) estimate a double hurdle fertilizer adoption model for Ethiopia, using self reported information on why farmers did not purchase fertilizer. They use self-reported information to determine a sample separation into farmers who want to use fertilizers but lack credit, versus those who find it unprofitable. They find that credit is a major supply side constraint to adoption. Most of the CIMMYT studies also cover self reported credit constraints, for example, Salasya et al (1998) look at the role of credit in adoption decisions Western Kenya.

The final strand of literature on Kenya I describe is experimental. There are several impact assessment studies[5] and field trials at experiment stations in Kenya, most of which show large increases in yields from using hybrid maize and fertilizers. Other than the work done by the Kenya Agricultural Research Institute on experiment stations, one of the early experimental studies was the Fertilizer Use Recommendations Project (FURP), conducted across 70 sites in the early 1990's in conjunction with the Kenya Maize Database Project (MDBP). FURP recorded yields about half of those on experimental stations (KARI (1993)). The focus of these experiments was to understand optimal rates of fertilizer use in comparison to survey data on actual use from the MDBP (see Corbett (1998)). Hassan et al. (1998) use these data and find that both adoption and varietal turnover rates are much higher (and diffusion faster) in high potential areas. They blame poor extension services, bad infrastructure and poor seed distribution for the low adoption rates in the marginal areas. Hassan, Murithi and Kamau

[5] There are studies that use survey data to understand the welfare impacts of improved technology use. For example, Karanja, Renkow and Crawford (2003) look at the differential impacts across high potential and marginal areas in Kenya, in terms of both efficiency and distribution. They find that adoption of technologies in high potential areas, relative to those in marginal areas, have large aggregate gains at the expense of poor distributional effects. Sserunkuuma (2002) looks at a partial equilibrium model of adoption of improved maize and fertilizers in Uganda. He uses price elasticities of demand and supply from secondary sources to estimate large consumer gains and large producer losses of a shift in supply due to the adoption of hybrid maize.

(1998) combine the data from the surveys and trials and find that farmers apply less fertilizer than is optimal, leading to an estimated 30% gap between current and optimal yields.

A more recent example is De Groote et al (2003) who look at an ex-ante impact assessment of the Insect Resistant Maize for Africa (IRMA) project that develops GM maize varieties that are more resistant to stem borers. They estimate the surplus from a shift in supply due to the decreased crop loss (measured experimentally) as a result of introducing a maize variety that is more resistant to stem borers. Estimated crop losses amounted to 13.5% with an estimated value of $80 million. The results imply high returns to such genetic technologies[6].

Similarly, Duflo, Kremer and Robinson (2003) run controlled experiments in the field to measure returns from hybrid seed and fertilizer. They find that the average rate of return for investing in top-dressing fertilizers is between 28% and 134% for an eight month investment. They study diffusion and find a significant negative impact on neighbors of the program. They find that farmers learn via demo trials, distributed kits, but not through announcements of government endorsement. They also started the Savings and Fertilizer Initiative (SAFI) as a commitment device for farmers. They find that farmers take up this program when it is offered at harvest time, but not later, pointing to these farmers being hyperbolic discounters.

3 Survey Data and Maize Cultivation in Kenya

I study maize as it is the main staple in Kenya[7], accounting for approximately 3.7 million acres of cropped land, with main season maize production ranging between 2.3 and 2.7 million MT, of which 75% is through small scale farming. Average maize yields are on the order of 0.8 MT per acre, although there is considerable cross sectional diversity. I use data from the Tegemeo Agricultural Monitoring and Policy Analysis (TAMPA) Project (April 1996-June 2005) between Tegemeo Institute at Egerton University, Kenya and Michigan State University, funded by the US Agency for International Development, Kenya[8]. This is a household level panel survey, representative of rural maize-growing areas, aimed at monitoring smallholder production patterns, consumption and incomes, as well as identifying policy agendas for farmers. The survey is geographically diverse, covering a large part of rural Kenya. It collects consumption, income, crop, production and credit data as well as village/community level information. Figure 1 shows a map of Kenya with the location of the villages covered.

Different modules of data were collected in different years of the survey, with a common core set. I have data for 1997, 1998, 2000, 2002 and 2004. The 1997, 2000 and 2004 surveys are similar, containing detailed agricultural input and output data (plus retrospective data for

[6]See http://www.cimmyt.org/ABC/InvestIn-InsectResist/htm/InvestIn-InsectResist.htm and Smale and De Groote (2003) for more information on the CIMMYT IRMA and GM projects in Kenya. An additional possible project still in the pipeline aims to introduce a new GM Vitamin A maize variety.

[7]McCann (2005) describes the fascinating history of maize in Africa, pointing out how, compared to other grains, maize not only gives more food per unit of land and labor, but also has the largest set of alternative uses. Nyameino, Kagira and Njukia (2003) cite that about 90% of Kenya's population depends on maize for income generation.

[8]See http://www.aec.msu.edu/agecon/fs2/kenya/index.htm for more information on the TAMPA project.

1995-1996), household consumption (not complete), income, demographics (age-sex composition, education and some health measures for all the individuals in a household), infrastructure, location, and some basic credit information. The panel sample covers about 1400 households, with an additional 800 households in the 2004 sample. The 1998 survey is a sub-sample of about 612 households, collecting similar information plus some quantitative data on fertilizer use and timing, and some soil quality measures. Kenya was strongly affected by El Nino in 1998 and so the 1998 sample is very different from the other years. The 2002 survey was a short proxy survey, but it includes data on hybrid maize use, including quantities and costs of the exact hybrid variety used.

To motivate my research questions, I outline some trends in my data over the period 1996-2004 (as the data permits) on the use of hybrid maize and fertilizers. Figure 2 shows the trends in adoption for the two seasons, defined here as the fraction of maize seed planted that is hybrid. Figure 3 looks at main season adoption patterns by province over the same period. Both Figures 2 and 3 illustrate the relative stability in adoption patterns over time with strong, persistent cross sectional differences. An identical pattern is evident if I look across wealth/asset or acreage quintiles. Similar trends hold up when looking at the use of inorganic fertilizers during the main season, shown in Figures 4a and 4b. Figure 4a shows the trends across provinces in the fraction of households that use inorganic fertilizers for maize production and Figure 4b shows the total value (in constant Kenyan shillings) of inorganic fertilizers used, by province. There is a lot more variation over time in the total value of inorganic fertilizers used, but both figures illustrate persistent cross sectional differences in use. Finally, for an idea of the more general patterns in my data, Figures 5a and 5b show main season yields of maize and the acreage planted to maize, respectively, over the same period. Of course, yields are not stable, with the sharp drop in yields around 1997/1998 the result of El Nino floods. Trends in acreage planted to maize actually stay rather stable over the period, even in the face of such large weather shocks.

Maize policy in Kenya has an interesting history over this period, Smale and Jayne (2003) provide an excellent review. In terms of technology releases, both hybrid seed and fertilizers have been around since the 1960's. In fact, more than twenty modern maize varieties of seed have been released by the government since 1955 (Ouma et al. (2002)). The period from 1965 to 1980 was impressive in terms of yields, hybrid variety 611 diffused in Western Kenya "at rates as fast as or faster than among farmers in the US corn belt during the 1930's-1940's" (Gerhart (1975)). Hybrid seed proved to have large yield advantages, especially when supplemented with the use of fertilizers. Smale and Jayne (2003) and Karanja (1996) attribute the impressive performance of this period to good germplasm, effective research, strong linkages between researchers and extension services, good seed distribution/enterprise (via Kenya Seed Company, still the predominant source of hybrid) and coordinated marketing of inputs and outputs. However, this quickly changed during the 1990's as the earlier policies of large subsidies, strong price supports, pan-territorial seed/output pricing and restrictions on cross district trade, resulted in large fiscal deficits[9]. Reform of the cereal sector began in 1988, followed by a wide liberalization in 1994[10], but some of the previous control policies of the government continued, probably to benefit politically important areas at the expense of marginal farmers (Smale and Jayne (2003)).

The government recommendations for maize cultivation deal with the type of seed (i.e. exactly which hybrid/improved variety seed) and the types and quantities of fertilizer to be applied. These recommendations vary by region, see Appendix Table A1, which also shows the release dates of hybrid varieties. Most of the agriculture is rain fed, with large variations in rainfall that make input use more risky and complicate plant breeding (see Hassan, Corbett and Njoroge (1998), Hassan, Onyango and Rutto (1998) and Mills, Hassan and Mwangi (1998)). However, the commonly held wisdom is that the further releases in hybrid seed in Kenya since the first releases have not shown big improvements in yields[11], unlike in India. This fact, along with the increases in agricultural intensification and shifting of maize to more marginal areas, is often blamed for stagnating yields.

Table 2a shows the summary statistics for my sample of households over the three periods of data I use (1997, 2000 and 2004), including some of the covariates I use. There is an increase in average yields over the period. There are also some interesting trends in fertilizer use. There are about 26 different types of fertilizers reported as being used over these three periods. Table 2a only shows three of the main ones (di-ammonium phosphate (DAP), calcium ammonium nitrate (CAN) and nitrogen phosphorus potassium (NPK)) and one with unusual trends (mono-ammonium phosphate (MAP)). Table 2b breaks out some of these variables by hybrid/non-hybrid sector for the three periods of data. Yields are lower across the board in the non-hybrid sector (the p-values on the t-tests are 0.000 in each period). A lot of the other variables look quite similar (p-values on the t-tests are often above 5%). However, fertilizer use and main season rainfall are significantly different across hybrid and non-hybrid sectors, though the latter may just be an indication of differential use of hybrid by province.

[9] Kenya National Cereals and Produce Board, the marketing board supporting these policies, managed to accrue immense losses in the 1980's, on the order of 5% of the country's GDP (Smale and Jayne (2003)).

[10] There are numerous studies of the cereal sector reforms and liberalization, see Jayne et al (2001). Karanja, Jayne and Strasberg (1998) look at the productivity impacts and Jayne, Myers and Nyoro (2005) at the effect of the National Cereals and Produce Board on maize prices over 1990-2004. Hassan, Mekuria and Mwangi (2001) show the five fold increase in private seed companies between 1992 and 1996, also documented by Kamau (2002) who points out important legislative and regulatory constraints during this time. Finally, Nyoro, Kiiru and Jayne (1999) look at the evolution of different types of maize traders post-liberalization and Wanzala et al (2001) describe in detail how the private sector has taken over the supply of fertilizers.

[11] Karanja (1996) states that "newly released varieties in 1989 had smaller yield advantages over their predecessors than the previously released ones... research yields were exhibiting a "plateau effect"". In numbers, KSII was followed by H611 (40% yield advantage), then H622 (16%) and then H611C (12%). Another example he gives is H626, which has a 1% yield advantage over H625, even though it was released eight years later.

4 Panel Models of Technology Adoption

4.1 Homogeneous Returns Model

The first research question that needs to be answered is what the returns to these technologies are for farmers. This usually involves estimating production or profit functions. Given my data is non-experimental, the choice to plant hybrid maize and/or use fertilizers can not be presumed to be exogenous. I first look at the benchmark homogeneous returns model, i.e. a household fixed effects model. This is the common approach to estimating farm yield or profit functions when panel data on farms is available. This model is motivated by a very specific form of heterogeneity, in particular fixed unobservables, such as farm management (see Mundlak (1961)), productivity of farmers, or soil quality (see Conley and Udry (2003)), among others. It is hoped that allowing for individual fixed effects can account for this heterogeneity.

The general model[12] is in the form of a switching regression framework where farmers *expected* yields in each sector are given by

$$y_{it}^N = \delta_t^N + \alpha_i + X_{it}'\gamma^N \tag{1}$$

$$y_{it}^H = \delta_t^H + \alpha_i + \beta + X_{it}'\gamma^H \tag{2}$$

where N and H represent the non-hybrid and hybrid sectors respectively so that y_{it}^N and y_{it}^H are the corresponding sector yields. The δ_t's are a set of time effects that are allowed to vary by sector. The α_i are household specific unobservables (fixed effects) and the X_{it} are covariates/controls. Here β represents the return in terms of yield to planting hybrid, conditional on covariates. Throughout, including the discussion of the model, for clarity of exposition I use a single technology (hybrid maize). I then adapt my model to account for the joint decision to use hybrid maize and inorganic fertilizer. The outcome variable y_{it} can be either a measure of yield (output per acre) in a production function approach, or a measure of profits in the standard adoption decision approach. When output prices are the same for the hybrid and non-hybrid maize output, these two outcomes differ only by (real) input costs. However, since the other inputs are often a choice variable(s), I will often estimate yield functions, even though strictly speaking the adoption decision will rely on comparison of the hybrid and non-hybrid profits.

In the model given by equations (1) and (2), the gain to growing hybrid maize is

$$B_{it} = \beta + (\delta_t^H - \delta_t^N) + X_{it}'(\gamma^H - \gamma^N) \tag{3}$$

[12]The simplest fixed effects model in the dummy endogenous variable/switching regressions model of expected yields would be

$$y_{it}^N = \delta_t + \alpha_i + X_{it}'\gamma$$
$$y_{it}^H = \delta_t + \alpha_i + \beta + X_{it}'\gamma$$

The causal gain to using hybrid maize in this framework is $y_{it}^H - y_{it}^N = \beta$, which is a constant, implying the mean, median and marginal effects are all constrained to be exactly equal.

The gain from hybrid maize can therefore be time varying due to $(\delta_t^H - \delta_t^N)$ and it can also vary across households as long as it is a function of the observable X's, but the unobservable, α_i, is restricted to affect yields identically in either sector.

I can then pool these sector-specific yields in a generalized expected yield equation of the form

$$y_{it} = h_{it} y_{it}^H + (1 - h_{it}) y_{it}^N \tag{4}$$

where h_{it} is a dummy variable for hybrid use by farmer i in period t. In principle, hybrid use could be a continuous variable as farmers plant quantities of hybrid seed. However, in my sample very few farmers actually plant both hybrid and traditional varieties in a given season. At the household level, therefore, hybrid use is essentially binary.

Substituting in equations (1) and (2) gives

$$y_{it} = \delta_t^N + \pi_t h_{it} + \alpha_i + \beta h_{it} + X_{it}' \gamma^N + X_{it}'(\gamma^H - \gamma^N) h_{it} + \varepsilon_{it} \tag{5}$$

where $\pi_t \equiv (\delta_t^H - \delta_t^N)$ and ε_{it} is the unanticipated component of yields and hence is mean zero.

This household fixed effects framework, however, is rather restrictive in the assumptions it imposes. For one, the household specific effect α_i can play a role in the adoption decision underlying h_{it}, but the gain from using hybrid (conditional on covariates), β, cannot explain the cross sectional variation in adoption decisions as it is constrained to be constant for all households. Apart from the permanent component in the outcome equation, α_i, the adoption decision cannot depend on *observed outcomes* except under restrictive assumptions on the transitory component ε_{it} (see Heckman and Robb (1985) and Ashenfelter and Card (1985)). Such assumptions can be motivated by myopia or ignorance of the potential gains from planting hybrid, but it is restrictive to impose this, especially in my scenario[13]. In addition, as emphasized by Card (1998), the fixed effects/homogeneous coefficients model requires that the selection bias for a given characteristic, such as farmer experience, soil quality, etc., must be of the same *sign* for all individuals. It does not allow, for example, selection into hybrid to be positive for people with low education, say, but negative for those with high education. The Roy (1951) model of choice between two activities is relevant here, even though it relies on comparisons of wages, or net benefits of jobs in various sectors. The conceptual framework it provides can easily be applied to productivities (as suggested by Mandelbrot (1962)) with modifications in the 'pricing' equations of skills across sectors. I do not introduce prices, but instead focus on heterogeneous production technologies across sectors.

The household fixed effects framework implies a restrictive set of assumptions on the adoption of seed varieties and the comparison of the adopters and non-adopters. It implies that the average and marginal returns to hybrid maize must be equal (of course, conditional on the X's), which is unlikely. The α_i allows some farmers to, say, be more productive overall (higher α_i), but the

[13] Hybrid maize was introduced in the 1960's, with widespread use of extension services to promote the technology. See Evenson and Mwabu (1998).

difference β across the hybrid and non-hybrid sectors is the same for all farmers. For various agronomic and economic factors, such as the slow spread of information and the credit constraints that are often alluded to in the literature, it is reasonable to allow for a distribution of returns to the technology that relies on both observed and unobserved factors. The homogeneous returns model may therefore not be appropriate for the question at hand. I will explicitly test this model later. For now, I turn to a description of a model where I relax the assumption of a constant β.

4.2 Heterogeneous Returns Model

The main contribution of this paper is to relax the assumption of a constant (across households) return to hybrid maize. I draw on recent empirical studies in labor economics that allow for alternatives to the single-threshold (or "index") model of fixed effects[14], and use a framework that allows for comparative advantage in the adoption decision. Advances in techniques have been made in the context of experimental data (for example, Heckman, Smith and Clements (1997)) and in cross sectional data where the covariate of interest is a stock variable like schooling. I use the panel nature of my data and build upon the approaches of Lemieux (1993, 1998) and Card (1996). To estimate this model, I generalize the multivariate regression/minimum distance approach of Chamberlain (1984) to allow for heterogeneous coefficients (which I refer to below as returns) that may be correlated with the adoption decision(s) in an arbitrary way.

Expected yields in each sector are now allowed to be affected by sector specific unobservables as follows:

$$y_{it}^N = \delta_t^N + \theta_i^N + X_{it}'\gamma^N \tag{6}$$

$$y_{it}^H = \delta_t^H + \theta_i^H + \beta + X_{it}'\gamma^H \tag{7}$$

Now, the gain from hybrid is a function of both observed and *unobserved* household characteristics:

$$B_{it} = \beta + (\delta_t^H - \delta_t^N) + (\theta_i^H - \theta_i^N) + X_{it}'(\gamma^H - \gamma^N) \tag{8}$$

Following Heckman and Honore (1991), Lemieux (1998) and others, I can define linear projections of the θ_i^H and the θ_i^N on $(\theta_i^H - \theta_i^N)$ as follows

$$\theta_i^H = b_H(\theta_i^H - \theta_i^N) + \tau_i \tag{9}$$

$$\theta_i^N = b_N(\theta_i^H - \theta_i^N) + \tau_i \tag{10}$$

where, by construction, the projection coefficients are $b_H = (\sigma_H^2 - \sigma_{HN})/(\sigma_H^2 + \sigma_N^2 - 2\sigma_{HN})$, $b_N = (\sigma_{HN} - \sigma_N^2)/(\sigma_H^2 + \sigma_N^2 - 2\sigma_{HN})$ and $\sigma_{HN} \equiv cov(\theta_i^H, \theta_i^N)$, $\sigma_H^2 \equiv Var(\theta_i^H)$, and $\sigma_N^2 \equiv Var(\theta_i^N)$[15]. In addition, τ_i (what I will refer to as absolute advantage) is, by construction,

[14]The fixed effects model described in the previous section can be thought of in terms of a selection model, whereby all the individuals who have an α_i greater than some value choose to adopt hybrid and the rest choose not to. This is what I mean by a single-threshold or index model: only one index guides the underlying selection process.

[15]Note that the τ_i's in equations (9) and (10) are the same. Subtracting equation (10) from equation (9)

orthogonal to what I call the comparative advantage gain in growing hybrid, $(\theta_i^H - \theta_i^N)$. I define θ_i to be

$$\theta_i \equiv b_N(\theta_i^H - \theta_i^N) \tag{11}$$

This is just a redefinition of the sector specific unobservables θ_i^H and θ_i^N. This identity illustrates how θ_i in my model is actually a description of the relationship between the unobservables in the two sectors and is hence my measure of comparative advantage. Allowing $\psi \equiv b_H/b_N$, equations (9) and (10) become

$$\theta_i^H = \psi\theta_i + \tau_i \tag{12}$$

$$\theta_i^N = \theta_i + \tau_i \tag{13}$$

The gain from using hybrid can now be expressed as

$$B_{it} = \beta + (\delta_t^H - \delta_t^N) + \phi\theta_i + X_{it}'(\gamma^H - \gamma^N) \tag{14}$$

where $\phi \equiv (\psi - 1)$ is the coefficient on the household specific comparative advantage component θ_i. Using the same generalized yield equation as earlier, I can substitute in equations (6) and (7) and use (12) and (13) to derive a specification for yields:

$$y_{it} = \delta_t^N + \pi_t h_{it} + \alpha_i + \beta h_{it} + X_{it}'\gamma^N + \phi\theta_i h_{it} + X_{it}'(\gamma^H - \gamma^N)h_{it} + \varepsilon_{it} \tag{15}$$

where $\pi_t \equiv (\delta_t^H - \delta_t^N)$, $\alpha_i \equiv \theta_i + \tau_i$ and ε_{it} is an unanticipated component of yields[16]. The coefficient on h_{it}, $\phi\theta_i$, depends on the unobserved household-specific effect θ_i, implying a random coefficient model. The problem here is that the coefficient on h_{it}, $\phi\theta_i$, is correlated with decisions to adopt. This poses a difficult econometric problem.

If we think of this in terms of the generalized Roy model framework, the household specific gain B_{it} is allowed to enter the latent index determining sector choice, h_{it}. So, the coefficient $\phi\theta_i$ is generally correlated with the dummy variable h_{it} if agents use their potential gains from growing hybrid in deciding whether or not to plant hybrid. This framework implies a *correlated* random coefficient (CRC) model that I show is a generalization of Chamberlain's (1984) correlated random effects (CRE) model, where only the intercepts are allowed to be correlated with h_{it}. The CRC model allows both the individual specific intercepts and slopes to be correlated with h_{it}. As I discuss in detail, this model can be estimated via methods similar to those introduced by Chamberlain (1984) if I consider only the special case of dummy endogenous regressors.

Notice that since θ_i and τ_i are uncorrelated by construction, and the individual specific slope is just $\phi\theta_i$, the covariance between individual specific slopes and intercepts in the yield function

implies that $\theta_i^H - \theta_i^N = (b_H - b_N)(\theta_i^H - \theta_i^N)$. So, for the τ_i's to be equal across sectors, $b_H - b_N$ must be equal to 1, which is easily shown: $b_H - b_N = \frac{\sigma_H^2 - \sigma_{HN} - \sigma_{HN} + \sigma_N^2}{\sigma_H^2 + \sigma_N^2 - 2\sigma_{HN}} = 1$.

[16] I discuss the stochastic properties and assumptions of this model in detail later.

is just:

$$cov(\alpha_i, \phi\theta_i) = \phi\sigma_\theta^2 \tag{16}$$

The structural coefficient ϕ therefore determines whether high intercept households are also high slope households. If $\phi > 0$, then $\psi > 1$, and the use of hybrid inflates the role of comparative advantage in the hybrid sector. In the long run, this would lead to greater inequality in yields in the overall economy. After discussing the estimation strategy and the results in the next two sections, I return to a discussion of comparative advantage and what these estimates of the $cov(\alpha_i, \phi\theta_i)$ actually mean. I describe how the covariance between slopes and intercepts above illustrates comparative advantage and I also outline the assumptions I am implicitly placing on a standard selection framework in my panel model. In addition, I show results from some of the more common selection correction methods, like the Heckman two-step, estimates of the average treatment effect (ATE), treatment on the treated (TT), marginal treatment effect (MTE) and local average treatment effect (LATE) for the normal non-random assignment treatment model.

5 Testing the Homogeneous Returns Model

5.1 Base Specifications

I start the empirical analysis by looking at some simple yield function specifications. Table 3 shows the OLS and household fixed effects specifications, both with and without covariates. The OLS estimates in the first column are extremely large and positive: households that plant hybrid maize tend to have higher maize yields. Also note the strong time trends in yields for my sample of households over this period. Adding province dummies in the second column of Table 3 decreases the coefficient on hybrid, as there are strong differences across provinces in both yields and hybrid use. In the third column, I add covariates to the specification. The purpose of these covariates is to control for other household variables that could affect yields, and that may be correlated with the use of hybrid maize. They include fertilizers (the results are robust to whether quantities or the total value are used), land preparation costs, seed quantities, variables that measure labor inputs, long term mean seasonal rainfall, current seasonal rainfall and land acreage. Adding these covariates decreases the OLS coefficient further, though it is still quite large at 0.56. The fourth and fifth columns of Table 3 report the household fixed effects results. The coefficient on hybrid decreases dramatically, though with covariates the difference between the OLS and fixed effects is less substantial. There is still a substantial return to hybrid maize even within households, controlling for this wide set of covariates, on the order of 0.15.

The rest of this section describes tests of the household fixed effects model, and some intuitive evidence of selection and heterogeneity in returns. I describe these tests and the results which do not show support for the homogeneous returns model. The household fixed effects estimates are consistent only under the assumption of strict exogeneity of the errors. Chamberlain's correlated random effects approach allows a test of this assumption (see Chamberlain (1984) and Jakubson

(1991)). I first describe this test and then examine other intuitive and descriptive evidence that highlights the role of selection and heterogeneity.

5.2 Two Period CRE Model

Chamberlain's CRE model provides a basis for testing the strict exogeneity of errors assumption underlying a standard fixed effects model. I illustrate the simple two period, no covariates CRE model, for which the data generating process is given by

$$y_{it} = \delta + \beta h_{it} + \alpha_i + u_{it} \tag{17}$$

A household fixed effects estimator of β is consistent under the following assumption of strict exogeneity of the errors, u_{it}:

$$E(u_{it}|h_{i1}, ..., h_{iT}, \alpha_i) = 0 \tag{18}$$

CRE illustrates how the fixed effects model is overidentified. Replace the fixed effect, α_i, by its *linear* predictor, based on the history of the covariates:

$$\alpha_i = \lambda_0 + \lambda_1 h_{i1} + \lambda_2 h_{i2} + v_i \tag{19}$$

where the projection error v_i is uncorrelated with h_{i1} and h_{i2} by construction and the λ's are the projection coefficients. Substituting this linear projection into the yield equation,

$$y_{it} = \delta + \beta h_{it} + \lambda_0 + \lambda_1 h_{i1} + \lambda_2 h_{i2} + v_i + u_{it} \tag{20}$$

Let $\epsilon_{it} = v_i + u_{it}$ where $E\left[\epsilon_{it} h_{i1}\right] = E\left[\epsilon_{it} h_{i2}\right] = 0$. For each time period, therefore:

$$y_{i1} = (\delta + \lambda_0) + (\beta + \lambda_1) h_{i1} + \lambda_2 h_{i2} + \epsilon_{i1} \tag{21}$$

$$y_{i2} = (\delta + \lambda_0) + \lambda_1 h_{i1} + (\beta + \lambda_2) h_{i2} + \epsilon_{i2} \tag{22}$$

These are the structural yield equations for each period. I estimate reduced form yield functions as follows:

$$y_{i1} = \delta_1 + \gamma_1 h_{i1} + \gamma_2 h_{i2} + \eta_{i1} \tag{23}$$

$$y_{i2} = \delta_2 + \gamma_3 h_{i1} + \gamma_4 h_{i2} + \eta_{i2} \tag{24}$$

It can be seen from equations (21) through (24) how the fixed effects model is overidentified. From the reduced form coefficients, $\gamma_1, \gamma_2, \gamma_3$ and γ_4, I can estimate the three structural parameters, λ_1, λ_2 and β using minimum distance. It is important to note here that estimating the CRE model does not require a specification of the conditional expectation of the α_i. Neither does it require knowledge of the *true* conditional expectation of the α_i.

The intuition behind the identification of the CRE model comes from the underlying assumption of the strict exogeneity of the errors. The idea is that if the fixed effects model is indeed valid, then the only way the history of h_{it} (both past and future values for any given time period) affect the current outcome is through the household level unobservable, α_i. This is testable with panel data as described above. The structural estimates, even in the two period case, are overidentified as there are four reduced form parameters from which to estimate the three structural estimates. The minimum distance estimator of the structural parameters is also the minimum χ^2 estimator if the weight matrix used is the variance covariance matrix of the reduced form coefficients. This is called the optimal minimum distance (OMD) estimator. If the identity matrix is used as the weight matrix instead, the estimates are referred to as equally weighted minimum distance (EWMD) estimates. The OMD estimates are efficient, but they can be biased in small samples. OMD can sometimes therefore be out-performed by EWMD (see Altonji and Segal (1996)). Throughout, I report both sets of estimates, as well as the χ^2 values on the OMD problem, which are just the value of the minimand in the OMD problem.

Estimates of the CRE model for three periods, both with and without covariates are shown in Table 4. This table shows both the reduced form and structural estimates for the various specifications. I show the reduced forms both with and without covariates. I estimate the structural parameters using minimum distance, comparing the OMD and EWMD estimates. The reduced forms in the upper panel of Table 4 give nine reduced form parameters (not including the constants or covariates), from which I use minimum distance to estimate the four structural estimates, shown in the lower panel of the table. Three of these structural estimates are the λ's from the linear projection of the fixed effects and the fourth is the estimate of the return to hybrid, β. The CRE estimates of β in all cases are very close to the household fixed effects estimates in Table 3, as expected. Also, the OMD and EWMD estimates of β are quite similar, all within sampling error of each other. The last column in the lower panel of Table 4 shows the χ^2 values on the overidentification test. In all cases, I can reject the null that the minimum distance restrictions hold. This overidentification test is an omnibus test. It therefore has low power against any *specific* alternative, but does have power against many alternatives. It is therefore not surprising I am able to reject the overidentifying restrictions. I now examine some other implications and tests of the household fixed effects model that illustrate the role of selection and heterogeneity.

5.3 Preliminary Evidence of Heterogeneity

To motivate my framework of heterogeneous returns, I begin by reporting some standard tests for heterogeneity (see Heckman, Smith and Clements (1997)). These are *purely* for illustrative purposes, as they assume that the data is experimental, i.e. that the samples of farmers in the hybrid and non-hybrid sectors are the same on average. This is an extremely strong assumption, and these tests are therefore only suggestive. Figures 6a, 6b and 6c show the marginal (conditional) distributions of yields, across the hybrid and non-hybrid sectors (i.e. across adopters and non-adopters) for each of the three periods of data I use. Note that in 2000, both the adopter and non-adopter distributions move to the right compared to 1997, but in 2004 the non-adopter distribution moves back to the left, closer to what it was in 1997.

Let the conditional yield distributions for the adopters and non-adopters of hybrid be $F^H(y^H|h=1)$ and $F^N(y^N|h=0)$ respectively. They bound the unknown joint distribution of interest, $F(y^H, y^N|h=1)$, via the Frechet-Hoeffding bounds as follows:

$$\begin{aligned} \max[F^H(y^H|h &= 1) + F^N(y^N|h=1) - 1, 0] \leq F(y^H, y^N|h=1) \\ &\leq \min[F^H(y^H|h=1), F^N(y^N|h=1)] \end{aligned} \tag{25}$$

These bounds allow us to bound the variance, $Var(y^H - y^N)$, but again only if we ignore the role of selection or allow selection to affect the joint distribution in very restrictive ways. We can then test whether the lower bound of $Var(\Delta y)$ is significantly different from zero as a test for heterogeneity in returns. An alternative test looks at the percentiles of the hybrid and non-hybrid yield distributions to see if each percentile of the two distributions differs by a common constant. Under no heterogeneity in returns, the percentiles of the two distributions should differ by the same constant, given assumptions about the dependence in the two distributions[17]. The null hypothesis is $H_0 : q(y^H) - q(y^N) = k$ for all q such that $0 \leq q \leq 100$. Figures 7a, 7b and 7c show the differences in percentiles of the returns distributions for each of the three periods, assuming perfect positive dependence, and Figures 8a and 8b show similar plots for my samples of joiners (farmers who start off a period not using hybrid and then use hybrid the next period) and leavers. Appendix Table A2 shows the full set of results for both cases of perfect positive and perfect negative dependence, and in all cases, the impact standard deviation is significantly different from zero, rejecting the null above.

5.4 Evidence of Selection

Of course, assuming away selection is hardly tenable, so in Table 5 I look for evidence of selection. I first split out the adoption history into dummies describing the transitions of households across technologies over the three periods. The idea is to look at whether households with different transition histories have different returns in terms of yields to planting hybrid (see Card and Sullivan (1988)). To understand transition histories, I define a joiner to be a farmer who does not plant hybrid the first period, but does the next, and a leaver to be a farmer who plants hybrid one period, but not the next. Similarly, I define a hybrid stayer to be a farmer who plants hybrid in both periods and a non-hybrid stayer to be one who plants traditional varieties in both periods. Under a household fixed effects model, the selection is reflected by the coefficients on the stayers and leavers in the periods in which they are *not* growing hybrid.

I therefore look at each pair of periods and compare the yields for the hybrid and non-hybrid stayers, the joiners, and the leavers in each of the two periods separately to learn about the extent of selection. For example, the first two columns in Table 5 look at the transitions of

[17] The two extreme dependence assumptions are perfect positive dependence (the individual at the 99th percentile in the hybrid distribution would be at the 99th percentile of the non-hybrid distribution had he not planted hybrid) and perfect negative dependence, where the percentile rankings are assumed to be reversed.

households over 1997-2000, with the omitted group being the non-hybrid stayers. The first and second columns of this table compare the yields in 1997 and 2000 for hybrid stayers, joiners and leavers in 1997 and 200 separately. If there was no selection at all (not even via a household fixed effect), we would expect the coefficient on the leavers in the yield equation for 1997 to be no different from the coefficient on the stayers, and also no different from the coefficient on the joiners in the yield equation for 2000. Similarly, the coefficient on joiners in the yield equation for 1997 would be no different from zero, as should the coefficient on the leavers in the yield equation for 2000. The coefficient on the leavers and joiners across the two yield equations therefore give us an idea of the extent of selection. Table 5 reports these estimates, not just for 1997-2000, but also for 2000-2004 and 1997-2004, in each case with and without covariates. In all three year pairs, the yields for the hybrid stayers are uniformly the largest and the leavers and joiners get very different yields.

The next step is to look for heterogeneity in returns to hybrid seed along observable dimensions. Table 6 reports results for yield functions, estimated separately for hybrid and non-hybrid households. I report both the OLS as well as household fixed effects specifications. We can see (perhaps unsurprisingly) that the returns to observables differ by technology choice, especially in the cases of fertilizers and rainfall. This holds for both the OLS as well as the household fixed effects specifications.

6 Estimating a Model with Heterogeneous Returns

The model of heterogeneous returns outlined earlier implies the following econometric model (without covariates) for the returns to hybrid maize:

$$y_{it} = \delta + \beta h_{it} + \alpha_i + \phi\theta_i h_{it} + \epsilon_{it} \tag{26}$$

For simplicity, I focus only on hybrid maize and leave out other inputs to explain the empirical strategy and discuss extensions later. The model above[18] can be estimated as per Lemieux (1998), using non-linear 2SLS. Instead, I extend the basic Chamberlain CRE approach to a scenario of correlated random coefficients. This is somewhat easier and allows a χ^2 overidentification test, similar to that described above. The next sections are devoted to describing my estimation and identification strategy, remembering the intuition of the CRE approach. I describe the extensions of the basic two period, no covariates model and then describe the estimation results.

[18] This empirical model is similar to models of individual specific heterogeneity in Heckman and Vytlacil (1997), Card (2000, 1998), Deschenes (2001, 2002), Carneiro, Hansen and Heckman (2001), Carneiro, Heckman and Vytlacil (2001), and Wooldridge (1997). Lemieux (1998) has a very similar scenario where he looks at whether the payoff to both observables and unobservables varies by union sector membership. He is able to estimate structural parameters similar to those identified above, using non linear two stage least squares.

6.1 Two Period CRC Model

For the simple two period, no covariates case, the yield function is

$$y_{it} = \delta + \alpha_i + \beta h_{it} + \phi \theta_i h_{it} + \varepsilon_{it} \tag{27}$$

The key identifying assumption here is that conditional on the comparative advantage component $\phi\theta_i h_{it}$ (and of course the covariates in the more general specifications), the unanticipated component of yields, ε_{it}, is not correlated with the decision to adopt. Remembering that $\alpha_i \equiv \theta_i + \tau_i$, where θ_i and τ_i are orthogonal, I re-write equation (27) as

$$y_{it} = \delta + \theta_i + \beta h_{it} + \phi \theta_i h_{it} + \tau_i + \varepsilon_{it} \tag{28}$$

Using the same idea as CRE, I linearly project the θ_i's onto the history of the hybrid decisions, as well as their interactions so that, by construction, the projection error is orthogonal to h_{i1} and h_{i2} individually, as well as to their product, $h_{i1}h_{i2}$[19]. The projection is therefore given by

$$\theta_i = \lambda_0 + \lambda_1 h_{i1} + \lambda_2 h_{i2} + \lambda_3 h_{i1} h_{i2} + v_i \tag{29}$$

Substituting, and since h_{it} is a dummy variable, the yield function is now

$$y_{it} = \delta+\lambda_0+\lambda_1 h_{i1}+\lambda_2 h_{i2}+\lambda_3 h_{i1}h_{i2}+\beta h_{it}+\phi\lambda_0 h_{it}+\phi\lambda_1 h_{i1}h_{it}+\phi\lambda_2 h_{i2}h_{it}+\phi\lambda_3 h_{i1}h_{i2}h_{it}+v_i+\phi v_i h_{it}+u_{it} \tag{30}$$

For each of the two time periods, therefore:

$$y_{i1} = (\delta+\lambda_0)+[\lambda_1(1+\phi)+\beta+\phi\lambda_0]h_{i1}+\lambda_2 h_{i2}+[\lambda_3(1+\phi)+\phi\lambda_2]h_{i1}h_{i2}+(v_i+\phi v_i h_{i1}+u_{i1}) \tag{31}$$

$$y_{i2} = (\delta+\lambda_0)+\lambda_1 h_{i1}+[\lambda_2(1+\phi)+\beta+\phi\lambda_0]h_{i2}+[\lambda_3(1+\phi)+\phi\lambda_1]h_{i1}h_{i2}+(v_i+\phi v_i h_{i2}+u_{i2}) \tag{32}$$

The corresponding reduced forms are:

$$y_{i1} = \delta_1 + \gamma_1 h_{i1} + \gamma_2 h_{i2} + \gamma_3 h_{i1}h_{i2} + \xi_{i1} \tag{33}$$

$$y_{i2} = \delta_2 + \gamma_4 h_{i1} + \gamma_5 h_{i2} + \gamma_6 h_{i1}h_{i2} + \xi_{i2} \tag{34}$$

From the six reduced forms coefficients from equations (33) and (34), $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5$ and γ_6, I can estimate the five structural parameters ($\lambda_1, \lambda_2, \lambda_3, \beta$ and ϕ) using minimum distance.

[19] The projection I use here is different from what CRE uses. If I use the simple CRE projection, $\theta_i = \lambda_1 h_{i1} + \lambda_2 h_{i2} + v_i$, and substitute this into the yield function,

$$y_{it} = \lambda_0 + \lambda_1 h_{i1} + \lambda_2 h_{i2} + \beta h_{it} + \phi\lambda_0 + \phi\lambda_1 h_{i1}h_{it} + \phi\lambda_2 h_{i2}h_{it} + v_i + \phi v_i h_{it} + u_{it}$$

Even though v_i, the projection error, is linearly uncorrelated with h_{i1} and h_{i2} individually, it is generally correlated with their product, $h_{i1}h_{i2}$, so that $E[v_i h_{i1} h_{i2}] \neq 0$. The projection I use must therefore include the interactions of the hybrid variables, not just their main effects.

The restrictions for the minimum distance problem are:

$$\begin{aligned}
\gamma_1 &= (1+\phi)\lambda_1 + \beta + \phi\lambda_0 \\
\gamma_2 &= \lambda_2 \\
\gamma_3 &= (1+\phi)\lambda_3 + \phi\lambda_2 \\
\gamma_4 &= \lambda_1 \\
\gamma_5 &= (1+\phi)\lambda_2 + \beta + \phi\lambda_0 \\
\gamma_6 &= (1+\phi)\lambda_3 + \phi\lambda_1
\end{aligned} \tag{35}$$

The structural parameters are overidentified and can therefore be estimated via minimum distance. Before I go on to discuss whether the individual θ_i's can be recovered from this estimation, I first discuss extensions to this basic model.

6.2 Extensions

I now consider extensions of the two period, no covariates CRC model described above to account for both more than two periods, as well as additional (potential endogenous) regressors. Consider the following extensions:

1. Covariates: all the identification arguments presented above carry through when the model includes covariates. The CRE model accommodates covariates/controls (and allows a test of whether the covariates are correlated with the fixed effects). This can then be generalized as above to the CRC case.

2. Three periods: this is also a simple extension. For space considerations, I do not show the 21 restrictions for the three period model. The problem does become heavily overidentified with only 9 structural parameters to estimate from the 21 reduced form coefficients.

3. More than one choice variable: the two-sector model presented above (hybrid/non-hybrid) can be extended to multiple sectors, but it quickly becomes cumbersome. An additional technology use sector, like fertilizers, can be incorporated by thinking of it as a four sector model (use both hybrid and fertilizers, neither, one or the other). I simplify it further. In my data, 76%, 79% and 71% percent of households respectively fit into one of two of the possible four sectors above, namely using both or neither. So, I think of the use of hybrid maize and fertilizers as necessarily a joint decision in an easier model that redefines sectors to this joint decision and looks at the heterogeneity in returns across these two sectors.

6.3 CRC Estimates

The standard procedures that correct for selection tend to be cross sectional in nature. They do not fully address the issues I am interested in, nor do they fully exploit the panel nature of my data (though I present them in the next section, purely for comparison). A better description of the selection mechanisms in a panel framework leads me to my CRC framework. This section describes various estimates of the CRC model. I report estimates for the pure hybrid model, both with and without covariates. In addition, I report some preliminary estimates for the joint hybrid-fertilizer extensions to the basic model.

Tables 7, 8a and 8b present the CRC reduced form and structural estimates. These tables report both the EWMD and OMD estimates for cases with and without covariates, as well as the χ^2 statistics on the overidentification tests for the OMD cases. The average returns, represented here by the β coefficient, are still on the order of the previous household fixed effects estimates, as expected. But, the coefficient of interest is ϕ, the comparative advantage, as it describes the covariance between the individual specific intercepts and slopes.

The upper panel in Table 7 presents the structural estimates of the CRC model for the two period case using data for 1997 and 2004, along with the χ^2 statistics on the overidentification of the restrictions for the OMD case. I present results both with and without covariates. The estimates are consistently negative, though with quite large standard errors in some cases. The lower panel of Table 7 present structural estimates of ϕ for the case where I allow the fertilizer covariate to enter the problem endogenously, i.e. to be correlated with the θ_i's. This generalizes the projection above to have the following form:

$$\begin{aligned}\theta_i &= \lambda_0 + \lambda_1 h_{i1} + \lambda_2 h_{i2} + \lambda_3 h_{i1} h_{i2} + \lambda_4 h_{i1} f_{i1} + \lambda_5 h_{i2} f_{i1} + \lambda_6 h_{i1} h_{i2} f_{i1} \\ &\quad + \lambda_7 h_{i1} f_{i2} + \lambda_8 h_{i2} f_{i2} + \lambda_9 h_{i1} h_{i2} f_{i2} + \lambda_{10} f_{i1} + \lambda_{11} f_{i2} + v_i\end{aligned}$$

where f_{it} for $t = 1, 2$ represents the use of fertilizer in each period. I allow the other covariates to enter the problem such that they are exogenous to the θ_i's[20]. The estimates of ϕ in this second panel of Table 7 are also consistently negative.

The first three columns in Table 8a illustrate the CRC reduced forms for the three period model, one for each of the periods 1997, 2000 and 2004, without covariates. The last three columns show similar reduced forms, this time controlling for covariates. The reduced forms contain all the possible interactions of the three hybrid histories, which implies a total of 21 reduced form estimates. These can then be mapped onto the 9 structural estimates: 7 λ's from the projection of the θ_i's, the average return to hybrid, β, and the comparative advantage coefficient, ϕ, which are shown in Table 8b.

As can be seen, the estimate of ϕ varies across the specifications. In the no covariates specifications, both the OMD and EWMD estimates of ϕ are positive and marginally significant. However, once we account for covariates the OMD estimate switches sign to become negative, though insignificant. The EWMD estimate, however, is strongly positive.

Table 9 reports the structural estimates for the joint fertilizer-hybrid problem as described in

[20] There is some justification for this. Using the estimates from the pure hybrid problem in Tables 7 and 8, I am able to back out the distribution of predicted θ_i's and correlate this with the covariates. Of all the covariates, only the correlations between the θ_i's and fertilizer use are important in magnitude.

the previous section. In my sample, there are farmers who use both technologies, one or neither. I define the technology sector to be the sample of farmers that use only both technologies (i.e. the relevant dummy variable for adoption here is coded as a 1 when the farmer used both fertilizers and hybrid maize and a 0 otherwise). Here both the OMD and EWMD with covariates are negative and significant. These estimates are quite different from the pure hybrid sector choice problem. The last section of this paper is devoted to discussing what all these estimates mean.

6.4 Recovering the Distribution of $\widehat{\theta_i}$

Before I move on to the conceptual discussion underlying my model and estimates, I look briefly at the distribution of the predicted θ_i's. Given the CRC structural estimates of the λ's from the previous section and the form of the projection given by equation (29), I can predict the θ_i's for a given history of hybrid use. I must assume that the projection in equation (29) describes the *true* conditional expectation of the θ_i's. In addition, once I have the distribution of comparative advantage, θ_i, I derive the distribution of the τ_i's via the yield function. Finally, this allows me to construct the α_i's since $\theta_i = \alpha_i + \tau_i$. The three distributions are shown in Figures 9a, 9b and 9c respectively.

Figure 9a shows the distribution of the predicted θ_i's in my sample. These are in fact averages of the individual true θ_i's for every history and interaction of histories of hybrid use. We can see that the distributed of predicted θ_i's for adopters lies to the right of that for non-adopters. Figure 9b shows the distributions of τ_i for adopters and non-adopters. They look quite similar. Of course, remember that τ_i and θ_i are orthogonal by construction. Finally, Figure 9c shows a similar picture for the α_i's. Since the α_i's are just the sum of the τ_i and θ_i, it is unsurprising that the distribution of α_i for adopters lies to the right of that for non-adopters.

7 Comparative Advantage, Sorting and Selection

The applied econometrics literature has seen an abundance of selection correction procedures, like the familiar Heckman two-step estimator and other control function estimators. This section describes the ideas of absolute and comparative advantage in a general selection model, and describes how they relate to the model I estimated. In addition, I briefly outline results from estimating a standard selection model, along with the various treatment effects (ATE, TT and MTE). I start with the simple definitions of comparative advantage and absolute advantage in each sector, before I outline a selection model and relate that back to my framework of heterogeneous returns.

The definitions and implications of absolute and comparative advantage as sorting mechanisms have been outlined by various labor economists (Willis (1986), Sattinger (1993) and Green (1991)). I draw on Carneiro and Lee (2004) who illustrate the concepts of absolute and comparative advantage very cleanly for selection models. Comparative advantage describes a situation where, for the individuals in the hybrid sector at any point in time, the returns are higher than for the individuals in the non-hybrid sector had they planted hybrid. In addition, had the hybrid individuals not planted hybrid, their returns would be lower than the returns of the individuals actually planting traditional varieties. Sorting on absolute advantage in each sector implies (but is not a consequence of) comparative advantage.

In terms of the simple Roy model in yields, $h_i = 1$ if $y_i^H > y_i^N$ and $h_i = 0$ if $y_i^H \leq y_i^N$. This implies that

$$\frac{y_i^H}{y_i^N} > 1 \text{ for } h_i = 1 \text{ and } \frac{y_i^H}{y_i^N} \leq 1 \text{ for } h_i = 0 \tag{36}$$

More generally, for any two individuals i and j in the hybrid and non-hybrid sectors respectively,

$$\frac{y_i^H}{y_i^N}_{h_i=1} > \left(\frac{y_j^H}{y_j^N}\right)_{h_j=0} \tag{37}$$

Equation (37) implies sorting based on comparative advantage. Absolute advantage of farmers in *each* sector, on the other hand, implies

$$\left(y_i^H\right)_{h_i=1} > \left(y_j^H\right)_{h_j=0} \text{ and } \left(y_i^N\right)_{h_i=1} < \left(y_j^N\right)_{h_j=0} \tag{38}$$

If absolute advantage in each sector is satisfied, as defined in equation (38), then automatically the condition in equation (37) is satisfied. The yield maximization rule imposed by the simple Roy model in equation (36) imposes comparative advantage but not necessarily absolute advantage in each sector. The generalized Roy model introduces costs and income/yield maximization is replaced by utility maximization where $h_i = 1$ if $u^H(y_i^H) - C^H > u^N(y_i^N) - C^N$ where C^i represents sector i specific costs. Now, *any* pattern of sorting is possible and you need the joint distribution $f(y^H, y^N)$ to understand the patterns of sorting and comparative advantage (the standard program evaluation problem). I relate this to my estimates of ϕ in the following section.

7.1 Cross Sectional Selection Model

I illustrate a simple one period (purely for simplicity) selection problem, along the lines of Heckman (2001), Heckman and Li (2003), Carneiro and Heckman (2002), etc. for the adoption of hybrid maize. Let y_i^H and y_i^N be the actual (not expected) yields for individuals in the hybrid and non-hybrid sectors respectively, given by

$$\begin{aligned} y_i^H &= X_i'\gamma^H + u_i^H \\ y_i^N &= X_i'\gamma^N + u_i^N \end{aligned} \tag{39}$$

where u_i^H and u_i^N are sector specific unobservables. Using the same form of a generalized yield equation as earlier for one period, and substituting,

$$\begin{aligned} y_i &= (X_i'(\gamma^H - \gamma^N))h_i + X_i'\gamma^N + [u_i^N + (u_i^H - u_i^N)h_i] \\ &= [X_i'(\gamma^H - \gamma^N) + (u_i^H - u_i^N)]h_i + X_i'\gamma^N + u_i^N \\ &\equiv \mu_i h_i + X_i'\gamma^N + u_i^N \end{aligned} \tag{40}$$

I can write the selection rule in terms of a binary choice selection rule, based on the latent variable, h_i^*, as follows:

$$h_i^* = p_i(Z_i) + u_i^s \tag{41}$$

where $h_i = 1$ if $h_i^* \geq 0$ and $h_i = 0$ otherwise. u_i^s is the selection error and Z includes the X's from the yield equations as well as other covariates that determine the selection equation (i.e. the exclusion restrictions). Think of the p_i as the propensity score function in the most general case. I will work with the linear case below. In the strict Roy model, the selection process is given by the straight difference in yields so that h_i^* is just $y_i^H - y_i^N$. The generalized Roy model introduces costs so that the the selection process is given by something closer to equation (41). The higher the selection error, u_i^s, in equation (41), the more likely the individual is to plant hybrid maize.

I can now relate this to my model of heterogeneous returns and outline what I assume about the stochastic structure in the selection process for identification. I assume the errors have a factor structure. In particular, the u_i's in the hybrid and non-hybrid sectors have the following factor structure:

$$u_i^N = \theta_i + \tau_i \tag{42}$$

$$u_i^H = \psi\theta_i + \tau_i \tag{43}$$

The θ_i is defined as in equation (11). I substitute this factor structure into equation (40), allowing $\phi \equiv (\psi - 1)$ and $\alpha_i \equiv \theta_i + \tau_i$ to get

$$\begin{aligned} y_i &= [X_i'(\gamma^H - \gamma^N) + (u_i^H - u_i^N)]h_i + X_i'\gamma^N + u_i^N \\ &= X_i'(\gamma^H - \gamma^N)h_i + X_i'\gamma^N + \phi\theta_i h_i + \alpha_i \end{aligned} \tag{44}$$

which is exactly my basic specification in equation (28). This assumption on the errors having a factor structure is similar to the factor assumptions made in Carneiro, Hansen and Heckman (2003). The parameter I am interested in is ϕ as it signs the covariance of individual specific slopes and intercepts in my model, $cov(\alpha_i, \phi\theta_i)$. In the selection model it maps onto a covariance between the errors in the two sectors,

$$\begin{aligned} cov(\alpha_i, \phi\theta_i) &= cov(u_i^N, (u_i^H - u_i^N)) \\ &= cov(u_i^N, u_i^H) - \sigma_{u^N}^2 \end{aligned} \tag{45}$$

7.2 Are the Adoption Decisions Unconstrained?

This literature on selection with heterogeneity also describes the treatment effects under non-random assignment (see Heckman (2001)). Heckman, Tobias and Vytlacil (2001) outline the relevant treatment effects here: the ATE, TT, MTE and LATE. Carneiro and Heckman (2002) illustrate the choice of college driven by comparative advantage in a simple Roy model (with no costs) where the individuals with the higher returns to schooling (i.e. the higher μ_i in equation (40) above) select into college and those with the lowest returns do not. The average returns of individuals in college is thus higher than the return for the person at the margin. However, it is possible to reverse this such that the marginal return is greater than the average when costs are sufficiently positively correlated with returns (i.e. the individuals with high returns are the individuals that also face high costs). If we assume the model is Gaussian, we can estimate the standard cross sectional selection parameters, along with the relevant ATE, TT and MTE. Say the yields in the hybrid and non-hybrid sectors are given by equations (39) above, and the selection equation is specified parametrically as

$$h_i^* = Z_i^{'}\pi + u_i^s \tag{46}$$

where

$$h_i(Z) = 1[h_i^*(Z_i) \geq 0] = 1[Z_i^{'}\pi + u_i^s \geq 0] \tag{47}$$

Drawing on Heckman, Tobias and Vytlacil (2001), the three treatment effects I look at, the ATE, TT and MTE, are given by

$$ATE(x) = E(\Delta|X = x) = x(\gamma^H - \gamma^N) \tag{48}$$

$$TT(x, z, h(Z) = 1) = x(\gamma^H - \gamma^N) + (\rho_H\sigma_H - \rho_N\sigma_N)\frac{\phi(z\pi)}{\Phi(z\pi)} \tag{49}$$

$$MTE(x, u^s) = x(\gamma^H - \gamma^N) + (\rho_H\sigma_H - \rho_N\sigma_N)u^s \tag{50}$$

where $\Delta = y^H - y^N$ is the yield gain from hybrid, $\rho_H\sigma_H$ and $\rho_N\sigma_N$ correspond to the coefficient on the selection term in a two-step selection model (assuming that $Var(u^s) = 1$) so that $\sigma_H = Var(u_i^H)$, $\sigma_N = Var(u_i^N)$, $\rho_H = Cov(u_i^H, u_i^s)$ and $\rho_N = Cov(u_i^N, u_i^s)$. In addition, $\frac{\phi(z\pi)}{\Phi(z\pi)}$ is the inverse mills ratio/selection correction term, where $\phi(.)$ and $\Phi(.)$ represent the normal probability and cumulative distribution functions respectively. The X's and Z's are as defined above. Under the factor assumptions, I can derive

$$\rho_H = \frac{Cov(\psi\theta_i, u_i^s)}{\sqrt{\psi\sigma_\theta^2 + \sigma_\tau^2}}$$

Using a similar expression for ρ_N, I can find an expression for $\rho_H\sigma_H - \rho_N\sigma_N$:

$$\begin{aligned}\rho_H\sigma_H - \rho_N\sigma_N &= (\psi - 1)Cov(\theta_i, u_i^s) \\ &= \phi Cov(\theta_i, u_i^s) \qquad (51)\end{aligned}$$

where $\sigma_\theta^2 = Var(\theta_i)$ and $\sigma_\tau^2 = Var(\tau_i)$. I use equation (51) in interpreting my results.

The treatment effects described above are extremely simple to compute. They involve a standard two-step control function procedure where I first run a probit of selection, use the estimates to back out the selection terms (mills ratio terms). The next step is to look at the sector specific yield functions, while controlling for the appropriate selection term. The ATE uses the estimates of the coefficients on the X's from this second step. The TT adjusts this estimated ATE for the selection into treatment (here, planting hybrid). And, finally, the MTE looks at the treatment effect as a function of the unobservables in the selection equation: it describes whether people who are more likely to use hybrid for unobservables reasons (the u_i^s) have higher or lower returns from planting hybrid. If the coefficient on u^s in equation (50) is negative, it implies that farmers with unobservables that make them the least likely to use hybrid have the highest yield return to planting hybrid. The MTE is very important to understanding whether heterogeneity is important and the role of unobservables that make farmers more (less) likely to use hybrid.

These treatment effects are shown in Table 10 for every cross section of my data. I need an exclusion restriction (or instrument), i.e. a variable that enters the selection equation (47), but not the yield functions in (39). I use a variable that describes a household's access to fertilizers and hybrid seed. It asks households what the distance between their homestead and the closest stockist of fertilizers is. This is usually not the actual distance to where fertilizers are purchased for the households that use inorganic fertilizers. Instead, this distance measure gives an idea of the access and availability of the technologies. I do not report the full selection corrected estimates for space considerations. Of importance is the MTE, to understand how it relates to the ATE. I evaluate the ATE's at the mean of the X's: they are all extremely large and positive, ranging from 0.756 in 2000 to 2.097 in 1997. The TT estimates are consistently smaller, ranging from 0.677 to 1.374. The MTE slope, meanwhile, is consistently negative across all three samples: -1.769 (with a standard error of 0.775) in 1997, -0.222 (0.370) in 2000 and -0.911 (0.251) in 2004. In two of these cases, the slope of the MTE function with respect to u^s is negative and significantly different from zero. This implies that farmers with unobservables that make them most likely to use hybrid get the lowest returns from hybrid. In addition, I can estimate the LATE (i.e. the IV estimate) using the same exclusion restriction as in the two-step procedure just described. These estimates of the LATE, reported in the lower panel of Table 10, are extremely large (on the order of 150%) especially when compared to the earlier OLS and household FE estimates.

Finally, in Table 11, I look at the OLS and household FE estimates of the return to hybrid and the prevalence of hybrid use across what I refer to as predicted fertility quantiles. Using the first period of my data, and just the sample of non-hybrid farmers, I estimate the expected yield based on my entire set of covariates. I predict what the expected yield under traditional varieties would be for those planting hybrid. This is an index of fertility for my entire sample of farmers, estimated from the non-hybrid farmers in 1997. I then look across quantiles of this estimated fertility measure. I compare the OLS and FE estimates of the hybrid return and the prevalence of hybrid across these quantiles. The results are reported in Table 11. It is clear that across these quantiles, the OLS estimates fall: the highest return is for the farmers in the lowest predicted fertility quantile. However, there is still a significant return even at the highest quantile. The FE estimates do not change across the quantiles as consistently as the OLS estimates. They first rise a little and then drop, but the estimate for the upper fertility quantile is zero, suggesting the causal effect of the technology goes to zero. Similar patterns hold out when looking at quantile regressions.

How do these estimates relate to my estimates of ϕ and what do they mean? In the strict Roy model without costs, $u_i^s = u_i^H - u_i^N = \theta_i$. So, from equation (51) above, $\rho_H \sigma_H - \rho_N \sigma_N = \phi \sigma_\theta^2$ so a negative MTE slope must imply that $\phi < 0$. But, in the generalized Roy model with costs, the $u_i^s = \theta_i - C_i$ where C_i is a measure of the relative costs in the hybrid sector. This implies

$$\rho_H \sigma_H - \rho_N \sigma_N = \phi \sigma_\theta^2 - \phi Cov(\theta_i, C_i) \tag{52}$$

Since $\rho_H \sigma_H - \rho_N \sigma_N < 0$, if $\phi > 0$ it implies that the $Cov(\theta_i, C_i)$ is quite large and positive (enough to outweigh the $\phi \sigma_\theta^2$). On the other hand if $\phi < 0$, then the $Cov(\theta_i, C_i)$ is either positive and small or negative. In terms of a generalized Roy model, therefore, the sign of ϕ describes something about the covariance between the household level unobservable comparative advantage and the relative costs of planting hybrid maize.

Given this battery of results, it is worth tying them all together. Note that both from the agronomy of hybrid maize, as well as the kernel density plots in Figure 6, it is clear that what hybrid does is increase productivity (observed and unobserved) by some amount, but this increase declines as you move rightward through the distribution, such that the expected return from hybrid is zero in the right tail. The difference in hybrid and non-hybrid yields is the average return β plus ϕ multiplied by the productivity. When $\phi < 0$, the marginal gain is negative. So, the farmers in the left tail have the largest gain – this is seen from the two period estimates of ϕ as well as the negative selection estimates above. If selection is only as per the Roy Model, we would see the largest adoption rates in the left tail, but this ignores the possibility of larger costs and greater constraints for these farmers. The LATE estimates using distance suggest however that distance *is* a constraining factor, at least for the low yield households. The mirror image of this is that there seems to be "over-adoption" for the farmers in the right tail, which may indicate that risk may be an important factor in the adoption decision, since the mean *gross* (not even net) returns to hybrid in the upper quartile are almost nil. To account for risk, it would be necessary to allow the variance of the transitory component in yields to play a role.

The risk issue aside, there does seem to be a need to incorporate a second index to explain why the percentage of hybrid use is increasing along quartiles. There is also a need to understand why the LATE estimates are greater than the mean θ_i's.

8 Conclusion

This paper examines the adoption decisions and benefits to hybrid maize in Kenya in a framework that is in stark contrast to the empirical technology innovation literature in the past two decades. Rather than think of adoption decisions as based on learning and information externalities, I instead focus on a framework that recognizes the large disparities in farming and input supply characteristics across the maize growing areas of Kenya. Within this framework, I find that returns to hybrid maize indeed vary greatly (although much less so once hybrid and fertilizer are considered as a joint decision). Furthermore, those farmers who are on the margin of adopting and disadopting (and who do so during my sample) experience very little change in yields, a finding consistent with my framework, yet harder to reconcile with a pure learning model.

Experimental evidence from both (controlled) experimental stations as well as field trials point to average high, positive returns to agricultural technologies. However, these experiments say very little else about the returns, other than demonstrating some irrationality driving adoption to account for patterns of low, non-increasing adoption rates. My framework of heterogeneous returns allows me to estimate not just average returns, accounting for selection, but also have an idea of what the distribution of returns looks like across my sample of farmers. Once I account for comparative advantage in the selection process, I find that a large component of the returns to hybrid maize are realized by households selecting into the hybrid sector. There is extremely strong evidence of heterogeneity in returns to the technology, with comparative advantage playing an important role in yield determination. In addition to estimates of comparative advantage trends across different technologies, this paper compares the panel estimation proposed to the more standard selection correction procedures. I am able to estimate average and marginal treatment effects, the latter illustrating the role of unobservables. Average returns are extremely large. I find, however, that households use their expected gain from planting hybrid to select into the hybrid sector so that marginal returns are lower than the average returns.

Within the existing technological inputs (i.e. hybrid varieties), the contrast between the results treating hybrid adoption as a single decision versus the more general framework treating the hybrid and fertilizer as a joint decision indicate that cost factors such as access to fertilizer may serve as barriers to beneficial hybrid adoption. At the more macro level, however, the low marginal returns indicate that encouraging further adoption of the existing hybrid varieties will do little to alter aggregate maize supply since the farmers adoption and non-adoption decisions are already doing this at the micro level. If increasing maize supply itself is the policy objective, emphasis should be placed on the development of new hybrid varieties as in India and Latin America, and not on only adopting the current set of varieties already in Kenya.

References

Altonji, Joseph, and Lewis Segal, (1996), "Small Sample Bias in GMM Estimation of Covariance Structures", *Journal of Business and Economic Statistics*, 14, pp. 353-366.

Ashenfelter, Orley, and David Card, (1985), "Using the Longitudinal Structure of Earnings to Estimate the Effects of Training Programs", *Review of Economics and Statistics*, 67(4), pp. 648-660.

Bandiera, Oriana, and Imran Rasul, (2003), "Complementarities, Social Networks and Technology Adoption in Mozambique", Working Paper, London School of Economics and University of Chicago.

Banerjee, Abhijit, (1992), "A Simple Model of Herd Behavior", *Quarterly Journal of Economics*, 107(3), pp. 797-817.

Besley, Timothy, and Anne Case, (1993), "Modelling Technology Adoption in Developing Countries", *AER Papers and Proceedings*, 83(2), pp. 396-402.

Boozer, Michael, and Stephen Cacciola, (2001), "Inside The 'Black Box' of Project Star: Estimation of Peer Effects Using Experimental Data", Economic Growth Center Discussion Paper 832, Yale University, June 2001.

Card, David, and Daniel Sullivan, (1988), "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment", *Econometrica*, 56.

Card, David, (1991), "The Effect of Unions on the Distribution of Wages: Redistribution or Relabelling?", Working Paper No. 287, Industrial Relations Section, Princeton University.

Card, David, (1996), "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis", *Econometrica*, 64.

Card, David, (1998), "The Causal Effect of Education on Earnings", in the Handbook of Labour Economics, eds. Orley Ashenfelter and David Card.

Card, David, (2000), "Estimating The Returns to Schooling: Progress of Some Persistent Econometric Problems", NBER Working Paper No. 7769, Cambridge, MA.

Carneiro, Pedro, and James Heckman, (2002), "The Evidence on Credit Constraints in Post-Secondary Schooling", *Economic Journal*, 112 (October), pp. 705-734.

Chamberlain, Gary, (1984), "Panel Data", in the Handbook of Econometrics, eds. Zvi Griliches and Michael Intriligator.

Conley, Timothy, and Christopher Udry, (2003), "Learning About A New Technology: Pineapple in Ghana", Working Paper, University of Chicago and Yale University.

Corbett, John, (1998), "Classifying Maize Production Zones in Kenya Through Multivariate Cluster Analysis", in *Maize Technology Development and Transfer: A GIS Application for Research Planning in Kenya*, ed. Rashid Hassan, CAB International, Wallingford, UK.

Croppenstedt, Andre, Mulat Demeke, and Meloria Meschi, (2003), "Technology Adoption in the Presence of Constraints: The Case of Fertilizer Demand in Ethiopia", *Review of Development Economics*, 7(1), pp. 58-70.

David, Paul, (2003), "Zvi Griliches on Diffusion, Lags and Productivity Growth... Connecting the Dots", Paper Prepared for the Conference on R&D, Education and Productivity, Held in Memory of Zvi Griliches (1930-1999), August 2003.

De Groote, Hugo, William Overholt, James Ouma, and S. Mugo, (2003), "Assessing the Potential Impact of Bt Maize in Kenya Using a GIS Based Model", Working Paper, CIMMYT (International Maize and Wheat Improvement Center), Paper Presented at the International Agricultural Economics Conference, Durban, South Africa.

De Groote, Hugo, Cheryl Doss, Stephen Lyimo, and Wilfred Mwangi (2002), "Adoption of Maize Technology in East Africa: What Happened to Africa's Emerging Maize Revolution?" Paper presented for FASID Forum V, *Green Revolution in Asia and its Transferability to Africa*, December 2002.

Doss, Cheryl, (2003), "Understanding Farm Level Technology Adoption: Lessons Learned from CIMMYT's Micro Surveys in Eastern Africa", Economics Working Paper No. 03-07, CIMMYT (International Maize and Wheat Improvement Center), Kenya.

Duflo, Esther, Michael Kremer, and Jonathan Robinson, (2003), "Understanding Technology Adoption: Fertilizer in Western Kenya, Preliminary Results from Field Experiments", Working Paper, MIT, Harvard University and Princeton University.

Duncan, Gregory, and Duane Leigh, (1985), "The Endogeneity of Union Status: An Empirical Test", *Journal of Labor Economics*, 3 (3), pp. 385-402.

Evenson, Robert, and Germano Mwabu, (1998), "The Effects of Agricultural Extension on Farm Yields in Kenya", Economic Growth Center Discussion Paper No. 798.

Feder, Gershon, Richard Just, and David Zilberman, (1985), "Adoption of Agricultural Innovations in Developing Countries: A Survey", *Economic Development and Cultural Change.*

Food and Agriculture Organization, FAOSTAT Online Database, Available at http://faostat.fao.org/faostat/collections?subset=agriculture.

Foster, Andrew, and Mark Rosenzweig, (1995), "Learning by Doing and Learning from Others: Human Capital and Technological Change in Agriculture", *Journal of Political Economy*, 103, pp. 1176-1209.

Gerhart, John Deuel, (1975), "The Diffusion of Hybrid Maize in Western Kenya," Ph.D. Dissertation, Princeton University.

Gibbons, Robert, Lawrence F. Katz, Thomas Lemieux, and Daniel Parent, (2002), "Comparative Advantage, Learning and Sectoral Wage Determination", Scientific Series, CIRANO.

Green, David, (1991), "A Comparison of the Estimation Approaches for the Union-Nonunion Wage Differential", Discussion Paper No. 91-13, Department of Economics, University of British Columbia, Vancouver, Canada.

Griliches, Zvi, (1957), "Hybrid Corn: An Exploration in the Economics of Technological Change", *Econometrica*, 25, pp. 501-522.

Hall, Bronwyn, (2004), "Innovation and Diffusion", NBER Working Paper No. 10212.

Hassan, Rashid, Mulugetta Mekuria, and Wilfred Mwangi, (2001), "Maize Breeding Research in Eastern and Southern Africa: Current Status and Impacts of Past Investments Made by the Public and Private Sectors, 1966-1997", Mexico, D.F.: CIMMYT.

Hassan, Rashid, John Lynam, and Peter Okoth, (1998), "The Spatial Sampling Frame and Design for Farmer and Village Surveys", in *Maize Technology Development and Transfer: A GIS Application for Research Planning in Kenya*, ed. Rashid Hassan, CAB International, Wallingford, UK.

Hassan, Rashid, John Corbett, and Kiarie Njoroge, (1998), "Combining Geo-Referenced Survey Data With Agroclimatic Attributes to Characterize Maize Production Systems in Kenya", in *Maize Technology Development and Transfer: A GIS Application for Research Planning in Kenya*, ed. Rashid Hassan, CAB International, Wallingford, UK.

Hassan, Rashid, Ruth Onyango, and J.K. Rutto, (1998), "Relevance of Maize Research in Kenya to Maize Production Problems Perceived by Farmers", in *Maize Technology Development and Transfer: A GIS Application for Research Planning in Kenya*, ed. Rashid Hassan, CAB International, Wallingford, UK.

Hassan, Rashid, Kiairie Njoroge, Mugo Njore, Robin Otsyula, and A. Laboso, (1998), "Adoption Patterns and Performance of Improved Maize in Kenya", in *Maize Technology Development and Transfer: A GIS Application for Research Planning in Kenya*, ed. Rashid Hassan, CAB International, Wallingford, UK.

Hassan, Rashid, Festus Murithi, and Geofry Kamau, (1998), "Determinants of Fertilizer Use and the Gap Between Farmers' Maize Yields and Potential Yields in Kenya", in *Maize Technology Development and Transfer: A GIS Application for Research Planning in Kenya*, ed. Rashid Hassan, CAB International, Wallingford, UK.

Heckman, James, and Bo Honore, (1990), "The Empirical content of the Roy Model", *Econometrica*, 58 (5), pp. 1121-1149.

Heckman, James, (1990), "Varieties of Selection Bias", *AEA Papers and Proceedings*, 80 (2), pp. 313-318.

Heckman, James, Jeffrey Smith, with the assistance of Nancy Clements (1997), "Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts", *Review of Economic Studies*, 64, pp. 487-535.

Heckman, James, and Edward Vytlacil, (1997), "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling", *Journal of Human Resources*, 33(4), pp. 974-987.

Heckman, James, Justin Tobias, and Edward Vytlacil, (2001), "Four Parameters of Interest in the Evaluation of Social Programs", *Southern Economic Journal*, 68 (2), pp. 210-223.

Heckman, James, (2001), "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture", *Journal of Political Economy*, 109 (4), pp. 673-748.

Jakubson, George, (1991), "Estimation and Testing of the Union Wage Effect Using Panel Data", *Review of Economic Studies*, 58 (5), pp. 971-991.

Jayne, Thom S., Robert J. Myers, and James Nyoro, (2005), "Effects of Government Maize Marketing and Trade Policies on Maize Market Prices in Kenya", Working Paper, Michigan State University and Tegemeo Institute.

Jayne, Thom S., Takashi Yamano, James Nyoro, and Tom Awuor, (2001), "Do Farmers Really Benefit from High Food Prices? Balancing Rural Interests in Kenya's Maize Pricing and Marketing Policy", Tegemeo Draft Working Paper 2B, Tegemeo Institute, Kenya.

Just, Richard, and David Zilberman, (1983), "Stochastic Structure, Farm Size and Technology Adoption in Developing Agriculture", *Oxford Economic Papers*, 35, pp. 307-328.

Kamau, Mercy, (2002), "An Overview of the Kenyan Seed Industry in a Liberalized Environment: A Case Study of Maize Seed", Working Paper, Tegemeo Institute, Presented in Stakeholder Consultation Meeting, 2002.

Karanja, Daniel, (1996), "An Economic and Institutional Analysis of Maize Research in Kenya", MSU International Development Working Paper No. 57, Department of Agricultural Economics and Department of Economics, Michigan State University, East Lansing, MI.

Karanja, Daniel, Thomas Jayne, and Paul Strasberg, (1998), "Maize Productivity and Impact of Market Liberalization in Kenya", KAMPAP (Kenya Agricultural Marketing and Policy Analysis Project), Department of Agricultural Economics, Egerton University, Kenya.

Karanja, Daniel, Mitch Renkow, and Eric Crawford, (2003), "Welfare Effects of Maize Technologies in Marginal and High Potential Regions of Kenya", *Agricultural Economics*, 29 (3), pp. 331-341.

Kumar, P., and T. Stengos, (1985), "Measuring the Union Relative Wage Impact: A Methodological Note", *Canadian Journal of Economics*, 18, pp. 182-189.

Lemieux, Thomas, (1993), "Estimating the Effects of Unions on Wage Inequality in a Two-Sector Model with Comparative Advantage and Non-Random Selection", Working Paper, Centre de Recherche et Developpement en Economique (CRDE) and Departement de Sciences Economiques, Universite de Montreal, Publication No. 9303.

Lemieux, Thomas, (1998), "Estimating the Effects of Unions on Wage Inequality in a Two-Sector Model with Comparative Advantage and Non-Random Selection", *Journal of Labor Economics*, 16(2), pp. 261-291.

Lynam, John, and Rashid Hassan, (1998), "A New Approach to Securing Sustained Growth in Kenya's Maize Sector", in *Maize Technology Development and Transfer: A GIS Application for Research Planning in Kenya*, ed. Rashid Hassan, CAB International, Wallingford, UK.

Makokha, Stella, Stephen Kimani, Wilfred Mwangi, Hugo Verkuijl, and Francis Musembi, (2001), "Determinants of Fertilizer and Manure Use for Maize Production in Kiambu District, Kenya", Mexico, D.F.: CIMMYT (International Maize and Wheat Improvement Center).

Mandelbrot, Benoit, (1962), "Paretian Distributions and Income Maximization", *Quarterly Journal of Economics*, 76, pp.57-85.

Manski, Charles, (2000), "Economic Analysis of Social Interactions", *Journal of Economic Perspectives*, 14(3), Summer 2000, 115-36.

Manski, Charles, (1993), "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies*, 60, 1993, 531-542.

McCann, James C., (2005), *Maize and Grace: Africa's Encounter with a New World Crop, 1500-2000*, Harvard University Press, USA.

Mills, Bradford, Rashid Hassan, and Peterson Mwangi, (1998), "Estimating Potential Benefits from Research and Setting Research Priorities for Maize in Kenya", in *Maize Technology Development and Transfer: A GIS Application for Research Planning in Kenya*, ed. Rashid Hassan, CAB International, Wallingford, UK.

Moffitt, Robert, (2000), "Policy Interventions, Low-Level Equilibria and Social Interactions", in *Social Dynamics*, eds. Steven Durlauf and Peyton Young, Brookings Institution Press, Washington, DC.

Moser, Christine, and Christopher Barrett, (20030, "The Complex Dynamics of Smallholder Technology Adoption: The Case of SRI in Madagascar", Working Paper, Cornell University.

Mugo, Frida, Frank Place, Brent Swallow, and Maggie Lwayo, (2000), "Improved Land Management Project: Socioeconomic Baseline Study", Nyando River Basin Baseline Data Report, ICRAF, Kenya.

Munshi, Kaivan, (2003), "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution", *Journal of Development Economics*, 73, pp. 185-213.

Mundlak, Yair, (1961), "Empirical Production Function Free of Management Bias", *Journal of Farm Economics*, 43(1) pp. 44-56.

Mwangi, Wilfred, John Lynam, and Rashid Hassan, (1998), "Current Challenges and Strategic Future Choices for Maize Research and Policy in Kenya: A Synthesis of the Maize Data Base Project Methods and Results", in *Maize Technology Development and Transfer: A*

GIS Application for Research Planning in Kenya, ed. Rashid Hassan, CAB International, Wallingford, UK.

Nyameino, David, Bernard Kagira, and Stephen Njukia, (2003), "Maize Market Assessment and Baseline Study for Kenya", RATES (Regional Agricultural Trade Expansion Support) Program, Nairobi, Kenya.

Nyoro, James K., Maria Wanzala, and Tom Awour, (2001), "Increasing Kenya's Agricultural Competitiveness: Farm Level Issues", Working Paper No. 4.

Nyoro, James K., M. W. Kiiru, and Thom S. Jayne, (1999), "Evolution of Kenya's Maize Marketing Systems in the Post-Liberalization Era", Working Paper, Tegemeo Institute, Paper Presented at the 4th Agricultural Transformation Workshop held in Nairobi, Kenya, June 1999.

Ouma, James, Festus Murithi, Wilfred Mwangi, Hugo Verkuijl, Macharia Gethi, and Hugo De Groote, (2002), "Adoption of Maize Seed and Fertilizer Technologies in Embu District, Kenya", Mexico, D.F.: CIMMYT (International Maize and Wheat Improvement Center).

Ovuka, Mira, and Anders Ekbom, (1999), "Farmers' Resource Levels, Soil Properties and Productivity in Kenya's Central Highlands", in D.E. Stott, R.H. Mohtar and G.C. Steinhardt (eds. 2001), "Sustaining the Global Farm", Selected Papers from the 10th International Soil Conservation organization Meeting, held May 24-29, 1999.

Place, Frank, and Brent Swallow, (2000), "Assessing the Relationship Between Property Rights and Technology Adoption on Smallholder Agriculture: A Review of Issues and Empirical Methods", CAPRi Working Paper No. 2, ICRAF, Kenya.

Robinson, Chris, (1989), "The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Models", *Journal of Political Economy*, 97 (3), pp. 639-67.

Rogers, Everett, (1995), "Diffusion of Innovations", Fourth Edition, The Free Press, New York, NY.

Roy, A., (1951), "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers*, 3, pp. 135-146.

Salasya, M.D.S., W. Mwangi, Hugo Verkuijl, M.A. Odendo, and J.O. Odenya, (1998), "An Assessment of the Adoption of Seed and Fertilizer Packages and the Role of Credit in Smallholder Maize Production in Kakamega and Vihiga Districts, Kenya", Mexico, D.F.: CIMMYT (International Maize and Wheat Improvement Center).

Sanders, John, Barry Shapiro, and Sunder Ramaswamy, (1996), "The Economics of Agricultural Technology in Semiarid Sub-Saharan Africa", John Hopkins University Press, Baltimore, MD.

Schultz, Theodore, (1964), *The Economic Value of Education*, Columbia University Press, New York, NY.

Schultz, Theodore, (1964), *Transforming Traditional Agriculture*, Yale University Press, New Haven, CT.

Smale, Melinda, and Thom Jayne, (2003), "Maize in Eastern and Southern Africa: "Seeds" of Success in Retrospect", EPTD Discussion Paper No. 97.

Smale, Melinda, and Hugo De Groote, (2003), "Diagnostic Research to Enable Adoption of Transgenic Crop Varieties by Smallholder Farmers in Sub-Saharan Africa", *African Journal of Biotechnology*, 2 (12), pp. 586-595.

Sserunkuuma, Dick, (2002), "The Adoption and Impact of Improved Maize Varieties in Uganda", Paper Prepared for the Symposium on The Green Revolution in Asia and its Transferability in Africa, Department of Agricultural Economics and Agribusiness, Makerere University, Uganda.

Sunding, David, and David Zilberman, (2001), "The Agricultural Innovation Process: Research and Technology Adoption in a Changing Agricultural Sector", Chapter 4, *Handbook of Agricultural Economics*, Volume 1, eds. B. Gardner and G. Rausser.

Tittonell, Pablo, (2003), "Soil Fertility Gradients in Smallholder Farms of Western Kenya, Their Origin Magnitude and Importance", MSc Thesis, Department of Plant Sciences, Wageningen University, Netherlands.

Udry, Christopher, (1996), "Gender, Agricultural Productivity and the Theory of the Household", *Journal of Political Economy*, 105(5).

Vella, F., and M. Verbeek, (1998), "Whose Wages Do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men", *Journal of Applied Econometrics*, 13, pp. 163-168.

Wanzala, Maria, Thom Jayne, John Staatz, Amin Mugera, Justus Kirimi, and Joseph Owuor, (2001), "Fertilizer Markets and Agricultural Production Incentives: Insights from Kenya", Tegemeo Institute Working Paper No. 3, Nairobi, Kenya.

Weir, Sharada, and John Knight, (2000), "Adoption and Diffusion of Agricultural Innovations in Ethiopia: The Role of Education", Working Paper, Center for the Study of African Economies, Oxford University.

Wekesa, E., Wilfred Mwangi, Hugo Verkuijl, K. Danda, and Hugo De Groote, (2003), "Adoption of Maize Technologies in the Coastal Lowlands of Kenya", Mexico, D.F.: CIMMYT.

Wooldridge, Jeffrey, (1997), "On Two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model", *Economics Letters*, 56, pp. 129-133.

World Bank, World Development Indicators, WDI Online Database, Available at http://www.worldbank.org.

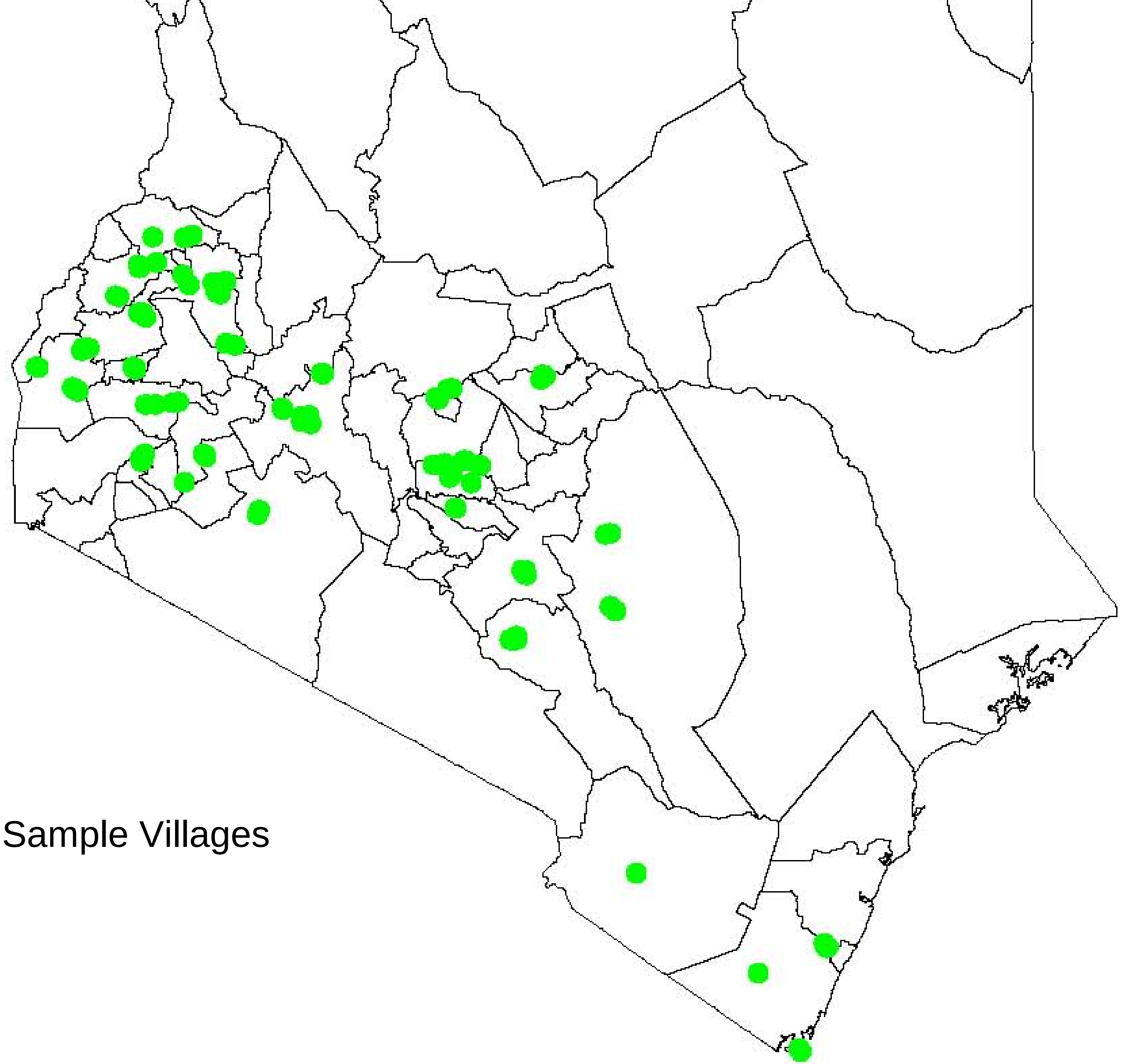

Figure 1
Location of Sample Villages

Figure 2
Hybrid Maize Adoption Patterns, by Season

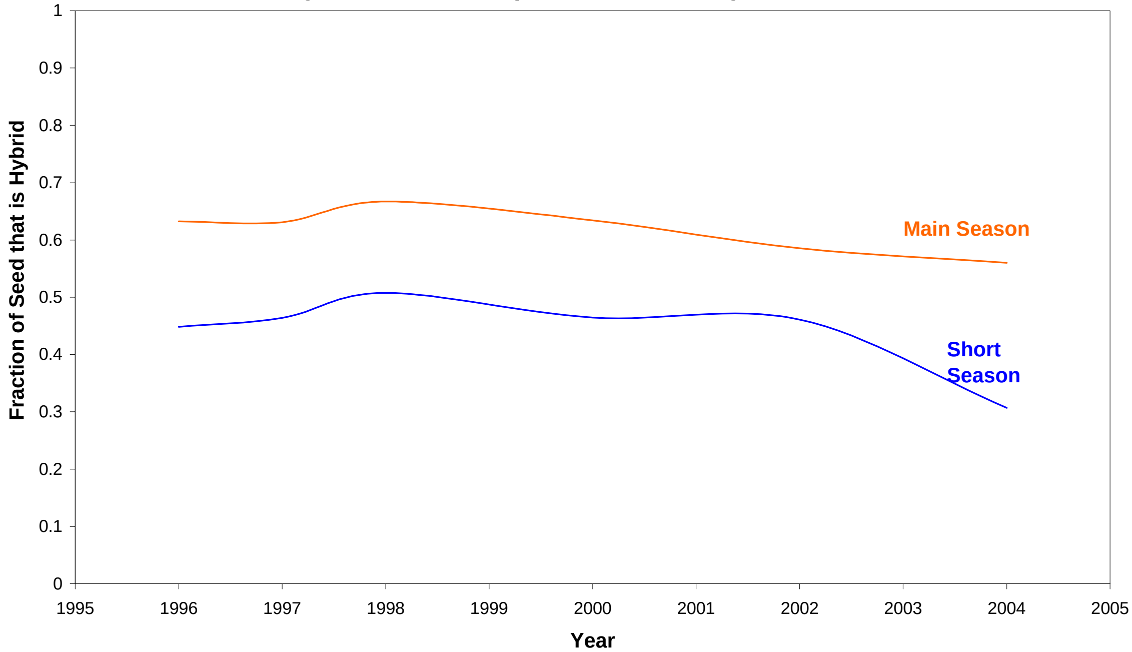

Figure 3
Hybrid Maize Adoption Patterns, by Province

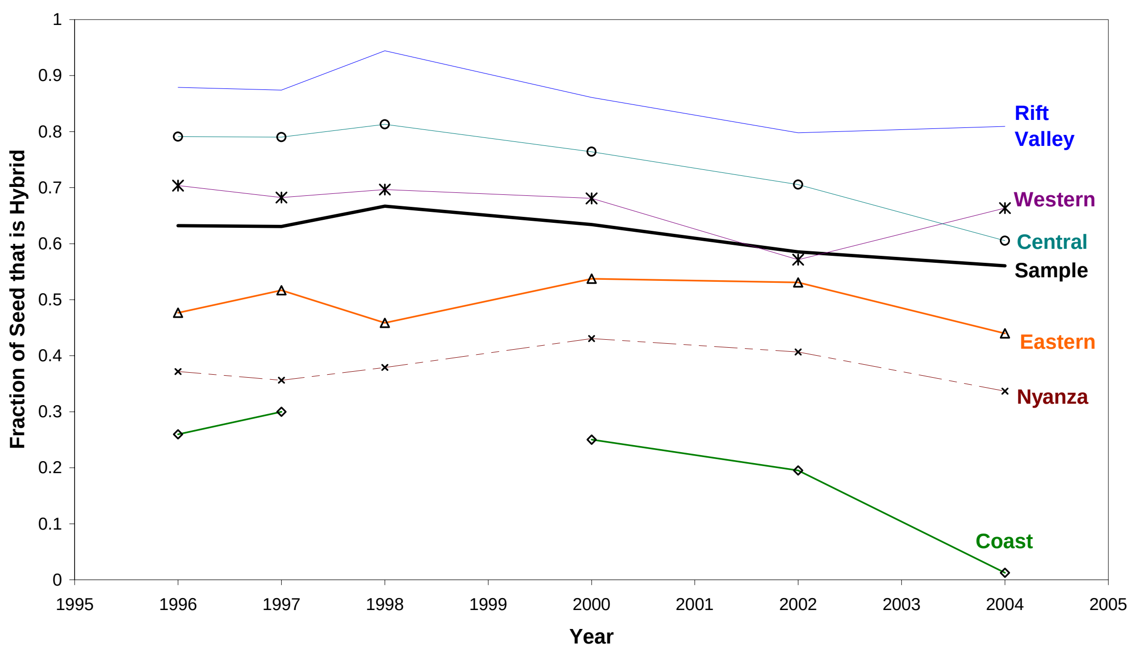

Figure 4a
Fraction of HH's Using Inorganic Fertilizer, by Province

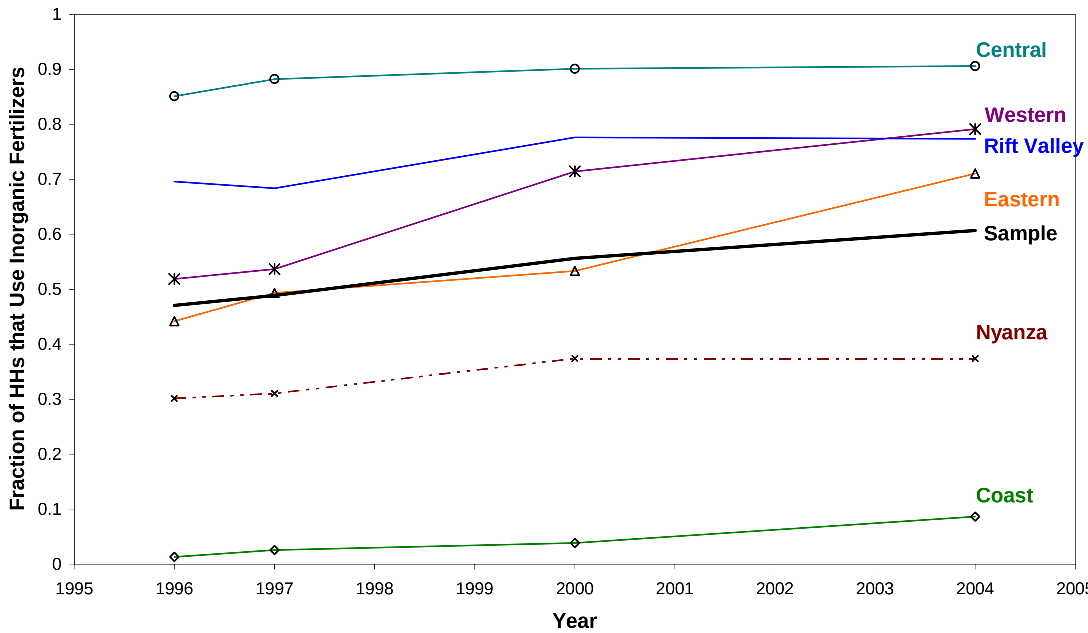

Figure 4b
Real Expenditure on Inorganic Fertilizer, by Province

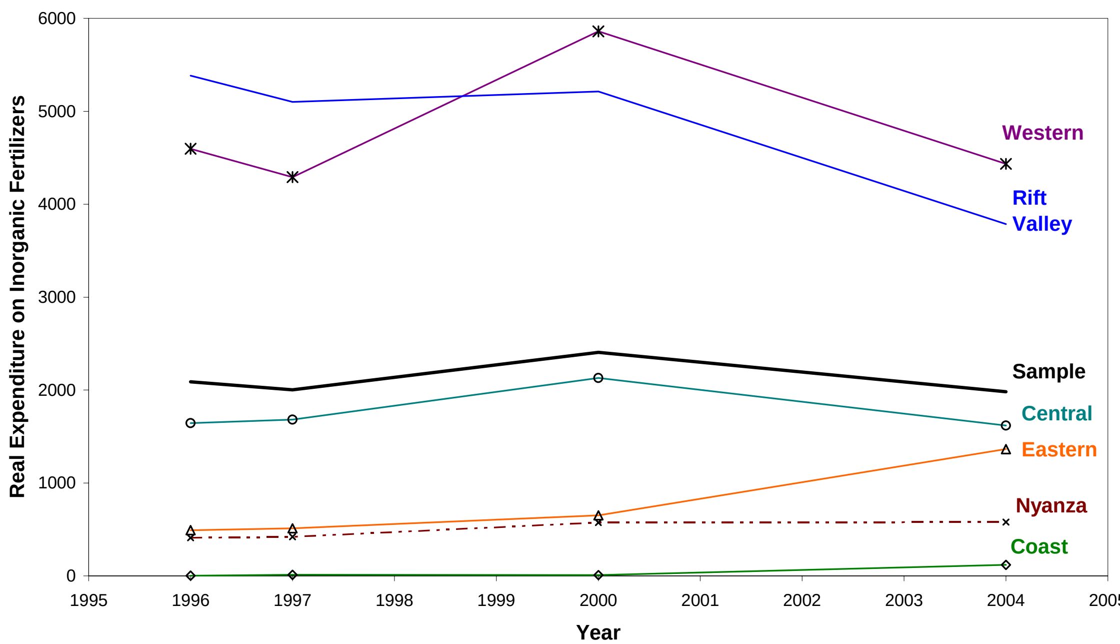

Figure 5a
Maize Yield, by Province

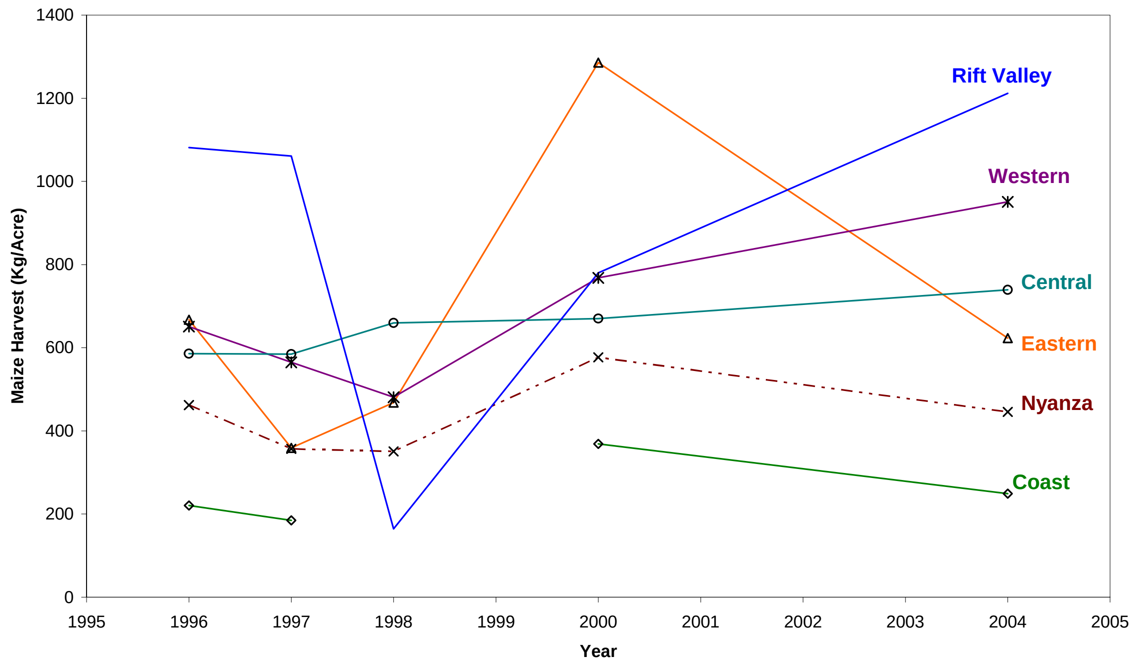

Figure 5b
Maize Acreage, by Province

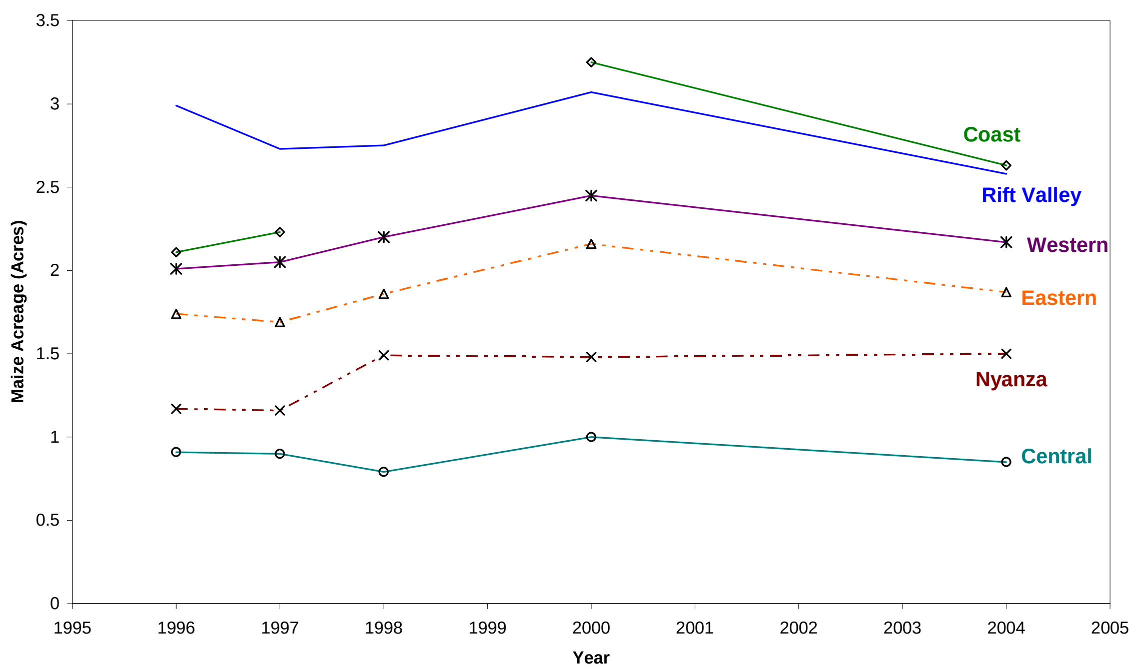

Figure 6a
Marginal Distributions of Yields by Sector, 1997

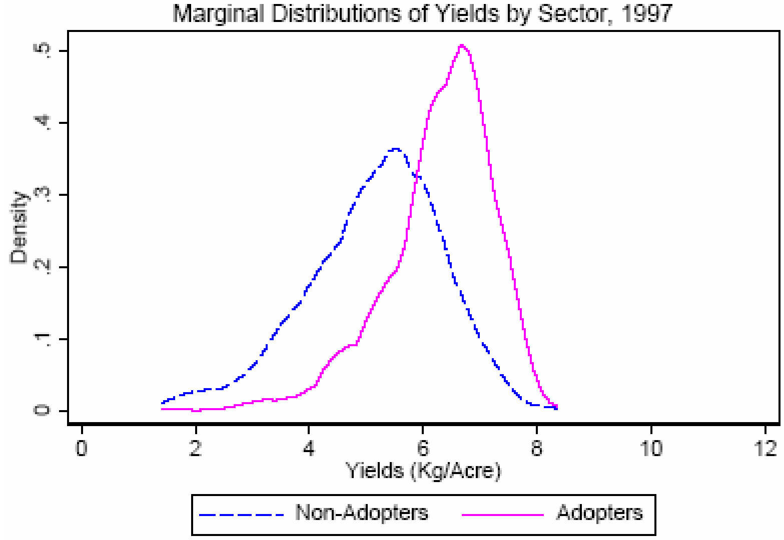

Figure 6b
Marginal Distributions of Yields by Sector, 2000

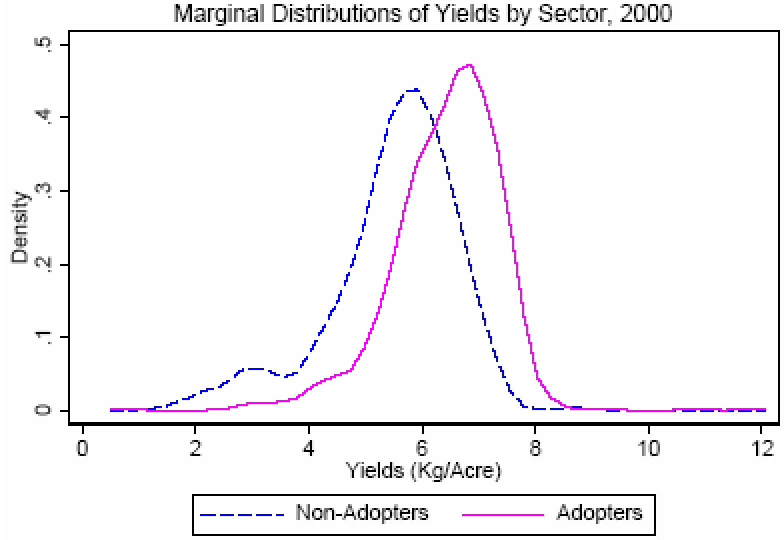

Figure 6c
Marginal Distributions of Yields by Sector, 2004

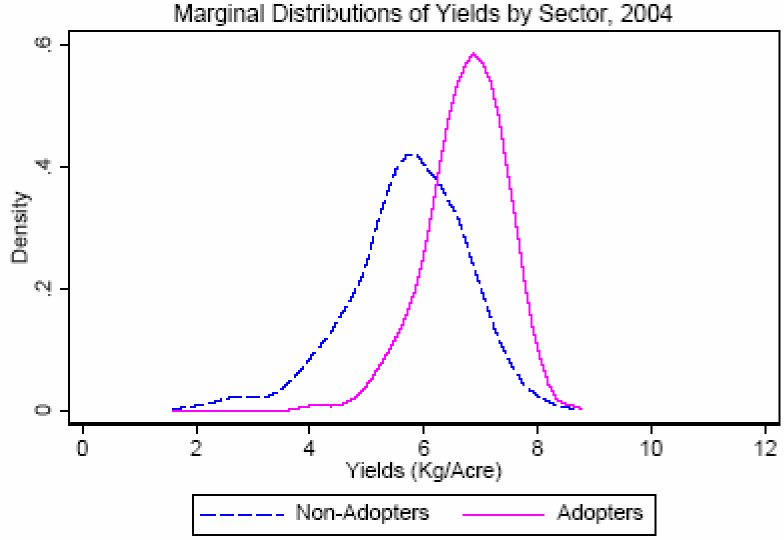

Figure 7a

Differences in Yields (Perfect Positive Dependence), 1997

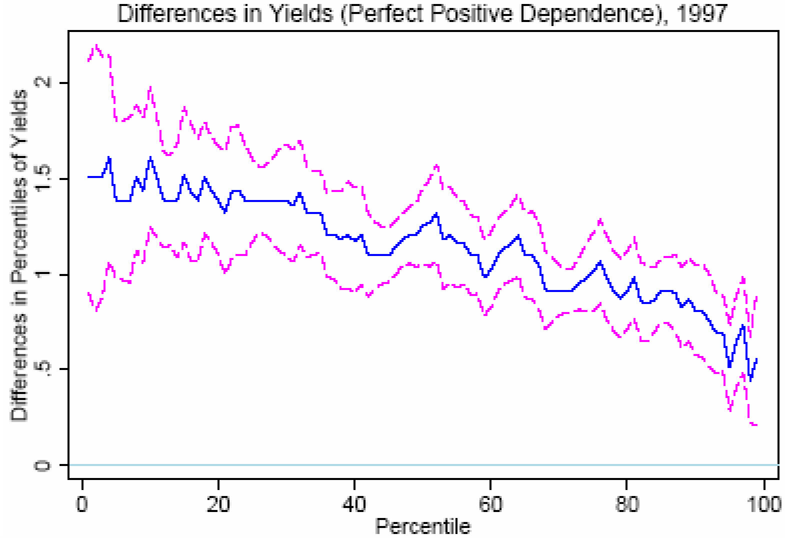

Figure 7b

Differences in Yields (Perfect Positive Dependence), 2000

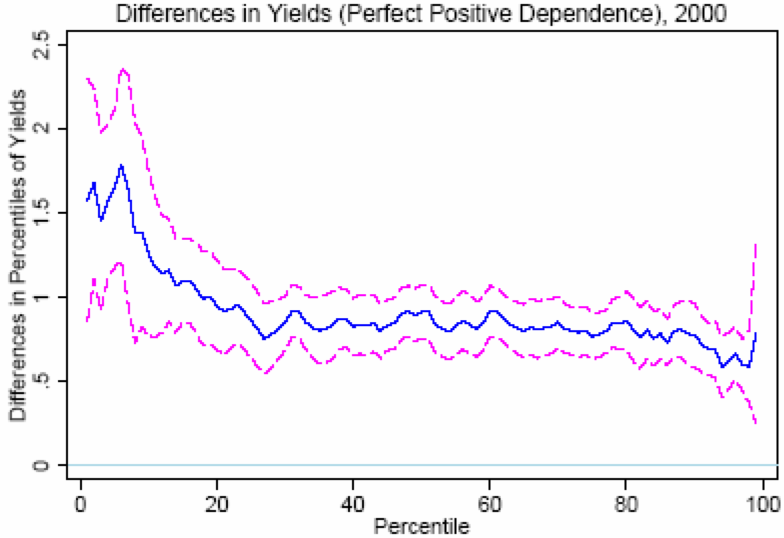

Figure 7c

Differences in Yields (Perfect Positive Dependence), 2004

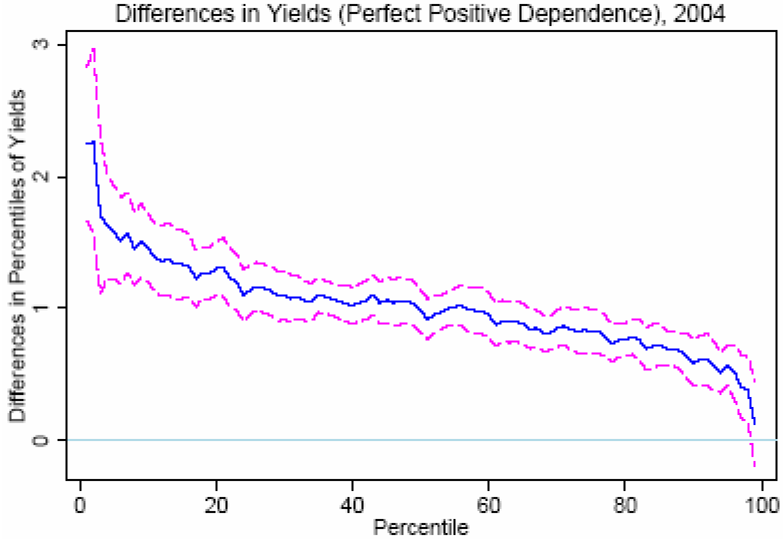

Figure 8a

Differences in Yields for Joiners (Positive)

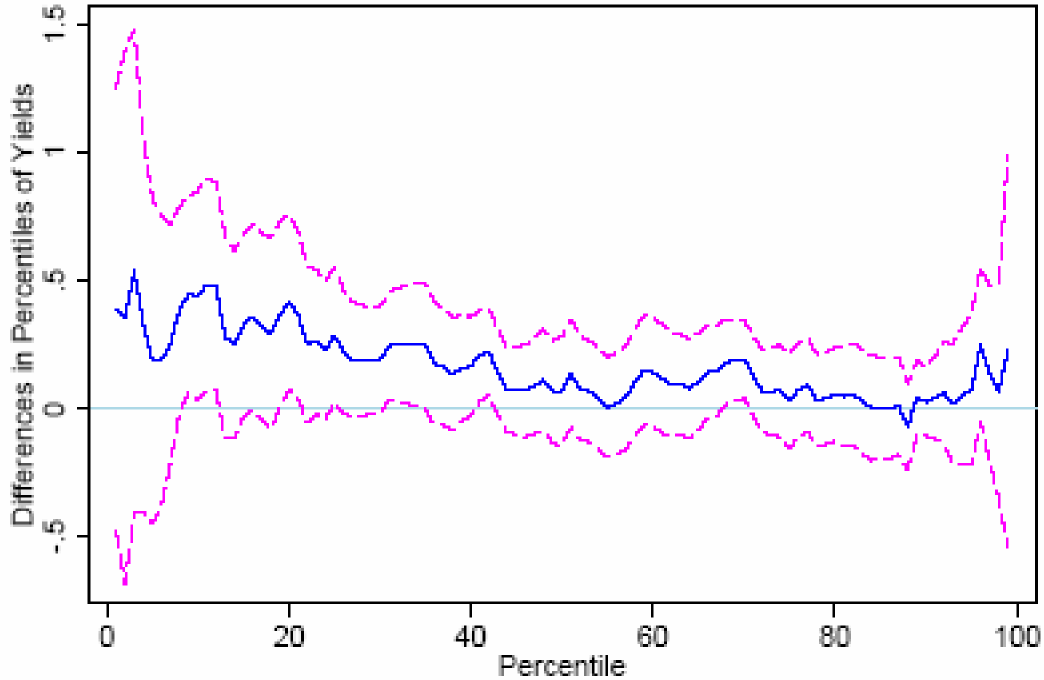

Figure 8b

Differences in Yields for Leavers (Positive)

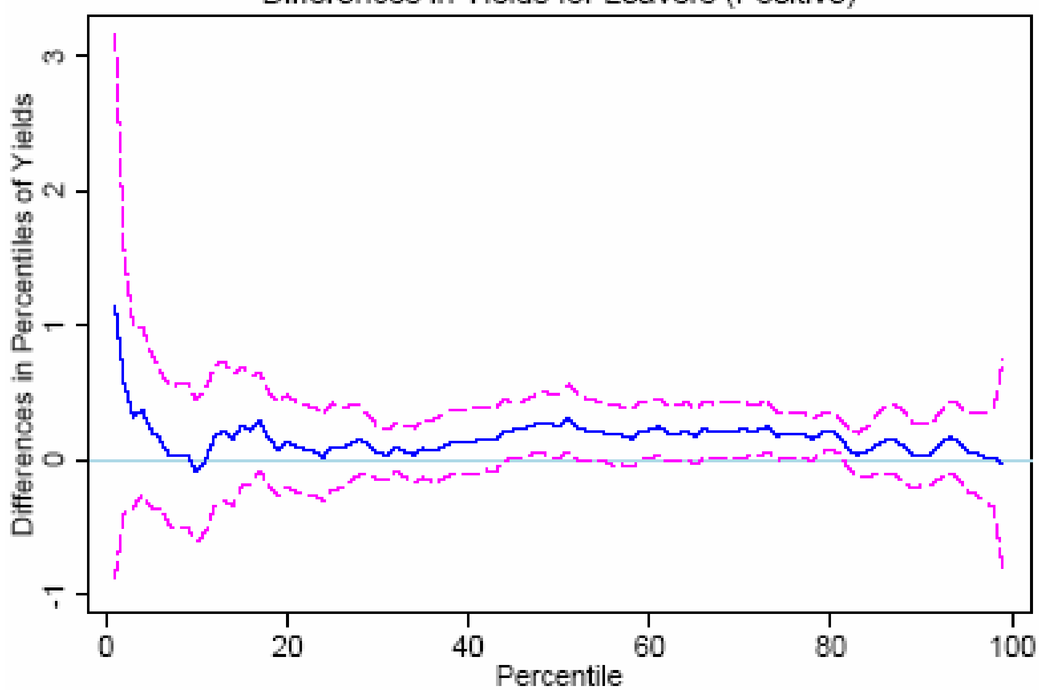

Figure 9a
Distribution of Comparative Advantage

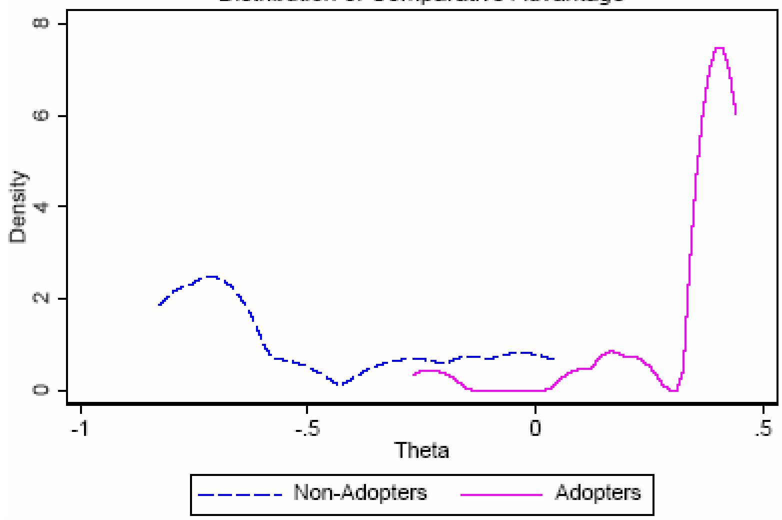

Figure 9b
Distribution of Tau

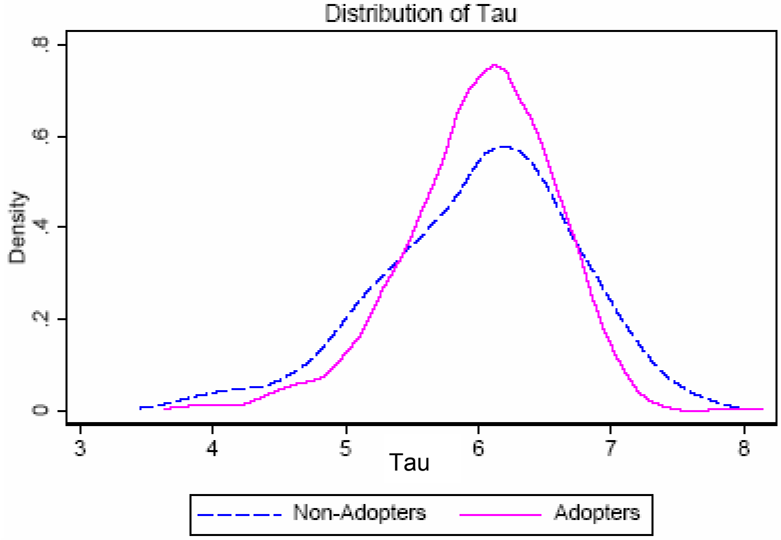

Figure 9c
Distribution of Alpha

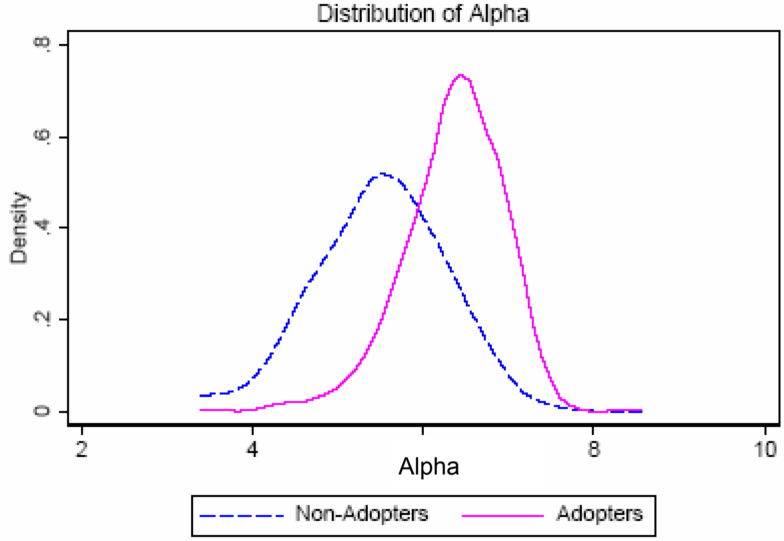

Table 1
Changes in Yields (Hg/Ha), by Decade

	1961-1970	1971-1980	1981-1990	1991-2004
Kenya				
Maize	-0.251	-0.159	0.009	-0.279
Wheat	5.323	3.029	-1.462	-0.734
India				
Maize	3.744	3.194	3.400	3.815
Wheat	4.675	1.097	3.349	1.208
Rice	1.030	1.878	3.682	0.878
Mexico				
Maize	2.238	4.836	1.037	1.682
Wheat	8.914	3.230	1.496	0.628

Source: FAOSTAT Online Database

Table 2a: Summary Statistics

	1997 Sample	2000 Sample	2004 Sample
Yield (Log Maize Harvest Per Acre)	5.907 (1.153)	6.125 (1.092)	6.350 (0.977)
Acres Planted	1.903 (3.217)	2.313 (3.948)	1.957 (2.685)
Total Seed Planted (Kg per Acre)	9.575 (7.801)	9.331 (6.805)	9.072 (6.863)
Total Purchased Hybrid Planted (Kg per Acre)	6.273 (6.926)	5.918 (6.636)	5.080 (5.260)
Total Local Seed Planted (Kg per Acre)	2.653 (6.326)	2.868 (6.129)	3.120 (7.318)
Hybrid (dummy)	0.658 (0.475)	0.676 (0.468)	0.604 (0.489)
Fertilizer (Kg DAP per Acre)	20.300 (38.444)	24.577 (79.919)	24.610 (34.001)
Fertilizer (Kg MAP per Acre)	1.566 (10.165)	0 (0)	0.308 (4.538)
Fertilizer (Kg CAN per Acre)	6.473 (24.727)	9.819 (75.081)	8.957 (21.702)
Fertilizer (Kg NPK per Acre)	4.256 (20.096)	2.365 (12.172)	1.217 (7.870)
Total Fertilizer Expenditure (KShs per Acre)	1361.7 (2246.3)	1346.5 (3347.1)	1354.6 (1831.2)
Land Preparation Costs (Kshs per Acre)	2168.1 (4895.7)	813.66 (1222.9)	808.99 (1080.2)
Main Season Rainfall (mm)	620.83 (256.43)	599.10 (264.21)	728.11 (293.29)

Notes: Standard deviations in parentheses. KShs represents Kenyan shillings (the current exchange rate is on the order of 75 shillings to the US dollar). DAP is di-ammonium phosphate, MAP is mono-ammonium phosphate, CAN is calcium ammonium nitrate and NPK is nitrogen phosphorus potassium.

Table 2b: Summary Statistics, by Sector

	1997 Sample		2000 Sample		2004 Sample	
	Hybrid	**Non-Hybrid**	**Hybrid**	**Non-Hybrid**	**Hybrid**	**Non-Hybrid**
No. of Households	791	411	813	389	726	476
Yield (Log Maize Harvest Per Acre)	6.296 (0.934)	5.158 (1.167)	6.423 0.948)	5.503 (1.113)	6.751 (0.692)	5.738 (1.030)
Total Maize Acres Cultivated	1.982 (3.557)	1.753 (2.428)	2.453 (4.334)	2.021 (2.970)	2.087 (3.029)	1.758 (2.042)
Total Seed Planted (Kg per Acre)	9.669 (6.569)	9.394 (9.750)	20.628 (41.724)	9.926 (7.507)	8.746 (4.156)	9.569 (9.608)
Fertilizer (Kg DAP per Acre)	28.755 (44.115)	4.028 (13.266)	34.068 (95.214)	4.740 (14.638)	37.148 (37.294)	5.488 (13.909)
Fertilizer (Kg CAN per Acre)	9.087 (29.715)	1.442 (7.152)	13.781 (90.809)	1.537 (9.479)	12.708 (24.961)	3.235 (13.622)
Land Preparation Costs (KShs per Acre)	2006.0 (1316.2)	2480.1 (8168.4)	909.99 (1318.9)	614.43 (964.13)	918.44 (1095.2)	642.06 (1036.1)
Total Expenditure on Fertilizer (KShs per Acre)	1922.3 (2542.9)	282.64 (740.53)	1848.2 (3941.6)	297.80 (730.18)	1893.3 (1964.7)	533.09 (1211.4)
Main Season Rainfall (mm)	651.70 (228.82)	561.44 (293.88)	611.67 (270.44)	572.83 (249.00)	825.41 (215.20)	579.69 (332.05)

Notes: Standard deviations in parentheses.

Table 3
Basic OLS and Fixed Effects Specifications
Dependent Variable is Yields (Log Maize Harvest Per Acre)

	OLS, Pooled with Year Dummies	OLS, with Year & Province Dummies	OLS, with Covariates	FE, with Year Dummies	FE, with Covariates
Hybrid	1.024	0.731	0.556	0.114	0.149
	(0.033)	(0.034)	(0.035)	(0.052)	(0.049)
Year = 1997	-0.199	-0.205	-0.272	-0.216	-0.289
	(0.039)	(0.037)	(0.035)	(0.033)	(0.032)
Year = 2004	0.299	0.278	0.131	0.233	0.089
	(0.039)	(0.037)	(0.038)	(0.033)	(0.036)
Constant	5.432	5.046	4.466	6.048	-1.951
	(0.036)	(0.072)	(0.085)	(0.042)	(4.931)
R-Squared	0.228	0.317	0.424	0.071	0.197
No. of Obs	3606	3606	3606	3606	3606

Notes: Throughout, standard errors are in parentheses. Omitted year is 2000.
Henceforth, covariates include real expenditure on fertilizers, land preparation costs, seed quantities, labor variables, province dummies, year dummies and rainfall variables (average seasonal rainfall and current seasonal rainfall).

Table 4: CRE Model
Standard CRE Model Reduced Forms and Structural Estimates

	Without Covariates			With Covariates		
Dependent Variable:	Yields, 1997	Yields, 2000	Yields, 2004	Yields, 1997	Yields, 2000	Yields, 2004
Hybrid, 1997	0.574 (0.075)	0.345 (0.084)	0.467 (0.067)	0.401 (0.168)	0.252 (0.071)	0.351 (0.057)
Hybrid, 2000	0.323 (0.081)	0.424 (0.086)	0.230 (0.074)	0.233 (0.071)	-0.022 (0.176)	0.161 (0.058)
Hybrid, 2004	0.680 (0.080)	0.490 (0.081)	0.632 (0.070)	0.407 (0.071)	0.360 (0.073)	0.292 (0.147)
R-squared	0.311	0.218	0.312	0.448	0.325	0.473

	Standard CRE Model Structural Estimates				
	β	λ_1	λ_2	λ_3	$\chi^2(5)$
OMD, Without covariates	0.071 (0.054)	0.498 (0.053)	0.281 (0.056)	0.502 (0.059)	100.87
EWMD, Without covariates	0.117 (0.054)	0.475 (0.053)	0.275 (0.056)	0.503 (0.059)	-
OMD, With covariates	0.147 (0.046)	0.362 (0.042)	0.208 (0.044)	0.328 (0.045)	16.315
EWMD, With covariates	0.168 (0.046)	0.344 (0.042)	0.210 (0.044)	0.338 (0.045)	-

Table 5: Selection
Returns by Hybrid History (Joiners, Leavers and Stayers)
Dependent Variable is Yields (Log Maize Harvest Per Acre)

Variable	**1997-2000**		**2000-2004**		**1997-2004**	
	Yield 1	**Yield 2**	**Yield 1**	**Yield 2**	**Yield 1**	**Yield 2**
Without Covariates:						
Hybrid Stayers	1.411	1.109	1.132	1.172	1.505	1.280
	(0.070)	(0.070)	(0.067)	(0.057)	(0.066)	(0.056)
Leavers	0.746	0.281	0.426	0.332	0.809	0.648
	(0.112)	(0.110)	(0.095)	(0.081)	(0.094)	(0.079)
Joiners	0.563	0.430	0.403	0.693	1.007	0.883
	(0.105)	(0.104)	(0.127)	(0.108)	(0.114)	(0.096)
With Covariates:						
Hybrid Stayers	0.766	0.637	0.719	0.593	0.868	0.652
	(0.073)	(0.075)	(0.073)	(0.061)	(0.074)	(0.063)
Leavers	0.466	0.044	0.367	0.159	0.509	0.350
	(0.097)	(0.100)	(0.086)	(0.067)	(0.085)	(0.069)
Joiners	0.316	0.335	0.258	0.339	0.491	0.499
	(0.089)	(0.094)	(0.115)	(0.092)	(0.101)	(0.084)

Notes: Standard errors in parentheses. A hybrid stayer is defined as a farmer who plants hybrid in both periods under consideration. A leaver is a farmer who plants hybrid the first period, not the second. A joiner is a farmer who plants traditional varieties the first period and hybrid the second. The omitted group is non-hybrid stayers, i.e. farmers who plant traditional varieties both years.

Table 6: Heterogeneity by Observables
Returns in the Hybrid/Non-Hybrid Sector
Dependent Variable is Yields (Log Maize Harvest Per Acre)

Variable	OLS, with Covariates		FE, with Covariates	
	Hybrid	Non-Hybrid	Hybrid	Non-Hybrid
Acreage (x 100)	0.784	-5.427	-4.353	-8.360
	(0.438)	(1.125)	(0.917)	(2.013)
Total Seed Planted (Kg per Acre)	3.418	2.176	2.854	2.397
(x 100)	(0.279)	(0.307)	(0.323)	(0.401)
Land Preparation Costs per Acre (KShs)	0.487	0.216	0.523	0.151
(x 10,000)	(0.129)	(0.031)	(0.170)	(0.068)
Total Fertilizer Expenditure per Acre	0.430	1.855	0.245	1.176
(KShs) (x 10,000)	(0.055)	(0.305)	(0.064)	(0.534)
Main Season Rainfall (mm)	3.879	12.286	9.297	12.706
(x 10,000)	(1.399)	(2.261)	(1.416)	(2.599)
Average Main Season Rainfall	1.551	0.458	-7.527	12.557
(x 1,000)	(0.227)	(0.625)	(19.43)	(11.152)
Number of Observations	2330	1276	2330	1276

Notes: Standard errors in parentheses. All regressions include covariates, as described earlier. The specifications are run separately for the hybrid and non-hybrid sectors.

Table 7: Comparative Advantage CRC Two Period Model Structural Estimates

	With Only Hybrid is Endogenous			
	Without Covariates		With Covariates	
	EWMD	OMD	EWMD	OMD
λ_1	0.786 (0.068)	0.743 (0.068)	0.648 (0.059)	0.596 (0.059)
λ_2	0.963 (0.093)	1.030 (0.093)	0.711 (0.085)	0.772 (0.085)
λ_3	-5.4185 (30.55)	-1.622 (0.724)	-1.098 (1.284)	-0.930 (0.256)
β	3.034 (16.76)	1.008 (0.274)	0.804 (0.324)	0.733 (0.048)
φ	-1.104 (0.700)	-1.558 (0.533)	-1.740 (2.235)	-1.966 (0.906)
χ^2	-	71.63	-	63.75

	With Both Fertilizer and Hybrid as Endogenous			
	Without Covariates		With Covariates	
	EWMD	OMD	EWMD	OMD
φ	-0.468 (0.145)	-0.824 (0.224)	-0.783 (0.691)	-1.807 (0.460)
χ^2	-	64.70	-	110.32

Notes: Standard errors in parentheses.

Table 8a: Comparative Advantage CRC Three Period Model
Reduced Form Estimates

	Without Covariates			With Covariates		
Dependent Variable:	Yields, 1997	Yields, 2000	Yields, 2004	Yields, 1997	Yields, 2000	Yields, 2004
Hybrid, 1997	0.687 (0.144)	0.174 (0.150)	0.574 (0.123)	0.501 (0.124)	0.139 (0.131)	0.439 (0.106)
Hybrid, 2000	0.175 (0.136)	0.232 (0.123)	0.115 (0.132)	0.100 (0.127)	0.440 (0.134)	0.058 (0.209)
Hybrid 2004	0.718 (0.197)	0.284 (0.184)	0.639 (0.159)	0.309 (0.166)	0.244 (0.175)	0.365 (0.142)
Hybrid 1997*2000	0.122 (0.218)	0.258 (0.214)	0.069 (0.197)	0.045 (0.176)	0.143 (0.185)	0.103 (0.150)
Hybrid 1997*2004	-0.294 (0.276)	0.126 (0.302)	-0.225 (0.232)	-0.061 (0.230)	-0.035 (0.242)	-0.215 (0.197)
Hybrid 2000*2004	0.344 (0.246)	0.237 (0.259)	0.312 (0.217)	0.392 (0.213)	0.011 (0.225)	0.197 (0.182)
Hybrid 1997* 2000*2004	-0.180 (0.336)	-0.098 (0.371)	-0.158 (0.295)	-0.350 (0.289)	-0.046 (0.205)	-0.241 (0.247)
Constant	4.890 (0.076)	5.389 (0.069)	5.496 (0.065)	3.999 (0.133)	4.339 (0.140)	4.779 (0.114)
R-squared	0.315	0.222	0.316	0.489	0.367	0.479
No. of Observations	1202	1202	1202	1202	1202	1202

Notes: Standard errors in parentheses. Reduced form yield functions include the entire history of the hybrid variable, as well as all the possible interactions.

Table 8b: Comparative Advantage CRC Model
Structural Estimates for Hybrid Decision

	Comparative Advantage CRC Model			
	Without Covariates		**With Covariates**	
	OMD	**EWMD**	**OMD**	**EWMD**
λ_1	0.486 (0.056)	0.414 (0.056)	0.361 (0.048)	0.247 (0.045)
λ_2	0.168 (0.059)	0.133 (0.059)	0.165 (0.049)	0.140 (0.048)
λ_3	0.537 (0.109)	0.476 (0.099)	0.275 (0.094)	0.232 (0.067)
λ_4	0.072 (0.131)	0.058 (0.124)	0.081 (0.111)	-0.008 (0.093)
λ_5	-0.154 (0.157)	-0.154 (0.140)	-0.115 (0.132)	-0.142 (0.090)
λ_6	0.257 (0.159)	0.135 (0.145)	0.221 (0.137)	0.078 (0.112)
λ_7	-0.202 (0.210)	-0.108 (0.189)	-0.259 (0.178)	-0.116 (0.138)
β	0.123 (0.057)	0.290 (0.054)	0.093 (0.050)	0.295 (0.042)
φ	0.175 (0.127)	0.396 (0.176)	-0.144 (0.124)	0.685 (0.349)
$\chi^2(12)$	105.79	-	95.04	-

Notes: Standard errors in parentheses.

Table 9: Comparative Advantage CRC Model
Structural Estimates for Joint Hybrid-Fertilizer Decision

	Comparative Advantage CRC Model			
	Without Covariates		With Covariates	
	OMD	**EWMD**	**OMD**	**EWMD**
λ_1	0.773 (0.070)	0.650 (0.063)	0.465 (0.055)	0.441 (0.055)
λ_2	0.469 (0.074)	0.465 (0.070)	0.225 (0.057)	0.245 (0.057)
λ_3	0.621 (0.069)	0.614 (0.060)	0.216 (0.066)	0.206 (0.065)
λ_4	-0.388 (0.148)	-0.379 (0.126)	-0.165 (0.124)	-0.235 (0.124)
λ_5	-0.350 (0.117)	-0.352 (0.104)	-0.209 (0.122)	-0.233 (0.121)
λ_6	-0.217 (0.125)	-0.338 (0.111)	-0.116 (0.101)	-0.278 (0.096)
λ_7	0.168 (0.186)	0.234 (0.161)	-0.080 (0.176)	0.059 (0.185)
β	0.095 (0.052)	0.133 (0.072)	0.128 (0.053)	0.183 (0.070)
φ	0.229 (0.171)	0.828 (0.362)	-0.327 (0.170)	-0.246 (0.269)
$\chi^2(12)$	45.66	-	151.40	-

Notes: Standard errors in parentheses.

Table 10: Treatment Effect Estimates:
IV (LATE), Heckit, Selection Corrected (ATE, TT, MTE) Treatment Effects

Year	Heckman Two-Step Estimates		Implied Treatment Effects		
	Selection Correction λ Hybrid Sector	Selection Correction λ Non-Hybrid Sector	ATE	TT	MTE Slope
1997	-0.871	0.897	2.097	1.374	-1.769
	(0.171)	(0.756)			(0.775)
2000	-0.861	-0.639	0.756	0.677	-0.222
	(0.119)	(0.350)			(0.370)
2004	-0.910	0.001	1.346	0.830	-0.911
	(0.192)	(0.162)			(0.251)

IV Estimates (Conditional on Covariates)	
First Stage: Effect of Distance to Closest Fertilizer Seller (km) on Probability of Using Hybrid	Second Stage: Effect of Predicted Hybrid (from the First Stage) on Yields (LATE)
-0.516	1.543
(0.097)	(0.434)

Notes: Standard errors in parentheses.

Table 11: OLS and FE Estimates by Predicted Fertility Quantile

	Predicted Fertility Quantile			
	1	2	3	4
OLS Estimate	0.418 (0.082)	0.403 (0.072)	0.473 (0.062)	0.347 (0.077)
FE Estimate	0.150 (0.105)	0.267 (0.099)	0.155 (0.082)	0.056 (0.099)
Hybrid Prevalence	31.01	59.00	78.85	89.67
Number of Observations	903	900	903	900

Notes: Standard errors in parentheses.

Table A1
Release of Hybrid Maize and Recommended Applications of Fertilizer By Agroclimatic Zone

Zone	Improved Maize Seed (Hybrid and/or OPV)			Recommendations		
	Recommended Variety	Year of Release	Expected Yield	Nitrogen	P_2O_5	Alternatives
UM0-1, Upper Midlands	H614D	1986	75-100b/ha	60kg/ha	60kg/ha	130kg DAP + 141kg CAN or ASN + 130kg TSP
	H624	-				
	H625	1981				
	H626	1989				
	H627	1996/7				
LM1-2, Lower Midlands	H614D	1986	37-50b/ha	40kg/ha	40kg/ha	110kg + 120kg CAN
	H622	1963/5				
	H512	1970				
	H511	1963/8				
Coastal Lowlands	Coast Composite	1974	3.8t/ha	60kg/ha	46kg/ha	
	Pwani Hybrid 1	1989	4.8t/ha			
	Pwani Hybrid 4	1997	5.4t/ha			
Coffee Dairy Zone (UM2-3)	H513	1996/7	1.8t/ha	50kg/ha	50kg/ha	
	C5222	1996/7	1.8t/ha			
	PAN5195	1996/7	1.8t/ha			
	PHB3253	1996/7	1.8t/ha	Plus top dress fertilizer at the rate of 50kg/ha N and farmyard manure at the rate of 5t/ha		
	CG4141	1996/2000	1.4t/ha			
	H512	1970	1.8t/ha			
	H511	1968	1.5t/ha			
	EMCO92SR	-	1.5t/ha			
Maize Sunflower Zone (UM4/LM3-4)	DH1 (dryland hybrid)	1996/7	1.2t/ha			
	DH2	1996/7	1.2t/ha			
	Makueni Composite	1989	1.1t/ha			
	Katumani Composite B	1968	1.1t/ha			
	CG4141	1996/2000	1.2t/ha			

Source: Ouma et al (2002), Hassan (1998), Salasya et al (1998), Wekesa et al (2003), Kamau (2002) and Karanja (1996).
Additional Varieties (Yr Released): Kitale Synthetic II (1961), Katumani Synthetic II (1963), Katumani Composite A (1966), H611 (1964), H621 (1964), H631 (1964), H611C (1971), H612C (1966), H613C (1972), H614C (1976), H612D (1986), H613D (1986), H632 (1965), H525 (1981), KTSP94 (2000), KH60011D (2000), KH634A (2001), KH60015A (2001), KH60016A (2001), CG5051 (2000), PAN5355 (2000), H623 (2000), FS650 (2001), H6212 (2001), H6211 (2001), PAN99 (2001), PAN5243 (2001), PAN67 (2001), PHB30A15 (2001), H516 (2001), DH04 (2001), DH05 (2001), H628 (2001), PAN691 (2001), KSH516 (2001)

Table A2: Impact Distribution Percentiles

	Perfect Positive Dependence					Perfect Negative Dependence				
Percentile	**Cross Section Results**			**Fixed Effects**		**Cross Section Results**			**Fixed Effects**	
	1997	**2000**	**2004**	**Joiners**	**Leavers**	**1997**	**2000**	**2004**	**Joiners**	**Leavers**
5	1.386	1.644	1.579	0.186	0.218	-2.485	-2.227	-1.743	-3.015	-3.183
	(0.191)	(0.222)	(0.167)	(0.318)	(0.287)	(0.106)	(0.155)	(0.102)	(0.204)	(0.211)
10	1.609	1.260	1.458	0.440	-0.084	-1.358	-1.434	-1.101	-2.304	-2.658
	(0.186)	(0.255)	(0.132)	(0.204)	(0.265)	(0.149)	(0.112)	(0.095)	(0.179)	(0.217)
25	1.386	0.862	1.139	0.287	0.105	0.000	-0.386	-0.113	-1.071	-1.174
	(0.111)	(0.113)	(0.097)	(0.131)	(0.161)	(0.090)	(0.056)	(0.077)	(0.126)	(0.114)
50	1.253	0.916	0.976	0.068	0.258	1.253	0.916	0.976	0.068	0.258
	(0.099)	(0.077)	(0.083)	(0.106)	(0.121)	(0.085)	(0.067)	(0.072)	(0.106)	(0.121)
75	1.012	0.768	0.827	0.028	0.191	2.398	2.016	2.079	1.386	1.471
	(0.104)	(0.065)	(0.084)	(0.092)	(0.085)	(0.101)	(0.107)	(0.084)	(0.103)	(0.143)
90	0.811	0.777	0.584	0.027	0.043	3.778	3.470	3.143	2.771	2.617
	(0.111)	(0.093)	(0.088)	(0.066)	(0.114)	(0.137)	(0.235)	(0.111)	(0.131)	(0.214)
95	0.511	0.624	0.560	0.074	0.065	4.382	4.495	3.881	3.275	3.466
	(0.106)	(0.083)	(0.087)	(0.146)	(0.146)	(0.177)	(0.168)	(0.143)	(0.282)	(0.270)
Impact Std Deviation	0.261	0.240	0.340	0.129	0.138	1.988	1.885	1.634	1.867	1.986
	(0.043)	(0.057)	(0.045)	(0.043)	(0.064)	(0.049)	(0.056)	(0.043)	(0.076)	(0.087)
Outcome Correlation	0.994	0.993	0.997	0.996	0.992	-0.943	-0.907	-0.969	-0.947	-0.942
	(0.004)	(0.004)	(0.002)	(0.010)	(0.005)	(0.011)	(0.014)	(0.009)	(0.026)	(0.016)

Notes: The impact standard deviation is the standard deviation of the percentile differences. The outcome correlation is the correlation of the two percentiles of the two distributions. Bootstrap standard errors are in parentheses.