

NBER WORKING PAPER SERIES

SELECTION AND COMPARATIVE ADVANTAGE IN TECHNOLOGY ADOPTION

Tavneet Suri

Working Paper 15346

<http://www.nber.org/papers/w15346>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

September 2009

I am extremely grateful to Michael Boozer for innumerable discussions on this project. A special thanks to Tom Stoker, Joe Doyle, Billy Jack, Gustav Ranis, Paul Schultz and Christopher Udry. I would also like to thank Ken Chay, Robert Evenson, Michael Greenstone, Penny Goldberg, Koichi Hamada, Thomas Jayne, Fabian Lange, Ashley Lester, Anandi Mani, Sharun Mukand, Ben Polak, Steve Pischke, Roberto Rigobon, James Scott, T.N. Srinivasan, and seminar audiences at Berkeley Economics Department, Haas School of Business, Harvard/MIT, Hunter College, London School of Economics, the NBER Productivity Lunch, NEUDC, Oxford, Princeton, Sloan School of Management, Stanford, Stanford Graduate School of Business, University of California at San Diego, University of Virginia and Yale for all their comments. 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. A special thanks to Thomas Jayne and Margaret Beaver at Michigan State University for all their help. I am sincerely grateful to the Tegemeo Institute and Thomas Jayne for including me in the 2004 survey and to Tegemeo for their hospitality during the ...eld work. I would like to thank James Nyoro, Director of Tegemeo, as well as Tegemeo Research Fellows Milton 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 enumeration teams. An additional thanks to Margaret Beaver and Daniel Kariuki for their work on cleaning the data. The rainfall data come from the Climate Prediction Center, part of the USAID/FEWS project - a special thanks to Tim Love for his help with these data. I would like to acknowledge the financial support of the Yale Center for International and Area Studies Dissertation Fellowship, the Lindsay Fellowship and the Agrarian Studies Program at Yale. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

© 2009 by Tavneet Suri. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Selection and Comparative Advantage in Technology Adoption
Tavneet Suri
NBER Working Paper No. 15346
September 2009
JEL No. C33,O12,Q12

ABSTRACT

This paper investigates an empirical puzzle in technology adoption for developing countries: the low adoption rates of technologies like hybrid maize that increase average farm profits dramatically. I offer a simple explanation for this: benefits and costs of technologies are heterogeneous, so that farmers with low net returns do not adopt the technology. I examine this hypothesis by estimating a correlated random coefficient model of yields and the corresponding distribution of returns to hybrid maize. This distribution indicates that the group of farmers with the highest estimated gross returns does not use hybrid, but their returns are correlated with high costs of acquiring the technology (due to poor infrastructure). Another group of farmers has lower returns and adopts, while the marginal farmers have zero returns and switch in and out of use over the sample period. Overall, adoption decisions appear to be rational and well explained by (observed and unobserved) variation in heterogeneous net benefits to the technology.

Tavneet Suri
MIT Sloan School of Management
50 Memorial Drive
E52-457
Cambridge, MA 02142
and NBER
tavneet@mit.edu

1 Introduction

Food security is a major social and economic issue across many Sub-Saharan African economies. Agricultural yields (output per acre) have fallen in the last decade across many of these economies, despite the widespread availability of technologies that increase yields. For instance, Table 1 shows the falling yields of staple crops over the last decade in Kenya, compared with increasing yields in India and Mexico. Kenya has clearly not been able to take the same advantage of improvements in agricultural technologies as have India and Mexico during their Green Revolutions. This is not an isolated example, and is very worrisome for economic policies intended to enhance food production and agricultural incomes.

The leading edge of agricultural technology has improved over the past few decades. Field trials at experiment stations across Kenya have shown that hybrid maize and fertilizers can increase yields and profits of maize significantly, with increases ranging from 40% to 100% (see Gerhart (1975), Kenya Agricultural Research Institute, KARI (1993), Karanja (1996)).² Despite this, the aggregate adoption rates of hybrid maize and fertilizer remain far below 100%, with no sustained increases in adoption over the past decade. This contrasts with the experience in more developed countries; for example, adoption rates of hybrid in the US are close to 100% (Griliches (1980) and Dixon (1980)).

The fact that improved agricultural technologies are not universally adopted, and that adoption rates have remained persistently low over a long period of time, is the empirical puzzle that motivates this paper. I find that if farmer heterogeneity is properly taken into account, there is no puzzle. The farmers with high net returns to the technology adopt it, and the farmers with low returns do not. Persistent lack of adoption is a reflection of the distribution of (observable and unobservable) costs and benefits of the technology. The approach here models households' adoption decisions in an environment with household specific heterogeneity in the costs and benefits, and hence profits, to the technology. I estimate how the returns to the technology vary across farmers and then compare these returns to the adoption decisions of farmers. I find that farmers with low (or zero) returns to the technology are precisely the farmers who do not adopt it.

More specifically, I use a generalized Roy model along the lines of Heckman and Vytlacil (1997) where expected profits from hybrid maize are heterogeneous and drive adoption decisions. This theoretical model implies a yield function for maize with a correlated random coefficient structure, which is estimated using a generalization of Chamberlain's (1982, 1984) method for fixed effects models, applied here to a correlated random coefficients model. The estimation results imply a distribution of returns that can endogenously affect adoption de-

²The Fertilizer Use Recommendations Project (FURP) in Kenya also shows high returns to hybrid maize and fertilizer (see Hassan et al (1998)). These results are confirmed for fertilizer in the small sample of Duflo, Kremer and Robinson (2008a) for Western Kenya. In fact, Duflo, Kremer and Robinson (2008b) analyze the rates of return to fertilizer and show high, but variable, returns to fertilizer.

cisions. I therefore provide an empirical resolution of the puzzle that, despite high average returns to hybrid use, the marginal returns are low, and, given the infrastructure constraints faced by farmers, adoption decisions are on the whole rational.

The estimation strategy allows for two different forms of household-specific heterogeneity in maize production: absolute advantage (independent of the technology used) and comparative advantage, which measures the relative productivity of a farmer in hybrid over non-hybrid. The novel econometric contribution of this paper is to analyze the role of comparative advantage in adoption decisions, to empirically estimate its importance, and to estimate its distribution. I also discuss why this approach is preferable in development settings to related approaches used to study firms in developed countries, where the evolution of firm efficiencies is more likely to be important in consistently estimating output and investment equations.

I use an extensive panel dataset representative of maize-growing Kenya for the period 1996 to 2004. I first document the adoption patterns of households for both hybrid maize and fertilizer. Aggregate adoption is stable over the period, but, surprisingly, at least 30% of households switch into and out of hybrid use from period to period.³ The panel structure of the data allows me to identify the distribution of returns to hybrid maize and to estimate the correlated random coefficient structure of the underlying yield functions. The panel aspect also allows me to construct counterfactual distributions of returns for all the farmers in the sample, including those who do not use hybrid. It is important for identification that some farmers switch in and out of use of the technology, and also that there are farmers in the sample who use hybrid maize in every period. I discuss how the identification assumptions used are supported by the timing of maize production, other characteristics of the production process and the fact that shocks to yields like rainfall are observed in the data.

I find strong evidence of heterogeneity in the returns to hybrid maize, especially in the sample excluding the two districts (out of a total of twenty two) with high adult mortality. I estimate the distribution of returns, and compare the mean of this distribution to estimates from standard approaches that fail to take full account of farmer heterogeneity, such as OLS, IV and fixed effects models.⁴ OLS estimates of the average return are in the range of 50-100%, models with fixed effects suggest returns of close to 0% and IV estimates (using supply constraints as instruments) are on the order of 150-200%. The estimated mean gross return from my approach is 60%, but some farmers have returns as high as 150%, while there are many who have returns either close to zero or (in some cases) negative.

The estimated returns control for input use, but do not account for other costs (such as

³This lack of a trend with underlying switching of use from period to period is seen in other parts of Africa. Dercon and Christiaensen (2005) find identical patterns for fertilizer use in Ethiopia. Duflo, Kremer and Robinson (2008b) also find a lot of switching behavior in fertilizer use in Western Kenya.

⁴When there is heterogeneity in returns, IV estimates a local average treatment effect, i.e. the returns for only the sub-population that is affected by the instrument. This could explain why OLS and IV are so different here and why different instruments may give different estimates.

the costs of accessing the technology) and are therefore gross, rather than fully net returns. The joint distribution of estimated returns and adoption decisions displays some remarkable features. There are three sub-groups of farmers in the sample. A small group of farmers has extremely high counterfactual returns to hybrid (about 150%), yet they choose not to adopt. This is rather striking and seems to deepen the initial puzzle, but is well explained by supply/infrastructure constraints, such as long distances to seed and fertilizer distributors. These farmers have higher costs to using hybrid as they have poor access to input suppliers. Their overall net counterfactual returns are therefore rather low. A comparatively larger group of farmers have lower, though still high, returns and they adopt hybrid in every period. Finally, a third group of farmers has essentially zero net returns and they switch in and out of use of hybrid from period to period. These are marginal farmers who switch easily when subject to shocks to the cost of and access to hybrid seed and fertilizer.

The heterogeneity in returns to this technology has important implications for policy. For one, encouraging complete adoption of a technology that has a large average return among existing adopters may be very inefficient due to the much lower returns for the non-adopters. In addition, knowing the distribution of returns to a technology allows for focused policy interventions that can be cost effective. For example, for farmers who would have high returns but are constrained on the supply side, alleviating their constraints by targeted distribution of inputs and infrastructure improvements could improve yields dramatically. Similarly, the unconstrained farmers would benefit from the development of new hybrid strains. This research also illustrates the importance of the distributional consequences of policy and sheds light on the ‘scaling up’ of policy (see Attanasio et al (2003)).⁵

The literature provides a number of explanations for low adoption, ranging from learning models (Foster and Rosenzweig (1995)), informational barriers, credit constraints, taste preferences, differences in agroecological conditions and local costs and benefits, to time inconsistency and the lack of effective commitment devices (see Duflo, Kremer and Robinson (2008a)). The dominant approach has been to frame adoption decisions in a learning environment where benefits and costs to the technologies are homogeneous but unknown, and are learned over time. The approach in this paper is in contrast to much of the literature. Since these technologies have been available for many years and are well known and understood⁶ (as in Duflo, Kremer and Robinson (2008a, 2008b)), the approach here assumes the benefits and costs are known *ex ante*, but are spatially heterogeneous across farmers.

⁵The questions posed here cannot be answered with an experimental which randomizes the technology across farmers without specific assumptions. By contrast, I am interested in how farmers choose what technologies to use. With experimental data, one can test for the presence of heterogeneous returns, but estimating the distribution of returns requires assumptions about the underlying selection process, which is randomized away in such an experiment (see Heckman, Smith and Clements (1997)).

⁶For example in my sample, about 90% of households have used hybrid maize at some point in the past. The mean number of years since first use for hybrid maize is 19.4 years and for fertilizer 20.5 years.

The rest of this paper is structured as follows. Section 2 outlines some relevant empirical literature. Section 3 describes the institutional context and the data. Section 4 discusses a theoretical framework for adoption decisions, clarifying the role of comparative advantage and the identification assumptions needed to estimate the yield function. It also describes the empirical model and its estimation. Section 5 discusses some descriptive baseline regressions, such as OLS, fixed effects, IV, various treatment effects and preliminary evidence that heterogeneity in returns is relevant. Section 6 presents the results from the correlated random coefficient model and the associated distribution of returns. Section 7 discusses implications of the results and a host of alternative models. Section 8 concludes.

2 Literature Review

This section briefly summarizes some of the empirical literature on technology adoption, focusing on a few studies in Sub-Saharan Africa (see Suri (2006) for a more detailed review).⁷ I describe the literature that has looked at the roles of heterogeneity, credit constraints, learning externalities and finally the more recent experimental research on similar technologies in Kenya. The seminal empirical paper on technology adoption is Griliches (1957) who looked at heterogeneity across local conditions in the adoption speeds of hybrid corn in Midwestern US and emphasized the role of expected profits and scale. He noted how the speed of adoption across geographical space depended on the suppliers of the technology and when the seed was adapted to local conditions. For Kenya, Gerhart (1975) tracked the adoption of hybrid maize in Western Kenya in the early 1970's. He highlighted the fast diffusion of hybrid and identified risk, education, credit availability, extension services and use of fertilizer as constraints.

There is a vast literature that describes the observable heterogeneity that drives adoption decisions. For example Schultz (1963) and Weir and Knight (2000) emphasized education. Various CIMMYT (The International Wheat and Maize Improvement Center) studies⁸ across Kenya highlighted the following: unavailability and untimely delivery of the technologies; labor and input use; costs; and unfavorable climactic conditions.⁹ There are a number of

⁷The literature on technology adoption is too vast to review here, excellent reviews are Sunding and Zilberman (2001), Sanders, Shapiro and Ramaswamy (1996), Rogers (1995), Feder, Just and Zilberman (1985), David (2003) and Hall (2004). For the theoretical literature, see Besley and Case (1993), Banerjee (1992) and Just and Zilberman (1983). For studies of land management practices, see Mugo et al (2000); for agricultural extension, see Evenson and Mwabu (1998); for property rights, see Place and Swallow (2000).

⁸See Doss (2003) and De Groote et al (2002) for a review of all the CIMMYT micro surveys in Kenya.

⁹For example, Makokha et al (2001) looked at fertilizer and manure use in Kiambu district. The main (self-reported) constraints to use were unavailability and untimely delivery, high labor costs and high prices of inputs. Ouma et al (2002) found that, in Embu district, gender, agroclimatic zone, manure use, hiring of labor and extension services were significant determinants of the adoption of improved seed and fertilizer. Wekesa et al (2003) looked at the adoption of various hybrids and fertilizer in the Coastal Lowlands where the non-availability and high cost of seed, unfavorable climatic conditions, perceptions of sufficient soil fertility, and lack of money were reasons for low use.

papers focusing on credit constraints, mostly using self-reports (see Croppenstedt, Demeke and Meschi (2003) for Ethiopia, Salasya et al (1998) for Western Kenya). Dercon and Christensen (2005) showed that the possibility of a poor harvest (and hence very low consumption) can account for the low use of fertilizer in Ethiopia. Their data showed similar adoption patterns to Kenya, with a lot of switching of technology use from period to period.

Much of the academic literature has focused on the learning externality described by Besley and Case (1993), which I find little evidence for. These papers mostly studied the Green Revolution in India where the learning externality was certainly key in the rapid growth of new hybrid varieties developed for India. Foster and Rosenzweig (1995) considered the adoption of high yielding varieties in India and found that farmers with more experienced neighbors are more profitable than those without. Munshi (2003) found these impacts to be heterogeneous across crops. Conley and Udry (2009) studied the adoption of fertilizer in the small-scale pineapple industry in Ghana and found evidence of social learning within information neighborhoods (defined by farmers as the people they discuss farming with).¹⁰

More recently, there has been a growing experimental literature. A number of field trials at Kenya Agricultural Research Institute (KARI) experiment stations, showing large increases in yields from hybrid seed and fertilizer. One of the early experimental studies was the Fertilizer Use Recommendations Project (FURP) in the early 1990's to understand optimal rates of fertilizer use (see Corbett (1998)). FURP recorded yields about half of those found at experiment stations (KARI (1993)). Hassan et al (1998) showed higher adoption and faster diffusion of hybrid in high potential areas, blaming poor extension, infrastructure and seed distribution in the marginal areas. Hassan, Murithi and Kamau (1998) found that farmers apply less fertilizer than optimal, leading to an estimated 30% yield gap.

De Groote et al (2003) considered an ex-ante impact assessment of the Insect Resistant Maize for Africa project that develops GM maize varieties that are more resistant to stem borers. They experimentally estimated a 13.5% crop loss due to these insects, valued at about \$80 million.¹¹ Duflo, Kremer and Robinson (2008a) ran experiments to understand the returns to fertilizer and the low use of fertilizer in Western Kenya. They found that the average rate of return for investing in top-dressing fertilizer is between 100% and 162% and that learning effects were extremely small. Their most significant contribution was to implement the Savings and Fertilizer Initiative (SAFI) as a commitment device for farmers. They showed that offering SAFI at harvest time (vs. planting) increases adoption by 17% and that this effect is about equivalent to a 50% subsidy to the fertilizer price. Duflo, Kremer and Robinson (2008b) looked at the heterogeneity in the returns to fertilizer. They found an

¹⁰Other studies of learning in Sub-Saharan Africa include Bandiera and Rasul (2003) for sunflower in Mozambique and Moser and Barrett (2003) for a rice production method in Madagascar.

¹¹See <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.

average annual return of about 70%, but zero returns to a combination of fertilizer and hybrid, and some evidence of heterogeneity, stating that “the returns to fertilizer are smaller when the control plot does better”. All of the empirical findings of Duflo, Kremer and Robinson (2008a, 2008b) are consistent with the evidence in this paper, though my explanation for the lack of adoption is different.

3 Institutional Context and Data

I now describe the relevant institutional detail and the data, which help motivate the model and empirical approach used. Maize¹² is the main staple in Kenya, with 90% of the population depending on it for income (Nyameino, Kagira and Njuki (2003)). Hybrid maize increases yields and can be more resistant to agricultural stress (Hassan et al (1998)). Both fertilizer and hybrid have been available since the 1960’s with more than twenty modern varieties of seed released since 1955, though later releases have not shown large yield improvements.¹³ In the data, about 70% of plots are planted with a hybrid variety released in 1986. Government recommendations for the types and quantities of hybrid seed and fertilizer vary across the country¹⁴, and both must be purchased each season. Hybrid seed replanted from the previous year’s harvest (i.e. recycled hybrid seed) has little yield advantage over non-hybrid. Most farmers (about 80% in the sample) use either both hybrid and fertilizer or neither. All inorganic fertilizer was imported and not locally produced over this period.

From 1965 to 1980, 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)), but this changed in the 1990’s.¹⁵ The most important government policy for this sector has been pan-territorial seed pricing. The hybrid seed was all produced by KARI on research stations and then distributed by Kenya Seed Company. The price of seed was not driven by market forces, shows little variance over time and was fixed across the country. In the data, only 1% of seed purchases are not from Kenya Seed Company. This pan-territorial seed

¹²McCann (2005) describes the fascinating history of maize in Africa. Smale and Jayne (2003) provide an excellent review of maize policy in Kenya.

¹³Karanja (1996) states “newly released varieties in 1989 had smaller yield advantages over their predecessors than the previously released ones... research yields were exhibiting a ‘plateau effect’”. Examples he gives are KSII, which was followed in time by H611 (with a 40% yield advantage), then H622 (16%) and then H611C (12%). H626 which had a 1% yield advantage over H625 was released eight years later.

¹⁴See Ouma et al (2002), Hassan (1998), Salasya et al (1998), Wekesa et al (2003) and Karanja (1996).

¹⁵Reform of the cereal sector began in 1988, followed by some liberalization in 1994. Smale and Jayne (2003) and Karanja (1996) attribute early successes to good germplasm, effective research, good distribution, coordinated marketing of inputs and outputs. This changed in the 1990’s as earlier policies of large subsidies, price supports, pan-territorial seed/output pricing, restrictions on cross district trade resulted in large fiscal deficits. The National Cereals and Produce Board accrued losses of 5% of GDP in the 1980’s. For more on the reforms, see Jayne et al (2001), Karanja, Jayne and Strasberg (1998), Jayne, Myers and Nyoro (2005), Nyoro, Kiiru and Jayne (1999) and Wanzala et al (2001).

pricing created poor incentives for suppliers to locate in distant areas or far from markets.¹⁶

3.1 Data

The dataset comes from the Tegemeo Agricultural Monitoring and Policy Analysis Project between Tegemeo Institute at Egerton University, Kenya and Michigan State University. It is a household level panel survey of Kenya, representative of rural maize-growing areas. Figure 1 shows a map of the survey villages and the population density across the country (darker shades represent greater density). Though data are available for the years 1997, 1998, 2000, 2002 and 2004, for most of the analysis in this paper, I use the 1997 and 2004 data. The 1997 and 2004 surveys collect detailed agricultural input and output data, consumption, income, demographics, infrastructure, and credit information.¹⁷ The panel sample covers just over 1200 households.

Figures 2 and 3 show the heterogeneity in hybrid and fertilizer use across provinces and over time. Figure 2 shows the stability in aggregate hybrid maize adoption over time and the persistence of cross-sectional differences.¹⁸ In principle, hybrid use could be a continuous variable as farmers plant quantities of hybrid, but only 2% farmers in the sample plant both hybrid and non-hybrid in a given season. I therefore take hybrid use to be binary throughout. Figure 3a shows the trends across provinces in the fraction of households that use inorganic fertilizer on maize and Figure 3b shows the total value (in constant Kenyan shillings) of inorganic fertilizer used, both showing similar persistent cross-sectional differences.¹⁹ Figures 4a and 4b show the distribution of yields for 1997 and 2004 by technology, illustrating that mean yields are much higher and the variance of yields lower in the hybrid sector, though both these summary measures could be artifacts of selection (as I discuss in Section 4).

Table 2a shows summary statistics for my sample of households for 1997 and 2004. Of the 26 different types of fertilizer used, Table 2a shows the three most popular (diammonium phosphate (DAP), calcium ammonium nitrate (CAN), and mono-ammonium phosphate (MAP)). Table 2b breaks out some of these variables by hybrid/non-hybrid use for 1997 and 2004. Yields are significantly lower across the board in the non-hybrid sector. Maize acreage and total seed planted are not different across sectors in 1997 and only just in 2004. Fertilizer, land preparation costs and main season rainfall are different across the two sectors. Finally, hired labor is only barely different in 2004, family labor only barely in 1997.

¹⁶The seed market was liberalized post 2004 and in late 2008, fertilizer was subsidized.

¹⁷The 2000 and 1998 surveys are similar, but data on family labor was not collected in 2000, and 1998 covers only a sub-sample of 612 households. The 2002 survey was a short proxy survey.

¹⁸The Coast province looks rather different in 2004. All the results in the paper are similar if the households in the Coast province are dropped from the sample (59 households). In addition, looking across wealth/asset or acreage quintiles, the pattern is identical with no systematic temporal variation in average adoption.

¹⁹Aggregate yields are not stable over time. There is a sharp drop in yields around 1997/1998, the result of El Nino. However, there are no dynamics in any of the main inputs. For example, the acreage farmers plant to maize is constant over this period. For more on these trends, see Suri (2006).

The land preparation costs and labor variables merit some discussion. The land preparation costs variable reported in the summary statistics covers costs that are incurred either before or right at the time of the seed choice and that do not include labor costs. The survey collects the labor data for each component of the production process, separately for hired and family labor. On average, about 75% of the total labor used for maize by these households is family labor, which is difficult to value (it should be valued at the shadow wage, which is not observable).²⁰ The main production activities that use labor are land preparation (about 23% of total labor), planting (10%), weeding (36%) and harvesting (30%). Harvesting includes all the post-harvest activities (threshing, winnowing, bagging, storage), which put the maize into a form fit for sale or consumption. The survey records the harvest of maize in bags, i.e. after all post-harvest activities.

Table 2c summarizes the labor variables by sector. Planting labor tends to be significantly different across the sectors since it is positively correlated with fertilizer use. Harvest and post-harvest labor are also higher in the hybrid sector, unsurprising as the yields of hybrid are higher. In terms of weeding, hired labor is only different across the two sectors in 2004, and family weeding only in 1997. Total weeding labor is not different across the sectors in either year. If I value family labor at the district hired labor wage, the dominant cost differences between hybrid and non-hybrid production come from seed and fertilizer costs, and not from labor cost differences. Finally, Table 2d shows the transitions of hybrid use in the data, with 30% of households switching in and out of use over 1997, 2000 and 2004. Including 2002, this fraction becomes 39%. In the raw data, households that always plant hybrid have the highest average yields, followed by households that switch in and out of use. Those that never use hybrid have the lowest average yields.

4 Theory and Empirical Framework

This section first describes the theoretical foundations underlying my empirical model and then the empirical framework. I am interested in estimating a distribution of returns to the hybrid technology to assess if the farmers that do not adopt hybrid are those that do not benefit from it. I do this by estimating the yield (production) function underlying farmers' adoption decisions. The exact structure of the adoption decisions is never estimated, so the adoption model below is for illustrative purposes. I base the model on a comparison of profits under hybrid and non-hybrid and it clearly illustrates the role of the fixed costs of acquiring hybrid seed and fertilizer in the adoption decision, which is important for identification. The model shows how the unobserved heterogeneity in the yield function is a key determinant in

²⁰When I impute the shadow wage for family labor as the district level hired labor wage, I find that about half my sample has negative profits, implying this is not the correct shadow wage, and that these households tend to undervalue family labor.

the profit comparison underlying the hybrid decision. The aim of this section is, therefore, to derive the fundamental of the yield function that is of interest: the distribution of farmers' comparative advantage in the production of hybrid maize, which is the correlated random coefficient in the yield function. I then discuss the underlying identification assumptions, their validity and the estimation of this yield function.²¹

4.1 Adoption Decisions Under Profit Maximization

I start with a simplified technology choice model where the farmer decides between hybrid and non-hybrid seed. In reality, the technology choice is a joint decision of using both hybrid and fertilizer or neither, which I elaborate on later. For ease of exposition, I use the hybrid/non-hybrid choice as a parsimonious representation of this joint choice.

The timing of the farmer's decisions is as follows. Each year, the farmer decides on his seed technology at the beginning of the growing season, just before the rains begin. The farmer makes his decision based on all his current information, his forward looking expectations as to the coming year's growing conditions, and the relative costs and benefits (differing productivities) of the two types of seed, which are assumed to be known to him.

The farmer is assumed to be risk-neutral and chooses a seed type to maximize profits per area of land. He compares $\pi_{it}^{*H}(p_{it}, a_{it}, b_{it}, \mathbf{w}_{it})$ and $\pi_{it}^{*N}(p_{it}, c_{it}, \mathbf{w}_{it})$, the maximized profit functions under hybrid and non-hybrid, respectively, where p_{it} is the expected output price of maize for both hybrid and non-hybrid maize²², a_{it} represents the (fixed) cost of obtaining hybrid seed (due to availability differences), b_{it} is per-unit cost of hybrid seed, c_{it} is the (very low, if not zero) per unit costs of replanting non-hybrid seed from the previous year's harvest and $\mathbf{w}_{it} \equiv (w_{1it}, w_{2it}, \dots, w_{Jit})$ represents the vector of input prices for the inputs X_{jit}^k , $j = 1, \dots, J$ and $k \in H, N$. The profit functions are:

$$\pi_{it}^H = p_{it}Y_{it}^H - (b_{it}s_{it} + a_{it}) - \sum_{j=1}^J w_{jit}X_{jit}^H \quad (1)$$

$$\pi_{it}^N = p_{it}Y_{it}^N - c_{it}s_{it} - \sum_{j=1}^J w_{jit}X_{jit}^N \quad (2)$$

where Y_{it}^H and Y_{it}^N are the yields of hybrid and non-hybrid maize respectively and s_{it} is the quantity of seed used. In addition, I have imposed the following restrictions based on the

²¹Note that profit functions themselves are difficult to estimate here due to the widespread use of family labor, which is hard to value. Households do not value family labor in the same way as hired labor. Below, I show that the yield function here is similar to a gross revenue function (as in Levinsohn and Petrin (2003)). As a robustness check, I estimate a model using the value of yields and valuing all the inputs where possible using their prices. The results are extremely similar. Finally, if I assume that family labor is valued at hired labor wages, I can estimate profit functions, which also show qualitatively similar results.

²²There is no distinction in the output market between hybrid and non-hybrid maize.

institutional context in Section 3. First, the quantity of seed used for a given area of land is the same whether it is hybrid or non-hybrid seed.²³ Second, since the seed price is fixed across space, b_t is indexed only by t and not i .

Therefore, the farmer chooses to plant hybrid when $\pi_{it}^{*H} > \pi_{it}^{*N}$ or when:

$$\left(Y_{it}^{*H} - \sum_{j=1}^J \frac{w_{jit}}{p_{it}} X_{jit}^{*H} \right) - \left(Y_{it}^{*N} - \sum_{j=1}^J \frac{w_{jit}}{p_{it}} X_{jit}^{*N} \right) > \frac{a_{it}}{p_{it}} + \frac{(b_t - c_{it})}{p_{it}} s_{it}^* \quad (3)$$

I can rewrite the second term on the right-hand side of the inequality in (3) as the normalized (by the price of maize) cost difference between hybrid and non-hybrid seed. In fact, it is the case that $s_{it}^* \approx s^*$, as evidenced in Tables 2a and 2b where both the quantity of seed used for a given area of land and the land area planted to maize are similar over time. This comes from standard seeding rates for maize and since land markets barely exist, there is little change in land ownership and land cropped to maize over time (see Suri (2006)).²⁴

From (3), the optimized profits from using hybrid over non-hybrid are greater when:

$$\left(Y_{it}^{*H} - \sum_{j=1}^J \frac{w_{jit}}{p_{it}} X_{jit}^{*H} \right) - \left(Y_{it}^{*N} - \sum_{j=1}^J \frac{w_{jit}}{p_{it}} X_{jit}^{*N} \right) > \frac{a_{it}}{p_{it}} + \frac{\delta_{it}^s}{p_{it}} \equiv A_{it} + \Delta_{it}^s \quad (4)$$

where $\delta_{it}^s = (b_t - c_{it})s^*$, $A_{it} = \frac{a_{it}}{p_{it}}$ and $\Delta_{it}^s = \frac{\delta_{it}^s}{p_{it}}$. In reality, the value of c_{it} , the cost of non-hybrid, is close to zero (relative to the hybrid cost) and the output price p_{it} does not vary much across space (time by province dummies explain 60% of the variation in p_{it}).

From the evidence on input use in Table 2b, the optimized quantities of inputs apart from fertilizer tend to be about equal under hybrid and non-hybrid. If, in addition, I assume that fertilizer is only used with hybrid (as in the data) and that a fixed amount is used per land allocated to hybrid in each year, then fertilizer costs are subsumed in b_t in equation (3) and hence in Δ_{it}^s .²⁵ If this is the case, then the farmer chooses hybrid if

$$\left(Y_{it}^{*H} - Y_{it}^{*N} \right) > A_{it} + \Delta_{it}^s \quad (5)$$

²³There are standard seeding rates for maize that do not vary by seed type. This is borne out by the empirical seeding rates not varying much over time and not varying across hybrid and non-hybrid sectors.

²⁴Note that the second term on the RHS of equation (3) can be ignored if the cost of obtaining hybrid seed, a_{it} , is much greater than the straight cost difference between hybrid and non-hybrid seed, which seems to be borne out in reality.

²⁵This is purely for expositional purposes, since the decision rule it leads to in equation (25) can be amended to include other input costs. Even if I allow the optimized inputs to differ by hybrid and non-hybrid for a given farmer (such as for harvest labor), this will create no bias in the empirical work since the estimated yield function does not depend on the *counterfactual* input use and conditions on inputs for the *observed* hybrid/non-hybrid choice. The other reason for taking the approach in the text is that for all inputs other than harvest/post-harvest labor, the assumptions are empirically true, and valuing labor is problematic in this context. Therefore, departures from the strict assumptions in the text should be thought of as captured in the error term ϑ_{it} in the decision rule in equation (25).

It is clear that adoption decisions based on profits depend fundamentally on yield comparisons. The above assumptions simply isolate the components of the adoption decision that are important to the identification and estimation of the yield function, which I now describe.

4.2 The Underlying Yield Functions

Underlying the profit functions, assume Cobb-Douglas production functions of the form:

$$Y_{it}^H = e^{\beta_t^H} \left(\prod_{j=1}^k X_{ijt}^{\gamma_j^H} \right) e^{u_{it}^H} \quad (6)$$

$$Y_{it}^N = e^{\beta_t^N} \left(\prod_{j=1}^k X_{ijt}^{\gamma_j^N} \right) e^{u_{it}^N} \quad (7)$$

These production functions for hybrid and non-hybrid maize have different parameters on the inputs, γ^H and γ^N , to allow for differential complementarity between the seed variety and the inputs (though the same set of potential inputs are used). u_{it}^H and u_{it}^N are sector-specific errors that may be the composite of time-invariant farm characteristics and time-varying shocks to production and the β 's are sector specific aggregate returns to yields. Taking logs,

$$y_{it}^H = \beta_t^H + x'_{it} \gamma^H + u_{it}^H \quad (8)$$

$$y_{it}^N = \beta_t^N + x'_{it} \gamma^N + u_{it}^N \quad (9)$$

I now place additional structure on the unobserved productivities and impose the following factor structure, as in Lemieux (1993, 1998) and Carneiro, Hansen, and Heckman (2003):

$$u_{it}^H = \theta_i^H + \xi_{it}^H \quad (10)$$

$$u_{it}^N = \theta_i^N + \xi_{it}^N \quad (11)$$

Farmers are assumed to know θ_i^H and θ_i^N , but not ξ_{it}^H and ξ_{it}^N , when making their seed choice. The transitory errors, ξ_{it}^H and ξ_{it}^N , are assumed to be uncorrelated with each other and with the X_{it} 's, and do not affect the hybrid decision (as per Heckman and Honore (1990)). It is clear that, in expected terms, $E(u_{it}^H - u_{it}^N) = (\theta_i^H - \theta_i^N)$ plays a role in the hybrid choice. I discuss this further in Section 4.4.

Following Lemieux (1998), since the relative magnitudes of the unobserved θ_i^H and θ_i^N are not identified, I define the relative productivity of a farmer in hybrid over non-hybrid as

$(\theta_i^H - \theta_i^N)$, using the following decomposition of θ_i^H and θ_i^N :

$$\theta_i^H = b_H(\theta_i^H - \theta_i^N) + \tau_i \quad (12)$$

$$\theta_i^N = b_N(\theta_i^H - \theta_i^N) + \tau_i \quad (13)$$

where 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 $\sigma_{HN} \equiv \text{cov}(\theta_i^H, \theta_i^N)$, $\sigma_H^2 \equiv \text{Var}(\theta_i^H)$, and $\sigma_N^2 \equiv \text{Var}(\theta_i^N)$.²⁶

The τ_i is farmer i 's absolute advantage: its effect on yields does not vary by technology, and it is (by construction of the linear projection) orthogonal to $(\theta_i^H - \theta_i^N)$. The gain, $(\theta_i^H - \theta_i^N)$, can be re-defined to be farmer-specific comparative advantage, θ_i , as:

$$\theta_i \equiv b_N(\theta_i^H - \theta_i^N) \quad (14)$$

Defining $\phi \equiv \frac{b_H}{b_N} - 1$, equations (12) and (13) become

$$\theta_i^H = (\phi + 1)\theta_i + \tau_i \quad (15)$$

$$\theta_i^N = \theta_i + \tau_i \quad (16)$$

I am interested in the structural parameter ϕ and the distribution of θ_i , fundamentals of the production function. The θ_i 's are the key unobservables that determine selection into hybrid.

Using this decomposition of u_{it}^H and u_{it}^N , I can rewrite equations (8) and (9) as²⁷

$$y_{it}^H = \beta_t^H + \tau_i + (\phi + 1)\theta_i + X_{it}\gamma^H + \xi_{it}^H \quad (17)$$

$$y_{it}^N = \beta_t^N + \tau_i + \theta_i + X_{it}\gamma^N + \xi_{it}^N \quad (18)$$

I use a generalized yield equation of the form,

$$y_{it} = h_{it}y_{it}^H + (1 - h_{it})y_{it}^N \quad (19)$$

and substituting in equations (17) and (18), I get

$$y_{it} = \beta_t^N + \theta_i + (\beta_t^H - \beta_t^N)h_{it} + X_{it}'\gamma^N + \phi\theta_i h_{it} + X_{it}'(\gamma^H - \gamma^N)h_{it} + \tau_i + \varepsilon_{it} \quad (20)$$

Equation (20) is my basic empirical specification - Section 4.5 describes the estimation in

²⁶The τ_i 's in equations (12) and (13) are the same. To see this, subtracting equation (13) from equation (12) gives $\theta_i^H - \theta_i^N = (b_H - b_N)(\theta_i^H - \theta_i^N)$. For the τ_i 's to be equal across sectors, $b_H - b_N$ must be equal to 1, which is easily shown: $b_H - b_N = \frac{\sigma_H^2 - \sigma_{HN} - \sigma_{HN} + \sigma_N^2}{\sigma_H^2 + \sigma_N^2 - 2\sigma_{HN}} = 1$.

²⁷Note that this decomposition of u_{it}^H and u_{it}^N implies a more complex Cobb-Douglas production function for hybrid maize, $Y_{it}^H = e^{\tau_i} e^{\theta_i^{\phi+1}} e^{\beta_t^H} \left(\prod_{j=1}^k X_{ijt}^{\gamma_j^H} \right) e^{\xi_{it}^H}$, and similarly for non-hybrid maize.

detail. Since, the coefficient on h_{it} , $\phi\theta_i$ (the fifth term in the equation), depends on the unobserved θ_i , this is a correlated random coefficient (CRC) model where the θ_i 's are correlated with the adoption decision.²⁸ I estimate two components of this model: the coefficient ϕ that describes how important differences in comparative advantage are and, second, the distribution of θ_i and hence the corresponding heterogeneous returns to hybrid.

Intuitively, in a specification like equation (20), θ_i measures farmer i 's relative productivity in hybrid over non-hybrid, i.e. his comparative advantage in hybrid. The θ_i 's are a fundamental of the production function, but also play a role in the adoption decision. If the farmers with high θ_i 's have *lower* gains to switching to hybrid from non-hybrid, then $\phi < 0$. In that case, relative to the mean θ_i in the population, a farmer would need a small θ_i in order to meet the adoption criterion. This is the notion of selection on the basis of comparative advantage, in that those with lower baseline productivities have larger gains to switching to the new technology. The coefficient ϕ therefore describes the sorting in the economy. If $\phi < 0$, there is less inequality in yields in this economy as compared to an economy where individuals are randomly allocated to a technology. On the other hand, if $\phi > 0$, then the self-selection process leads to greater inequality in yields. To see this, from equation (20), if we let $\mu_i = \theta_i + \tau_i$, then μ_i is a household specific intercept (average yield) and $\phi\theta_i$ is the household specific return to hybrid. Since θ_i and τ_i are uncorrelated by construction, the covariance between μ_i and $\phi\theta_i$ is $\phi\sigma_\theta^2$. The sign of ϕ is therefore the sign of the covariance between a household's overall average yield and its return to hybrid. A negative ϕ implies that farmers that do better on average do worse at hybrid and vice versa for a positive ϕ .

4.3 The Role of Fixed Costs in the Adoption Decision

In this section, I discuss the implications of profit maximization to illustrate how changes in infrastructure and access to seed and fertilizer distributors affect adoption decisions. This profit maximization approach is a generalization of the strict Roy model where adoption decisions are based purely on the outcome (yields). I combine the yield function decomposition of Section 4.2 with the profit maximization problem of 4.1 to discuss how the farm specific comparative advantage parameters combine with the costs of acquiring seed and fertilizer to determine the hybrid adoption decision.

Rewriting (4) in log-output and using (8) and (9), a farmer chooses hybrid if

$$E(u_{it}^H - u_{it}^N) > A_{it} + \Delta_{it}^s - (\beta_t^H - \beta_t^N) + \sum_{j=1}^J \left(\gamma_j^N x_{jit}^{*N} - \gamma_j^H x_{jit}^{*H} \right) \quad (21)$$

In the data, the revenue in hybrid is about double that in non-hybrid. 30% of this is due

²⁸Also note that equation (20) is a generalization of the household fixed effects model (see Suri (2006)).

to differential seed and fertilizer costs, 4% is land preparation cost differences, 7% is hired labor cost differences, family labor goes the other way and the rest is profit differences and costs of acquiring seed over and above the seed price. Apart from fertilizer, the optimized inputs x_{jit}^{*H} and x_{jit}^{*N} are not substantially different, so if $\gamma_j^H \simeq \gamma_j^N$ for all inputs j , then the last term in this expression is zero for all inputs except fertilizer.²⁹ If for fertilizer, I assume as above that fertilizer is used only with hybrid and in fixed proportions per area of land, then fertilizer is subsumed in the Δ_{it}^s . In this case, the adoption rule reduces to:

$$E(u_{it}^H - u_{it}^N) > A_{it} + \Delta_{it}^s - (\beta_t^H - \beta_t^N) \quad (22)$$

Substituting in equations (10) and (11), and using equations (15) and (16),

$$(\theta_i^H - \theta_i^N) > A_{it} + \Delta_{it}^s - (\beta_t^H - \beta_t^N) \quad (23)$$

$$\phi\theta_i > A_{it} + \Delta_{it}^s - (\beta_t^H - \beta_t^N) \quad (24)$$

Equation (24) implies that the technology choice will depend on (i) unobserved, farmer-specific, time-invariant comparative advantage θ_i that comes from the underlying production function, (ii) pure macroeconomic factors affecting the differential productivity of hybrid and non-hybrid seed ($\beta_t^H - \beta_t^N$), (iii) potentially time varying costs of obtaining hybrid, A_{it} , and (iv) the real relative purchase costs of hybrid seed, Δ_{it}^s . The key aspect of the A_{it} costs is that they affect the demand for hybrid seed, but not production and yields directly.³⁰

This framework illustrates how to empirically relate my comparative advantage estimates to observables. Let $\alpha_i \equiv E_t[A_{it}]$ be the population time mean of the real fixed costs of acquiring hybrid seed for each farmer, and let $\vartheta_{it} \equiv A_{it} - \alpha_i$. Re-writing equation (24),

$$\phi\theta_i - \alpha_i > \Delta_{it}^s - (\beta_t^N - \beta_t^H) + \vartheta_{it} \quad (25)$$

where α_i is the permanent component of the fixed costs (i.e. the average real fixed costs for a farmer over time) and ϑ_{it} the period to period fluctuations in these costs.

Clearly, $\phi\theta_i$ cannot be separately distinguished from the permanent component of fixed costs, α_i . In the empirics, I relate the estimated comparative advantage to permanent aspects of infrastructure, since both act in an equivalent way to drive adoption decisions *across* farmers. By contrast, the changes in the fixed costs, ϑ_{it} , drive adoption decisions for a *given*

²⁹It is important to emphasize again that the assumption that the term involving the optimized inputs is zero is purely for expositional convenience, in addition to it being roughly true in the data. The estimated yield equation will condition on the observed input intensities. So, even without this simplifying assumption, conditional on the observed input intensities, the counterfactual input intensities do not affect the observed outputs, so their role in the adoption decisions does not create bias in the empirical work.

³⁰ Δ_{it}^s , the relative per unit costs of hybrid seed normalized by the output price also play a role, as equation (24) makes clear. But, as this varies mostly across time and not by individual, it is absorbed by the time dummies I include in the empirical specifications.

farmer over time. Section 4.4.3 discusses the empirical evidence on these costs.

4.4 Identification of The Generalized Yield Function

The basic equation I estimate is equation (20):

$$y_{it} = \beta_t^N + \theta_i + (\beta_t^H - \beta_t^N)h_{it} + X'_{it}\gamma^N + \phi\theta_i h_{it} + X'_{it}(\gamma^H - \gamma^N)h_{it} + \tau_i + \varepsilon_{it}$$

In this section, I discuss the identification assumptions needed to estimate this and the justifications for them. While the projections in equations (15) and (16) only impose uncorrelatedness of the absolute advantage, τ_i , and the comparative advantage, θ_i , my empirical work does not require this level of generality. So, as in Lemieux (1993), I use the stronger sufficient assumption of mean independence of the composite error ($\tau_i + \varepsilon_{it}$) and the comparative advantage component (θ_i) and the histories of the regressors, i.e. that

$$E(\tau_i + \varepsilon_{it} | \theta_i, h_{i1} \dots h_{iT}, X_{i1}, \dots X_{iT}) = 0 \quad (26)$$

With respect to the absolute advantage component, τ_i , the assumption in equation (26) is not restrictive for two reasons. First, by definition, τ_i does not affect the differential return to growing hybrid (it is differenced out of $\theta_i^H - \theta_i^N$ as per equations (15) and (16) above). Second, Heckman and Honore (1990) discuss the identification of the Roy model and show that if the absolute advantage plays a role in adoption decisions, then the Roy model has no empirical content. This assumption is therefore simple the standard strict exogeneity requirement used in panel data models.³¹

The strict exogeneity assumption on the transitory part of the composite error, ε_{it} , is more restrictive, though since the data includes measures of shocks to yields, this assumption is more plausible than if I did not observe such shocks. In terms of the primitives of the model,

$$\varepsilon_{it} \equiv h_{it}\xi_{it}^H + (1 - h_{it})\xi_{it}^N \quad (27)$$

So, the mean independence assumption regarding ε_{it} implies that ξ_{it}^H and ξ_{it}^N from equations (17) and (18) do not affect the farmer's decision to use hybrid and, crucially, the farmer's switching behavior. I discuss the various types of possible shocks and how they relate to yields and hybrid decisions in the following two subsections. Given the long lag between planting

³¹The assumption in equation (26) implies that $E(\tau_i | \theta_i) = 0$ by the Law of Iterated Expectations. As shown in Section 4.5.1, θ_i can be written as the linear projection $\theta_i = \lambda_0 + \lambda_1 h_{i1} + \lambda_2 h_{i2} + \lambda_3 h_{i1} h_{i2} + v_i$. Given the mean independence of τ_i from the adoption history discussed above, the additional assumption that $E(\tau_i | \theta_i) = 0$ only implies $E(\tau_i | v_i) = 0$. This last assumption is not actually needed for the empirical work, as the relationship between these two error components is left unspecified - they can be correlated, for example, and my empirical approach is still consistent. So, the mean independence assumption in (26) is 'overly strong', but is useful for expositional convenience concerning the adoption behavior and the assumptions on the composite error term. I do make the strict exogeneity assumption regarding τ_i , as just described.

and harvest (an average of four or more months), I consider two ‘types’ of ξ_{it}^j ($j \in H, N$) transitory shocks to yields: those that occur after the seed choice has been made and those that occur before. I discuss each separately below and then the implications for the switching behavior.

4.4.1 Shocks Realized After the Technology Choice

Central to identification is the fact that the hybrid seed choice is made before the farmer experiences most of the agricultural shocks to yields. However, the shocks post planting may affect optimal input use differentially for hybrid and non-hybrid and hence pose a problem for identification. So, it is key that the most important transitory shock post planting, rainfall, is observable in the data and can be controlled for. There may be other unobservable shocks to yields that are correlated with inputs, such as weeds, pests, disease, etc. Hassan, Onyango and Rutto (1998) use survey data to rank farmers’ perceptions of the most important maize production shocks and what options are available to farmers to deal with these shocks. They find that inadequate and/or erratic rainfall is one of the most widespread and one of the two worst shocks to yields across all zones.³²

In addition, given the timing of production, the inputs that could be adjusted after the technology choice are only a sub-sample of the labor inputs since the empirical specification will allow for the joint hybrid-fertilizer decision. The main labor activities are land preparation, planting, weeding, harvest and post harvest. Land preparation and planting are costs incurred before the technology choice and harvest and post-harvest after all shocks are realized. The only input that could therefore be an issue is weeding labor. Table 2c shows that the use of weeding labor in the sample is not very different across hybrid and non-hybrid farmers. Duflo, Kremer and Robinson (2008b) compare plots that use hybrid and fertilizer, those that use just fertilizer and those that use non-hybrid and no fertilizer. Other than the cost of the seed and fertilizer, they state that farmers reported no differences in the time spent weeding across these plots and their field officers observed no differences in weeding.

Overall, it seems that shocks realized after the hybrid choice are well captured by the

³²I have also investigated other potential shocks. For example, Hassan, Onyango and Rutto (1998) also identify Striga (a weed) as an important issue in three agro-ecological zones (pests and most diseases do not rank high). Hassan and Ransom (1998) carefully analyze Striga and state there is “no conclusive evidence that local vs. improved maize, time of planting, and cropping pattern either encourage or discourage Striga”. The occurrence of Striga therefore seems to be uncorrelated with the hybrid decision and should not pose a problem for identification. As a robustness check, where I can match my households to these three zones, I drop these households and the results are extremely similar. Another example is temperature - maize grows best in a temperature range of 24-30 degrees C and has trouble germinating at temperatures above 38 (see Pingali and Pandey (1999) and McCann (2005)). All the maize areas in Kenya fit within this range of temperatures. Looking at data on monthly temperatures for 1956 to 2006 by latitude and longitude, the maximum observed temperature is 23 degrees. The warmest part of the country in my sample is probably the Coast Province. As a robustness check, I drop the households in the Coast Province from the analysis and it does not qualitatively change the results.

variables in my data since rainfall is observed and post planting input use is unlikely to be differential across technologies.

4.4.2 Shocks Realized Before the Technology Choice

Even though the identification assumptions seem robust to shocks realized after the technology choice, there could be an issue with shocks that happen between 1997 and 2004 and affect the use of hybrid as well as yields. An example may be changes in household structure, for example the death of adults in the household due to HIV, that affect the quality of labor as well as the decision to use hybrid. In all the empirical results reported, I control directly for household structure (excluding these controls does not affect the results, see Suri (2006)). In addition, all the results in the paper are reported for the full sample as well as for the sample that excludes two districts where much of the adult mortality was concentrated over this period and where HIV is prevalent. The results are stronger and more consistent in the latter case, implying that my description of adoption decisions is most relevant for the sample that excludes the two high mortality districts, what I refer to as the non-HIV sample.³³

4.4.3 The Switching Behavior in Practice

The identification assumptions just described translate directly into assumptions on what drives the hybrid switching behavior in the data. Essentially, the unobserved time-varying variables that drive the switching should not be correlated with yields. I argue that the switching behavior is driven by exogenous changes in the availability of seed and fertilizer. If hybrid seed and/or fertilizer are not easily available at the time of planting, this would manifest itself in a supply constraint or a higher cost of using hybrid, and the farmer would switch to non-hybrid.³⁴

There is a lot of qualitative evidence that this happens often. Farmers get their seed and fertilizer from very small local distributors, who get their supplies from slightly larger distributors, etc. In the survey, farmers who use non-hybrid were asked why they do not use improved seed varieties and 14% blamed unavailability of seed as one of top two reasons for this, 65% put high costs as one of the top two reasons. The 2000 survey (data not shown, see Suri (2006)) showed that one of the common fertilizers (MAP) was not used at all in the sample due to a macro level unavailability. Several of the CIMMYT and FURP studies mentioned earlier cited unavailability of inputs, issues that were also at the forefront of The

³³The survey does not necessarily identify the cause of death accurately so these are not necessarily all HIV related deaths. However, these two districts are well known to be the areas with extremely high HIV prevalence rates. So, it is likely that the high mortality rates in these two districts are a consequence.

³⁴The framework in Dercon and Christiaensen (2005), for example, shows the switching to be related to households' ability to smooth downside consumption risks. This would be compatible with my framework.

Africa Fertilizer Summit in 2006.³⁵ In addition, there are certainly changes in the pricing and availability of seed and fertilizer from period to period, especially since the Kenyan government plays such an active role in these markets.³⁶

Finally, Tables 2a and 2b show the changes in infrastructure between 1997 and 2004. These have been due to private investment in fertilizer distribution networks, with now more than 10 importers, 500 wholesalers and 7000 retailers in the country (see Ariga, Jayne and Nyoro (2006)). For example, the distance to the closest fertilizer seller fell from 6.3 km in 1997 to 3.5 km in 2004 (a fall from 4.9 km to 2.4 km for hybrid farmers and from 9.3 km to 5.1 km for the non-hybrid).

4.5 Estimating a Model with Heterogeneous Returns

I now describe how to estimate the CRC model in equation (20).³⁷ An alternative estimation framework would be to use contemporary IO methods. However, in the face of the correlated heterogeneity in equation (20), methods such as those of Olley and Pakes (1996) and Levinsohn and Petrin (2003) that focus on unobserved time-varying productivity are inconsistent since the returns to hybrid are heterogeneous and correlated with the decision to use hybrid. Also, in the contexts those papers examined, productivity shifts for existing firms and the productivities for the firms leaving and entering the industry, are of key importance in estimating responses to regulatory or trade environments. In my scenario, given the static but cross-sectionally heterogeneous nature of the returns to hybrid and the underlying productivities of Kenyan farmers, it is of less importance to deal with time-varying productivities, since the approach controls for the dominant time varying shocks to output effectively.

4.5.1 Empirical Identification of the CRC Model

The estimation strategy I use is a generalization of the Chamberlain (1982, 1984) correlated random effects approach. It parallels Chamberlain (1984) in how the model is identified and how the parameters of the model are estimated, most importantly ϕ . Later, I describe how to use these estimates to derive a distribution of the predicted θ_i 's. For simplicity, I first

³⁵For the Africa Fertilizer Summit 2006, see <http://www.africafertilizersummit.org/>.

³⁶The price of fertilizer varies tremendously over time since its supply and availability vary. It is crucial that the fertilizer be available right before planting. Since the early 2000's, the National Cereals and Produce Board (NCPB) has been importing fertilizer in bulk and sometimes selling it at lower prices to farmers. A Lexis Nexis search covered only one newspaper and returned 55 articles between 1999 and 2004 documenting the role of the government in fertilizer, seed and maize markets. The articles discuss changes in prices of the inputs in different parts of the country, often due to shortages (DAP in 1999 and 2004), delays in the imports by the NCPB (2004), a stop to MAP fertilizer aid from Japan (late 1990's on) and the dramatic changes in the structure of the institutions that help farmers get access to inputs.

³⁷This empirical model is similar to models of individual specific heterogeneity in Heckman and Vytlacil (1997), Card (2000, 1998), Deschênes (2001), Carneiro, Hansen and Heckman (2001), Carneiro and Heckman (2002), and Wooldridge (1997). Lemieux (1998) uses the same model to look at whether the return to union membership varies along observable and unobservable dimensions.

describe how to estimate the model without covariates:

$$y_{it} = \delta + \beta h_{it} + \theta_i + \phi \theta_i h_{it} + u_{it} \quad (28)$$

where $u_{it} \equiv \tau_i + \varepsilon_{it}$ and assuming $\beta_t^H - \beta_t^N \equiv \beta \forall t$. Relaxing this assumption empirically does not change the results.

To estimate equation (28), I eliminate the dependence of the observed θ_i 's on the endogenous input (h_{it}), by following Chamberlain to exploit the linear projection of θ_i on the full history of the inputs, though the projection I use is more general. The θ_i 's are projected onto not just the history of the hybrid decisions, but also their interactions so that, in the two period case, the projection error is orthogonal to h_{i1} and h_{i2} individually as well as to their product, $h_{i1}h_{i2}$ by construction.³⁸ The generalized linear projection used is:

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

In addition, the θ_i 's are normalized so that $\sum \theta_i = 0$.³⁹ This normalization implies that λ_0 can be written as a function of λ_1 , λ_2 and λ_3 from equation (29).

Here, the coefficient λ_3 is of crucial importance to the empirical identification and estimation of the model, and is where this is a generalization of Chamberlain's approach. The $h_{i1}h_{i2}$ interaction term is necessary in the projection to ensure that v_i is orthogonal to every possible history of hybrid use. Since all the h_{it} variables describing the use of hybrid are dummies, to estimate λ_3 it is necessary to have farmers in the sample that have planted hybrid in *both* periods. Also note that the projection does not have a behavioral interpretation, but is used purely to purge θ_i of its dependence on the full histories of the inputs.

Substituting the projection into the yield equation for each of the two time periods gives:

$$y_{i1} = (\delta + \lambda_0) + [\lambda_1(1 + \phi) + \beta + \phi\lambda_0]h_{i1} + \lambda_2 h_{i2} + [\lambda_3(1 + \phi) + \phi\lambda_2]h_{i1}h_{i2} + (v_i + \phi v_i h_{i1} + u_{i1}) \quad (30)$$

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

The corresponding reduced forms are:

$$y_{i1} = \delta_1 + \gamma_1 h_{i1} + \gamma_2 h_{i2} + \gamma_3 h_{i1} h_{i2} + \varsigma_{i1} \quad (32)$$

³⁸Chamberlain's CRE model uses the projection $\theta_i = \lambda_1 h_{i1} + \lambda_2 h_{i2} + v_i$, which, when substituted into the yield function, gives $y_{it} = \lambda_0 + \lambda_1 h_{i1} + \lambda_2 h_{i2} + \beta h_{it} + \phi \lambda_0 + \phi \lambda_1 h_{i1} h_{it} + \phi \lambda_2 h_{i2} h_{it} + v_i + \phi v_i h_{it} + u_{it}$. Even though v_i is uncorrelated with h_{i1} and h_{i2} individually (by the nature of the projection), it could be correlated with their product, $h_{i1}h_{i2}$, so that $E[v_i h_{i1} h_{i2}] \neq 0$. The CRC projection must therefore also include all the interactions of the hybrid histories.

³⁹This normalization is used so that β in equation (28) corresponds to the average return to hybrid as in a standard fixed effects model. This normalization fixes the average θ_i in the sample and does not change the ordering of the θ_i 's across the hybrid histories of farmers.

$$y_{i2} = \delta_2 + \gamma_4 h_{i1} + \gamma_5 h_{i2} + \gamma_6 h_{i1} h_{i2} + \varsigma_{i2} \quad (33)$$

Equations (32) and (33) give six reduced form coefficients $(\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6)$, from which the five structural parameters $(\lambda_1, \lambda_2, \lambda_3, \beta, \phi)$ can be estimated using minimum distance. The structural parameters are therefore overidentified given these minimum distance restrictions:

$$\begin{aligned} \gamma_1 &= (1 + \phi)\lambda_1 + \beta + \phi\lambda_0 \\ \gamma_2 &= \lambda_2 \\ \gamma_3 &= (1 + \phi)\lambda_3 + \phi\lambda_2 \\ \gamma_4 &= \lambda_1 \\ \gamma_5 &= (1 + \phi)\lambda_2 + \beta + \phi\lambda_0 \\ \gamma_6 &= (1 + \phi)\lambda_3 + \phi\lambda_1 \end{aligned} \quad (34)$$

There may seem to be six structural parameters, since λ_0 is in equations (29) and (34). However, given the normalization $\sum \theta_i = 0$, then $\lambda_0 = -\lambda_1 \overline{h_{i1}} - \lambda_2 \overline{h_{i2}} - \lambda_3 \overline{h_{i1} h_{i2}}$ where $\overline{h_{i1}}$ and $\overline{h_{i2}}$ are the averages of the adoption decisions in periods one and two, and $\overline{h_{i1} h_{i2}}$ is the average of their interaction.

From the restrictions above, for example, ϕ could be (inefficiently) estimated as a combination of the reduced form parameters in two ways, $\phi = \frac{\gamma_6 - \gamma_3}{\gamma_4 - \gamma_2}$ and $1 + \phi = \frac{\gamma_1 - \gamma_5}{\gamma_4 - \gamma_2}$. Note that if γ_2 and γ_4 are equal, then ϕ is not identified. Requiring γ_2 and γ_4 to be different means that the reduced form effect of the hybrid decision in period two on yields in period one has to be different from the reduced form effect of the hybrid decision in period one on yields in period two. Since these are reduced form coefficients, this could structurally arise from a number of mechanisms. One example directly from the theory above would be differential shocks to the costs in each period, i.e. different ϑ_{it} (see equation (25)) in each period. In Table 7, the null that $\gamma_2 = \gamma_4$ can clearly be rejected. In addition, in the extension discussed in Section 4.5.2 where the use of fertilizer is endogenized (the preferred model), the problem becomes heavily overidentified, and ϕ is a more complicated function of the underlying reduced form parameters.

4.5.2 Extensions to the Basic Model

I consider the following extensions to the basic model⁴⁰:

1. Covariates: all the identification arguments presented above generalize when covariates are included. Covariates can be either exogenous or endogenous. Exogenous covariates

⁴⁰ An extension to three periods is straightforward. The problem is heavily overidentified with 9 structural parameters to estimate from the 21 reduced form coefficients for the basic case. Models with three time periods are not presented here as the 2000 survey does not collect data on family labor inputs. See Suri (2006).

are uncorrelated with the θ_i 's and therefore enter only the reduced form equations (32) and (33). Endogenous covariates are correlated with the θ_i 's and hence also enter the projection in equation (29). In the case where fertilizer is the other endogenous covariate, the CRC projection generalizes to

$$\begin{aligned} \theta_i = & \lambda_0 + \lambda_1 h_{i1} + \lambda_2 h_{i2} + \lambda_3 h_{i1} h_{i2} + \lambda_4 h_{i1} f_{i1} + \lambda_5 h_{i2} f_{i1} + \lambda_6 h_{i1} h_{i2} f_{i1} \\ & + \lambda_7 h_{i1} f_{i2} + \lambda_8 h_{i2} f_{i2} + \lambda_9 h_{i1} h_{i2} f_{i2} + \lambda_{10} f_{i1} + \lambda_{11} f_{i2} + v_i \end{aligned} \quad (35)$$

where f_{it} for $t = 1, 2$ represents the use of fertilizer in each period. Adding more endogenous covariates does complicate the CRC model and it can become cumbersome.⁴¹ In this case, there are twenty two reduced form coefficients and only fourteen structural coefficients to be estimated, so the problem becomes heavily overidentified.

2. Joint choice variables: the two-sector model presented above (hybrid/non-hybrid) can be extended to multiple sectors. Since the use of hybrid seed and fertilizer are clearly a joint decision in the data, as a further check, the use of fertilizer is incorporated into the technology sector. I therefore also estimate a model that looks at the heterogeneity in returns in a technology sector, where a farmer is in the technology sector when he uses both hybrid and fertilizer, else he is not in the technology sector.

5 Descriptive Regressions

This section provides estimates of the returns to hybrid from models that assume homogeneity in these returns. I also present descriptive evidence that illustrates the importance of selection and heterogeneity in returns, and that there may be selection on these heterogeneous returns.

5.1 OLS and Fixed Effects Estimates

The OLS estimates of the yield functions, controlling for various inputs, are in the first columns of Table 3a. The estimates of the average return to hybrid are extremely large: households that plant hybrid have 54% to 100% higher yields. The fourth and fifth columns of Table 3a report the household fixed effects results. The coefficient on hybrid decreases to about 9%, indicating a substantial role for heterogeneity in the production function that is fixed across households.⁴² The simple household fixed effects estimates are consistent

⁴¹There is some justification for treating only fertilizer as endogenous with respect to θ_i . First, the summary statistics by sector in Table 2b show that there are not big differences across sectors in the use of other inputs. Second, using the estimates from the hybrid problem in Table 8a, I correlate the predicted θ_i 's with the inputs. Only the correlations between the θ_i 's and fertilizer are important in magnitude and/or significance.

⁴²The fixed effects framework imposes restrictive assumptions on the underlying adoption process. Apart from the fixed effect, the adoption decision cannot depend on observed outcomes except under restrictive

under the assumption of strict exogeneity. Chamberlain’s correlated random effects (CRE) approach illustrates how the fixed effects model is overidentified and testable with panel data.⁴³ Intuitively, the CRE model tests the fact that if the fixed effects model is valid, then the only way the history of h_{it} affects the current outcome is through a fixed effect that is the same in every period. Table 3b shows these tests. It shows both the reduced form and structural estimates for the CRE model. Covariates can be treated as either exogenous or endogenous (the latter are assumed to be correlated with the fixed effects). I report estimates where all the covariates are allowed to be endogenous.⁴⁴ The CRE estimates of the return to hybrid are close to the household fixed effects estimates in Table 3a, but the χ^2 values on the overidentification test allow me to reject the fixed effects model.⁴⁵ The last column of Table 3b shows estimates for the case where hybrid is interacted with all the covariates. Though the coefficient on hybrid is negative, when evaluated at the mean input levels, the mean return to hybrid is positive.

5.2 IV and Treatment Effect Estimates

This section presents IV and control function estimates of the returns to hybrid, using the Heckman two-step estimator (also see Garen (1984)). In particular, I present estimates of the average treatment effects (ATE), the treatment on the treated (TT), marginal treatment effects (MTE) and local average treatment effects (LATE) under non-random assignment (see Björklund and Moffitt (1987) and Heckman, Tobias and Vytlacil (2001)).⁴⁶ These approaches do not fully address the issues at hand and do not exploit the panel nature of the data, but they provide useful benchmarks.

With assumptions of normality, it is straightforward to estimate the selection corrected treatment effects using a two-step control function procedure. The first stage is a probit describing the hybrid adoption decision, from which selection correction terms are computed

assumptions on the transitory component of yields (see Ashenfelter and Card (1985)). Such assumptions can be motivated by myopia or ignorance of the potential gains from planting hybrid. Both of these seem unrealistic here since hybrid maize was introduced in the 1960’s, with widespread use of extension services to promote the technology (Evenson and Mwangi (1998)) and data in the survey instrument supports this as 90% of farmers have used hybrid seed at some point in the past.

⁴³With a DGP of $y_{it} = \delta + \beta h_{it} + \alpha_i + u_{it}$, the CRE model is based on the assumption of strict exogeneity, i.e. $E(u_{it}|h_{i1}, \dots, h_{iT}, \alpha_i) = 0$. For the CRE model, the minimum distance estimator is the minimum χ^2 estimator if the weight matrix used is the inverse of the variance covariance matrix of the reduced form coefficients. See Suri (2006) for more detail on this approach for the current setting.

⁴⁴If I treat only hybrid as endogenous and all the other covariates as exogenous, the results are similar.

⁴⁵This overidentification test is an omnibus test and has low power against any specific alternative. It is therefore not that surprising that I am able to reject the overidentifying restrictions.

⁴⁶The treatment effects are defined as follows. The ATE is given by $E((y^H - y^N)|X = x) = x(\gamma^H - \gamma^N)$, $TT(x, z, h(Z) = 1) = x(\gamma^H - \gamma^N) + (\rho_H \sigma_H - \rho_N \sigma_N) \frac{\phi(z\pi)}{\Phi(z\pi)}$ and $MTE(x, u_i^s) = x(\gamma^H - \gamma^N) + (\rho_H \sigma_H - \rho_N \sigma_N) u_i^s$ where $y^H - y^N$ is the yield gain from hybrid, u_i^H and u_i^N are as defined earlier, $\sigma_H^2 = Var(u_i^H)$, $\sigma_N^2 = Var(u_i^N)$, u_i^s is the error in the selection equation, $\rho_H = Corr(u_i^H, u_i^s)$ and $\rho_N = Corr(u_i^N, u_i^s)$. $\phi(\cdot)$ and $\Phi(\cdot)$ represent the normal probability and cumulative distribution functions respectively and $Var(u_i^s) = 1$.

and used as controls in second stage sector-specific yield functions. The ATE uses estimates from this second step. The TT adjusts the estimated ATE for the sample of those who actually plant hybrid. The slope of the MTE describes whether people who are more likely to use hybrid for unobservable reasons (i.e. have a higher error in the selection equation) have higher or lower returns from planting hybrid.

The exclusion restriction I use is the distance to the closest fertilizer store (not to where fertilizer is actually purchased), which proxies for the availability of the technologies. The treatment effects are shown in Table 4 separately for 1997 and 2004. The ATE's are all large and positive, ranging from 1.3 to 2.4, with smaller TT estimates. The MTE slope is consistently negative and significant: -0.99 in 2004 and -2.51 in 1997. A non-zero MTE slope implies heterogeneity in returns and the sign of the MTE slope provides information on the underlying selection process. The negative MTE slopes imply that the farmers who are more likely to use hybrid are those with the lower relative returns to using hybrid, i.e. there is negative selection on returns in hybrid.

I also estimate the IV (LATE) specification. An issue with IV estimates is that the results are often different if different instruments are used. This is attributed to underlying heterogeneity in the population where the separate instruments affect a different subset of the population, resulting in the different LATE estimates (i.e. an average of heterogeneous returns that is instrument dependent). However, the instrument itself does not always highlight which subset of the population it affects (Heckman (1997)). The lower panel of Table 4 reports two versions of IV. The first column reports an estimate of over 200%, which uses the distance to the closest stockist of fertilizer as the excluded instrument. The second column uses the interactions of this distance measure with dummies for the household's asset quintile as excluded instruments (the distance and asset quintile main effects are included in both stages). This strategy allows distance to be more of a constraint for poorer households. These estimates are on the order of 150%, still very large when compared to the earlier OLS and household fixed effects estimates, and indicate high returns to hybrid for the supply (of seed and fertilizer) constrained farmers.

Overall, the preliminary evidence from these approaches indicates large positive returns for at least some sub-populations of farmers, but also evidence of substantial heterogeneity in returns with negative selection into the use of hybrid.

5.3 Motivation for Heterogeneity in Returns

This section presents some preliminary regressions that further motivate heterogeneity in returns. Previous research has developed tests for heterogeneity in returns for experimental data (see Heckman, Smith and Clements (1997)). Since the data is not experimental, these

results are not reported here, but using such tests I do reject the null of no heterogeneity.⁴⁷

Table 5 presents some evidence of selection on the part of farmers in their adoption decisions. The adoption history of a farmer is split into a set of dummy variables to analyze whether farmers with different histories have different returns (see Jakubson (1991) and Card and Sullivan (1988)). A “joiner” is a farmer who does not plant hybrid the first period, but does the next, and a “leaver” is a farmer who plants hybrid the first period, but not the next. Similarly, “hybrid stayers” always use hybrid and “non-hybrid stayers” always use non-hybrid. Table 5 compares the yields for each of these groups in 1997 and 2004 separately. If there was no selection at all, we would expect the leavers in 1997 to be no different from the hybrid stayers in 1997, and no different from the joiners in 2004. Similarly, the joiners in 1997 and the leavers in 2004 should be no different from the omitted group (the non-hybrid stayers). I can reject most of these restrictions, implying that there is selection and leaver and joiners are distinct (unlike in a fixed effects model).

Finally, in Table 6, I look for heterogeneity in the returns to hybrid along observables, estimating OLS and fixed effects yield functions separately for hybrid and non-hybrid farmers. In the OLS case, the returns to some of the covariates are different, though not family or hired labor. In the case of the fixed effects specifications, it is only the return to fertilizer that is different across hybrid and non-hybrid sectors. The last row reports estimates of the return to hybrid (evaluated at the mean inputs), still showing a significant return to hybrid.

6 CRC Estimates

This section describes the results for the CRC model. I report estimates for the pure hybrid model described in detail above (with and without covariates). In addition, I report results for the endogenous covariates and joint hybrid-fertilizer technology sector models.

Tables 7, 8a, 8b and 8c present the CRC model reduced form and structural estimates. These tables report only the optimal minimum distance (OMD) estimates where the weight matrix used in the minimum distance problem is the inverse of the variance covariance matrix of the reduced form coefficients. If the minimum distance problem is overidentified, the χ^2 test statistic on the overidentification test is the value of the minimand in the OMD problem.⁴⁸

Throughout, the results show that the selection into hybrid is negative with the farmers having the lowest yields in non-hybrid having the highest returns to planting hybrid. The

⁴⁷These tests include looking at the Frechet-Hoeffding bounds, bounding the variance in the percentiles of the returns distribution and testing whether this bound is different from zero (see Suri (2006)).

⁴⁸Equally weighted minimum distance (EWMD) estimates use the identity matrix as the weight matrix and diagonally weighted minimum distance (DWMD) estimates use the OMD matrix with the off-diagonal elements set to zero (see Pischke (1995)). The OMD estimates are efficient, but may be biased in small samples and can therefore be out-performed by EWMD (see Altonji and Segal (1996)). For all the results in the paper, the EWMD and DWMD estimates are extremely similar to OMD, but the OMD estimates are asymptotically efficient, so only those are reported. See Suri (2006) for the EWMD and DWMD estimates.

estimated ϕ is consistently negative, which illustrates that the households that do better on average do relatively worse at hybrid as the sign of ϕ describes the sorting process. The CRC model results are therefore consistent with the earlier MTE results.

Table 7 presents the two period reduced forms for the CRC model. The first column shows the reduced form estimates without covariates as a benchmark. The second column controls for all the inputs and the third for all the inputs interacted with hybrid. The reduced forms are estimated via seemingly unrelated regressions with the most general variance-covariance estimation. The estimates in Table 7 are for the basic specification described above where hybrid is the only endogenous variable and the projection used is given by equation (29). Table 8a reports the structural OMD results and the χ^2 statistics for all three specifications. The estimates of ϕ are consistently negative. I also report the structural estimates for the case where the sample excludes two districts with very high adult mortality between 1997 and 2004⁴⁹, what I call the non-HIV sample. The results are much more precise and stable across specifications for this sample.

Tables 8b and 8c present structural estimates of ϕ for the case where fertilizer endogenously enters the farmer's decisions. In Table 8b, I report results for the case where there is a joint hybrid-fertilizer decision on the part of the farmer so that he is in the technology sector if he uses both hybrid and fertilizer⁵⁰, otherwise he is not. Results for ϕ are again consistently negative, and significant mostly in the specifications that exclude the two high mortality districts. Finally, in Table 8c, I allow for fertilizer to be endogenous and correlated with the θ_i 's, using the projection in equation (35). Again, the estimates of ϕ in Table 8c are consistently negative across all OMD specifications, again more so for the sample without the two high mortality districts.

Overall, to summarize these results, the null hypothesis that ϕ is equal to zero is often rejected and estimates of ϕ are consistently negative.⁵¹ These results hold predominantly for the non-HIV districts in the sample, i.e. after leaving out two of the overall 22 districts in the sample. For these households, the selection into hybrid is negative with the farmers having the lowest yields in non-hybrid having the highest returns to planting hybrid.

⁴⁹Thanks to an anonymous referee for pointing this out. Jayne and Yamano (2004) document how this mortality affects the value of high value crops produced, but not the value or productivity of maize.

⁵⁰The fertilizer decision I use regards chemical fertilizers. The results are similar if I include the use of manure as part of the fertilizer decision.

⁵¹The results without covariates and across all EWMD and DWMD specifications are similar. In addition, the following robustness checks give similar results: dropping the Coast Province, allowing for depreciation of assets used in land preparation (for example, a tractor, which only 2% of households own), valuing all the inputs where possible, and using estimated profits instead of yields.

6.1 Recovering the Distribution of $\hat{\theta}_i$

Given the CRC structural estimates of the λ 's and the form of the projection given by either equations (29) or (35), the θ_i 's can be predicted for a given history of hybrid (and fertilizer) use. However, to do so, I must assume that the projections describe the true conditional expectation of the θ_i 's. This is essentially an assumption only for the case of equation (35). For equation (29), since h_{it} is binary and each history is accounted for, the projection is essentially saturated (for example, polynomials in adoption would be redundant since adoption is a dummy variable). Using the predicted distribution of θ_i , the predicted τ_i 's can be derived via the yield function. Intuitively, the process of recovering the distribution of θ_i is building a counterfactual set of returns for the non-adopters using a weighted average of the experience of every 'type' of farmer, where 'type' refers to the farmer's hybrid history.

For the simple two period model (Table 8a), since the predicted θ_i 's are obtained from the projection in equation (29), the distribution of the predicted θ_i 's has just four mass points. There are only four possible hybrid histories: non-hybrid stayers, hybrid stayers, leavers and joiners. The results indicate that the non-hybrid stayers have the most negative $\hat{\theta}_i$, followed by the hybrid stayers, then the leavers and joiners. While the non-hybrid stayers have the lowest predicted θ_i 's, they have the highest returns to hybrid as $\phi < 0$.⁵²

Using the projection in equation (35), I can similarly estimate the distribution of θ_i using the estimates in Table 8c. Here, the distribution of the predicted $\hat{\theta}_i$'s is continuous since the predicted $\hat{\theta}_i$'s come from the projection in equation (35), which includes the amount of fertilizer used. To plot the distribution of the $\hat{\theta}_i$'s and the overall return to hybrid, I use the estimates from the third column of Table 8c, which excludes the two HIV districts and controls only for covariates. This is a simpler model to compute the overall return to hybrid since hybrid is not interacted with all the covariates. The results are not different across the last two columns in Table 8c, so I opt to use this simpler model for the purposes of plotting the distribution of the comparative advantage and the returns.

Figure 5a shows the means of the $\hat{\theta}_i$ by transition - again, the non-hybrid stayers have the most negative $\hat{\theta}_i$, followed by the hybrid stayers, then the leavers and joiners. Figure 5b shows the distribution of returns across the sample by transition. The distribution of returns is given by $\beta + \phi\theta_i$ where β is the structural coefficient on h_{it} in equation (20), i.e. the average return to hybrid. Since $\phi < 0$, the ordering of returns is the reverse of the ordering of θ_i 's, i.e. the non-hybrid stayers have the most negative θ_i 's, but the highest positive returns. The joiners and leavers have close to zero returns (they are the marginal farmers) and the hybrid stayers have lower positive returns. Note that the sign of ϕ determines the ordering of the magnitude of the returns across transitions, this ordering is by no means mechanical. Figure 5c shows the distribution of the predicted $\hat{\theta}_i$'s by transition. Finally, as a check, Figure

⁵²I do not report these estimates as they do not account for the endogeneity of fertilizer (see Suri (2006)).

5d shows the corresponding distribution of the predicted τ_i 's, which were constructed to be orthogonal to the hybrid choice.

7 Discussion

Figure 5b resolves one puzzle about low adoption rates: the hybrid joiners and leavers are marginal farmers in the sense of having very low returns as compared to the average return. However, while the empirical results resolve this one puzzle of why some farmers move in and out of hybrid despite high returns for the average farmer, the results indicate a further puzzle: the very large counterfactual returns to growing hybrid for the non-hybrid stayers. To examine this latter puzzle, I return to the decision rule in equation (25) and relate the estimated gross returns to observable supply and demographic characteristics of the farmers (just for the sample excluding the two HIV districts). Table 9 illustrates that the households with the lower predicted $\hat{\theta}_i$'s have much higher cost determinants. Recall that throughout the sample period, the price of hybrid seed was fixed across the country. This meant that seed suppliers had no incentives to locate far away from markets or roads as there was no compensating price increase.

The regressions in Table 9 therefore correlate the predicted $\hat{\theta}_i$'s with observable cost measures in the data.⁵³ The observables considered are the distance from the household to the closest fertilizer seller, distance to the closest matatu (public transport) stop, distance to the closest motorable road, distance to the closest extension services, a set of dummies for education of the household head, a dummy for whether the household tried to get credit, a dummy for whether the household tried to get credit but did not receive any, a dummy for whether the household received any credit and province dummies. Villages are large and heterogeneous and these measures of costs and infrastructure vary even within a village. The infrastructure and education variables correlate strongly with the predicted $\hat{\theta}_i$'s, but the credit variables do not. The last column shows similar results without province dummies.⁵⁴

These results indicate that while there are high potential returns to the farmers who never grow hybrid during the sample period, they face higher costs and supply constraints. The IV/LATE estimates in Table 4 also suggested that distance and infrastructure are constraining factors. From Table 4, the IV estimates are about 150%, which implies that those affected by the infrastructure instrument are the non-hybrid stayers since these IV estimates

⁵³As pointed out by an anonymous referee, given how the $\hat{\theta}_i$'s are estimated, it could induce a mechanical positive correlation between the estimated $\hat{\theta}_i$'s and the costs even if no such correlation exists. However, simulating the model for a range of values of ϕ showed that the bias is in the opposite direction to the correlations reported in Table 9 and must therefore be small. Whenever $\phi < 0$, for example, the bias would always mean a positive coefficient between costs and the estimated $\hat{\theta}_i$'s, whereas the correlations in Table 9 are all negative.

⁵⁴These results are all for the model with endogenous hybrid and fertilizer (estimates from Table 8c).

are of the same magnitude as the estimated returns for the non-hybrid stayers.

Thus, given existing seed supply locations and infrastructure constraints, the adoption decisions appear to be quite rational. Liberalizing the seed market and seed prices may have large benefits. In addition, while I do not have the data for a social return calculation to expanding supply, encouraging greater seed supply to the more remote areas of Kenya could have large benefits, but will also have large costs of expanding infrastructure.

7.1 Competing Explanations

Finally, I discuss some alternative explanations of adoption behavior that are popular in the literature. The first is a model of learning, where households are uninformed about the benefits of a technology and they experiment to learn what the returns are. Learning models give rise to the familiar S-shaped curves of adoption rates over time (see Griliches (1957)). In my data, adoption rates show no dynamics. Aggregate adoption remains constant over the span of almost ten years. In fact, there are no dynamics in any of the inputs, even after a large rainfall shock due to El Nino in 1998. Any conventional model of learning does not appear to be borne out by these aggregates, though there may have been learning effects closer to the introduction of hybrid in Kenya - Gerhart (1975) shows S-shaped curves in the 1970's, curves that flattened out at low aggregate adoption rates prior to 1997.

The switching between hybrid and non-hybrid from period to period may look like experimentation, but recall that hybrid and fertilizer have been available for a few decades now. In addition, 90% and 83% of households have used hybrid and fertilizer before, respectively. Table 2d does not show any systematic persistence in adoption; the cycling behavior is about as prevalent as the non-cycling. In the qualitative parts of the survey, farmers who were using traditional varieties were asked why they were not using hybrid and only about 0.3% cited experimenting or on trial. And, Duflo, Kremer and Robinson (2008a) find little evidence of learning in Western Kenya. Learning therefore seems to be unimportant in this setting.⁵⁵

There are different varieties of hybrid, so it may be possible that households are learning about these types and are switching to newer varieties over time. In this case, there should still be aggregate increases in adoption over time. For two periods of the survey (2002 and 2004), the data include exactly what type of hybrid variety was used. In 2002, 61.5% of maize plots were planted with hybrid 614, 6% each of 625 and 627 and 4% of 511. The release date of these varieties was 1986, 1981, 1996 and the 1960's respectively. In 2004, 69.5% of plots were planted with 614, 5% with 627, 4% each with 625 and 511. It seems that most households use rather old varieties of hybrid seed. As a further check, I categorize the

⁵⁵On the methodological side, a learning model with heterogeneity in returns would imply a random coefficient model with feedback. Chamberlain (1993) shows that such models suffer from identification problems. Gibbons et al. (2005) estimate a random coefficient model with learning, but they must rely on the structure of the learning process and an extremely long panel (the NLSY) to identify the model.

type of hybrid used in 2004 as new or not, where anything released after 1990 is defined as new. I cannot look at leavers as I do not know what hybrid type they used in 1997, and in 2004 they do not use hybrid. Of the hybrid stayers, about 11.5% use new hybrids and 12.9% of the joiners use new hybrids. These two means are not statistically significantly different from each other (the p-value is 0.702).

A second set of alternative explanations deals with the role of credit constraints. Credit constraints do not seem to be of first order importance for several reasons. First, a strikingly large number of households are able to get access to credit, especially for agricultural purposes, and of those that try to get credit, most get it. In 2004, 41% of all households tried to get credit and 83% got it. In 1997, these figures were collected for agricultural credit, where 33.7% of households tried to get agricultural credit and 90% got it. There are no significant differences between the hybrid stayers, non-hybrid stayers, leavers and joiners in the three credit variables (whether the farmer tried to get credit, whether he received it and whether he tried to get credit but did not receive it). The p-values on these F-tests are all above 0.20. Also, none of these credit variables correlate strongly with comparative advantage in Table 9. A caveat here is that these variables are not necessarily good measures of credit constraints, since households who may be constrained may not even try to get credit.

The patterns in Table 5 also do not fit a *pure* liquidity constraints story. Say the variation in adoption is completely explained by liquidity constraints. In this case, if a non-hybrid farmer does better than average in terms of yields, it will lead him to use hybrid in the next period. But, in that second period, he should have the average hybrid yield, such that non-hybrid households that have higher yields than average in the first period should have no different than average hybrid yields in the second period. So, for 1997-2004, the coefficient on joiners would be greater than zero in 1997, and equal to the coefficient on the hybrid stayers in 2004. Similarly, the leavers should do worse than the hybrid stayers in 1997 and no different than the non-hybrid stayers in 2004. These joint restrictions are rejected in Table 5. Liquidity constraints alone do not seem to explain the patterns in the data.

The final alternative hypothesis is differences in tastes between hybrid and non-hybrid (as in Latin America). In Kenya, hybrid and non-hybrid output are indistinguishable and sold at the same price. The survey asked farmers why they chose the variety they plant and an equal fraction of hybrid and non-hybrid farmers said it was because they preferred the taste. The differential tastes hypothesis does not seem to fit the Kenyan case.

8 Conclusion

This paper examines the adoption decisions and benefits of hybrid maize in the context of Kenya, where adoption trends have been relatively constant for the past decade, yet vary considerably across space. To examine this setting, in contrast to much of the literature

on technology adoption, I use a framework that abandons learning about a homogeneous technology, and that instead considers a model and empirical approach that allows for heterogeneous returns to hybrid that correlate with the hybrid adoption decision. My framework emphasizes the large disparities in farming and input supply characteristics across the maize growing areas of Kenya. Within this framework, I find that for the non-HIV districts in the sample, returns to hybrid maize vary greatly. Furthermore, in these districts, farmers who are on the margin of adopting and disadopting (and who do so during my sample period) experience little change in yields.

The experimental evidence for Kenya, along with simple OLS and IV estimates, has indicated average high, positive returns to these agricultural technologies. This has indicated a puzzle as to why adoption rates had flattened out well below 100% despite the seemingly large gains that could be had by adopting hybrid. The experimental and IV evidence appeared to indicate that maize growers in Kenya were leaving “money on the table” by not adopting hybrid seed. However, the experiments and IV results reveal little else about the returns. The framework of heterogeneous returns in this paper allows the estimation of not just average returns, but also the distribution of returns across the sample of farmers. In contrast, I find strong evidence of heterogeneity in returns to hybrid maize, with comparative advantage playing an important role in the determination of yields and adoption decisions.

My findings regarding the heterogeneity in returns have important implications for policy. Looking at a distribution of returns across the sample of farmers allows me to separate out farmers with low returns from those with high returns into a group that could be targeted by policy interventions. For a small group of farmers in the sample (only 20% of the sample), returns to growing hybrid (as opposed to their observed choice of non-hybrid) would be extremely high, yet they do not adopt hybrid maize. I show that these farmers have high fixed costs that prevent them from adopting hybrid, as they have poor access to seed and fertilizer distributors. In terms of policy, alleviating these constraints would increase yields for these farmers though this may not be a socially optimal intervention. Furthermore, liberalization of the seed supply market and of hybrid seed prices might encourage seed suppliers to locate in the more distant areas. For a large fraction of my sample, the returns to hybrid maize are smaller and these farmers choose to adopt. While I do not build risk into the choice framework used in this paper, farmers may use hybrid maize, even when the mean returns to hybrid are low, as it helps insure them against bad outcomes. For these farmers, since they do not seem to be constrained, they would benefit from improved R&D efforts in developing new hybrid strains and a biotechnology effort similar to that of, say, India, where releases of newer hybrids occurs often and leads to continual yield improvements. Finally, the heterogeneity in returns to the hybrid technology, on the whole, suggests quite rational and relatively unconstrained adoption of existing hybrid strains, in contrast to the evidence from the experimental and IV literatures.

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).
- Attanasio, Orazio, Costas Meghir, and Miguel Szekely, (2003), “Using Randomized Experiments and Structural Models for Scaling Up: Evidence from the PROGRESA Evaluation”, Mimeo, University College London.
- Bandiera, Oriana, and Imran Rasul, (2003), “Complementarities, Social Networks and Technology Adoption in Mozambique”, Working Paper, London School of Economics.
- Besley, Timothy, and Anne Case, (1993), “Modelling Technology Adoption in Developing Countries”, *AER Papers and Proceedings*, 83(2), pp. 396-402.
- Björklund, Anders, and Robert Moffitt, (1987), “The Estimation of Wage Gains and Welfare Gains in Self Selection Models”, *Review of Economics and Statistics*, pp. 42-49.
- Card, David, (2000), “Estimating The Returns to Schooling: Progress of Some Persistent Econometric Problems”, NBER Working Paper No. 7769, Cambridge, MA.
- Card, David, (1998), “The Causal Effect of Education on Earnings”, in the Handbook of Labour Economics, eds. Orley Ashenfelter and David Card.
- Card, David, (1996), “The Effect of Unions on the Structure of Wages: A Longitudinal Analysis”, *Econometrica*, 64.
- Card, David, and Daniel Sullivan, (1988), “Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment”, *Econometrica*, 56.
- Carneiro, Pedro, Karsten Hansen, and James Heckman, (2003), “2001 Lawrence R. Klein Lecture: Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice”, *International Economic Review*, 44 (2), pp. 361-422.
- Carneiro, Pedro, and James Heckman, (2002), “The Evidence on Credit Constraints in Post-Secondary Schooling”, *Economic Journal*, 112 (October), pp. 705-734.
- Carneiro, Pedro, and Sokbae Lee, (2004), “Comparative Advantage and Schooling”, Working Paper, University College London.
- Chamberlain, Gary, (1993), “Feedback in Panel Data Models”, *mimeo*, Department of Economics, Harvard University.
- Chamberlain, Gary, (1984), “Panel Data”, in Zvi Griliches and Michael Intriligator (eds.), *Handbook of Econometrics*, Amsterdam: North-Holland.

- Chamberlain, Gary, (1982), "Multivariate Regression Models for Panel Data", *Journal of Econometrics*, 18, pp. 5-46.
- 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).
- Dercon, Stefan and Luke Christiaensen, Luc, (2005), "Consumption Risk, Technology Adoption and Poverty Traps: Evidence from Ethiopia", *mimeo*.
- Dixon, Robert, (1980), "Hybrid Corn Revisited", *Econometrica*, 48 (6), pp 1451-1461.
- 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, (2008a), "Nudging Farmers to Use Fertilizer: Evidence from Kenya", Working Paper, 2008.
- Duflo, Esther, Michael Kremer, and Jonathan Robinson, (2008b), "How High Are Rates of Return to Fertilizer? Evidence From Field Experiments in Kenya", Prepared for 2008 American Economic Association Meetings.
- 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*.
- 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.

- Garen, John, (1984), "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable", *Econometrica*, 52 (5), pp. 1199-1218.
- 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 Differences", *Journal of Labor Economics*, 23, pp. 681-723.
- 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.
- Griliches, Zvi, (1980), "Hybrid Corn Revisited: A Reply", *Econometrica*, 48 (6), pp. 1463-1465.
- Hall, Bronwyn, (2004), "Innovation and Diffusion", NBER Working Paper No. 10212.
- 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.
- 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, 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, and Joel Ransom, (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.
- Heckman, James, and Xuesong Li, (2003), "Selection Bias, Comparative Advantage and Heterogeneous Returns to Education: Evidence from China in 2000", IFAU - Institute for Labour Market Policy Evaluation, Working Paper 2003:17.

- Heckman, James, Justin Tobias, and Edward Vytlačil, (2001), “Four Parameters of Interest in the Evaluation of Social Programs”, *Southern Economic Journal*, 68 (2), pp. 210-223.
- 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 Vytlačil, (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).
- Heckman, James, and Bo Honore, (1990), “The Empirical content of the Roy Model”, *Econometrica*, 58 (5), pp. 1121-1149.
- 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, MSU.
- 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.
- Jayne, Thomas, and Takashi Yamano, (2004), “Measuring the Impacts of Working-age Adult Mortality on Small-Scale Farm Households in Kenya”, *World Development*, 32(1), pp. 91-119.
- 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.
- 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.
- 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, CRDE and Departement de Sciences Economiques, Universite de Montreal.
- 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.
- Levinsohn, James, and Amil Petrin, (2003), “Estimating Production Functions Using Inputs to Control for Unobservables”, *Review of Economic Studies*, 70, pp. 317-341.
- 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.

- McCann, James C., (2005), *Maize and Grace: Africa's Encounter with a New World Crop, 1500-2000*, Harvard University Press, USA.
- Morons, Lawrence, James Mutunga, Francis Munene, and Boaz Cheluget, "HIV Prevalence and Associated Factors", DHS Report, Chapter 13.
- Moser, Christine, and Christopher Barrett, (2003), "The Complex Dynamics of Smallholder Technology Adoption: The Case of SRI in Madagascar", Working Paper, Cornell.
- Mugo, Frida, Frank Place, Brent Swallow, and Maggie Lwayo, (2000), "Improved Land Management Project: Socioeconomic Baseline Study", Nyando River Basin Report, ICRAF.
- 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.
- Nyameino, David, Bernard Kagira, and Stephen Njuki, (2003), "Maize Market Assessment and Baseline Study for Kenya", RATES Program, Nairobi, Kenya.
- Nyoro, James K., M. W. Kiiru, and Thom S. Jayne, (1999), "Evolution of Kenya's Maize Marketing Systems in the Post-Liberalization Era", Tegemeo Working Paper, June 1999.
- Olley, Steven and Ariel Pakes, (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry", *Econometrica*, 64, pp. 1263-1297.
- Ouma, James, Festus Murithi, Wilfred Mwangi, Hugo Verkuyl, 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).
- Pischke, Jörn-Steffen, (1995), "Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study", *Journal of Business and Economic Statistics*, 13 (3).
- 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, Free Press, New York.
- Roy, A., (1951), "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers*, 3, pp. 135-146.
- Salasya, M.D.S., W. Mwangi, Hugo Verkuyl, 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", CIMMYT.

- Sanders, John, Barry Shapiro, and Sunder Ramaswamy, (1996), “The Economics of Agricultural Technology in Semiarid Sub-Saharan Africa”, John Hopkins University Press.
- Sattinger, Michael, (1993), “Assignment Models of the Distribution of Earnings”, *Journal of Economic Literature*, 31 (2), pp. 831-880.
- Schultz, Theodore, (1963), *The Economic Value of Education*, Columbia University Press.
- 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.
- 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.
- Suri, Tavneet, (2006), “Selection and Comparative Advantage in Technology Adoption”, Economic Growth Center Discussion Paper No. 944, Yale University.
- Tittonell, Pablo, (2003), “Soil Fertility Gradients in Smallholder Farms of Western Kenya, Their Origin, Magnitude and Importance”, Thesis, Wageningen University, Netherlands.
- Wanzala, Maria, Thom Jayne, John Staatz, Amin Mugeru, 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, CSAE, Oxford University.
- Wekesa, E., Wilfred Mwangi, Hugo Verkuil, K. Danda, and Hugo De Groote, (2003), “Adoption of Maize Technologies in the Coastal Lowlands of Kenya”, Mexico, D.F.: CIMMYT.
- Willis, Robert, (1986), “Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions”, in Orley Ashenfelter and Richard Layard (eds.), *Handbook of Labor Economics*, Amsterdam: North-Holland.
- 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.

Figure 1
Population Density and Location of Sample Villages

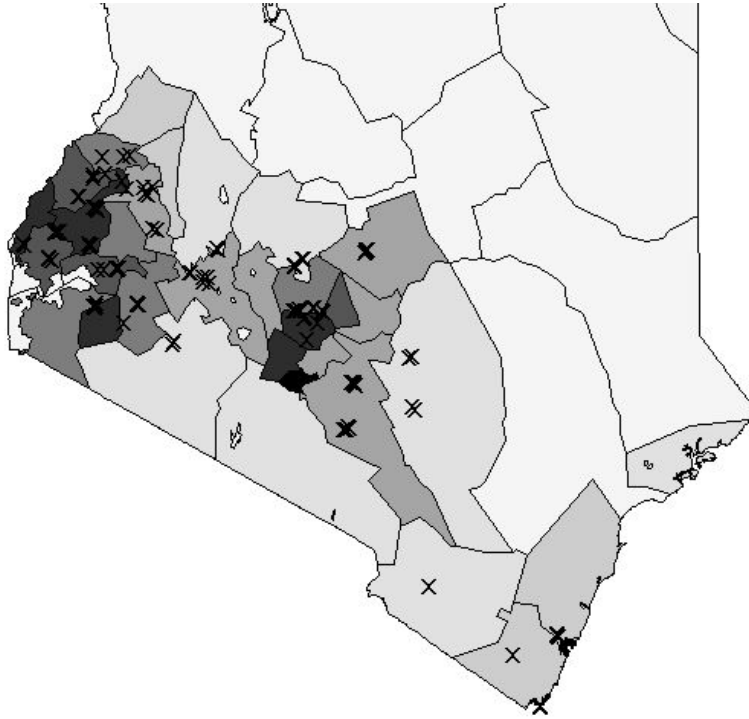


Figure 2
Hybrid Maize Adoption Patterns, by Province

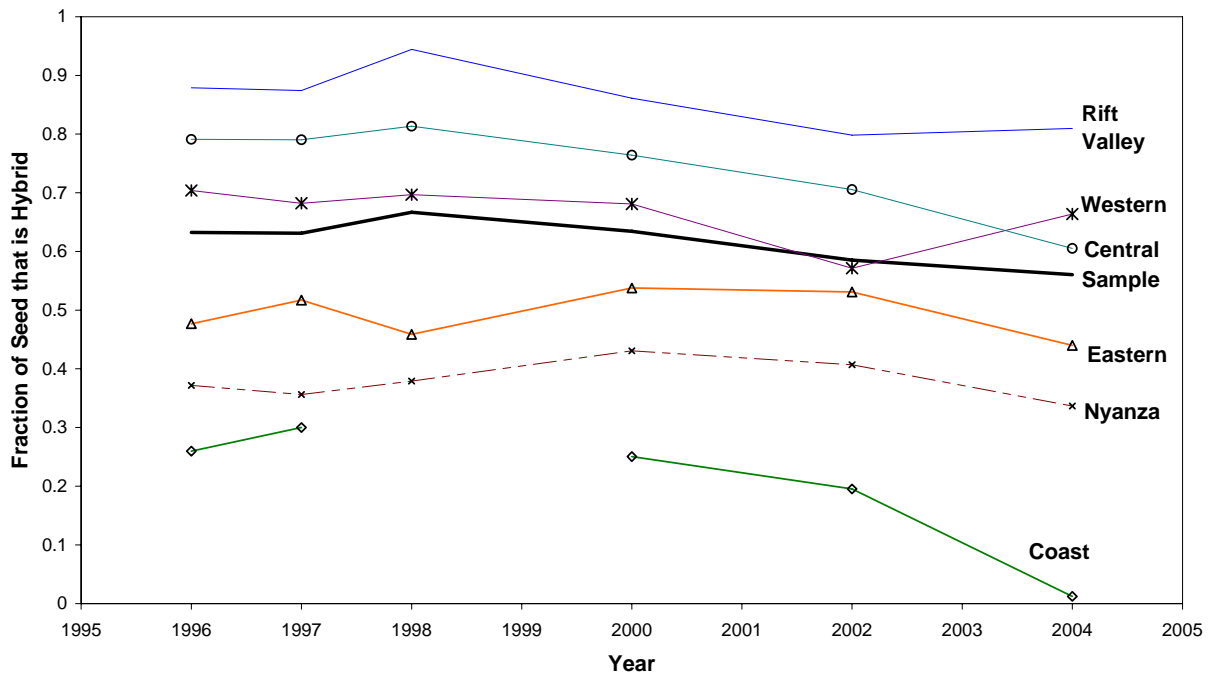


Figure 3a
Fraction of HH's Using Inorganic Fertilizer, by Province

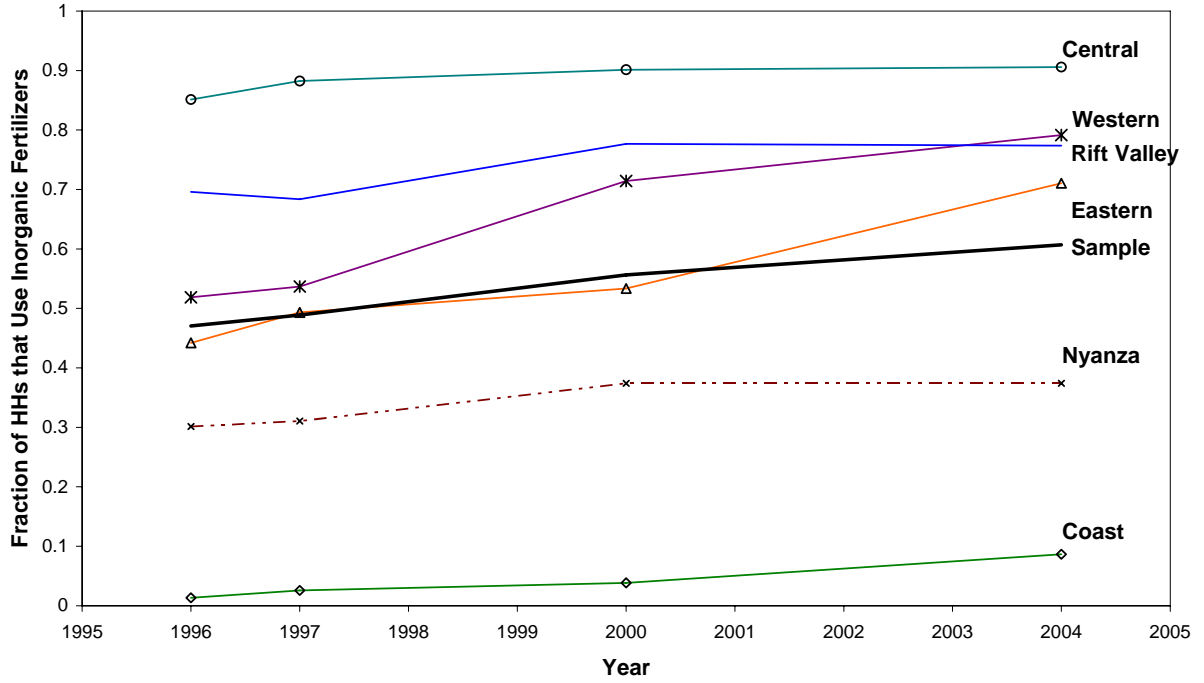


Figure 3b
Real Expenditure on Inorganic Fertilizer, by Province

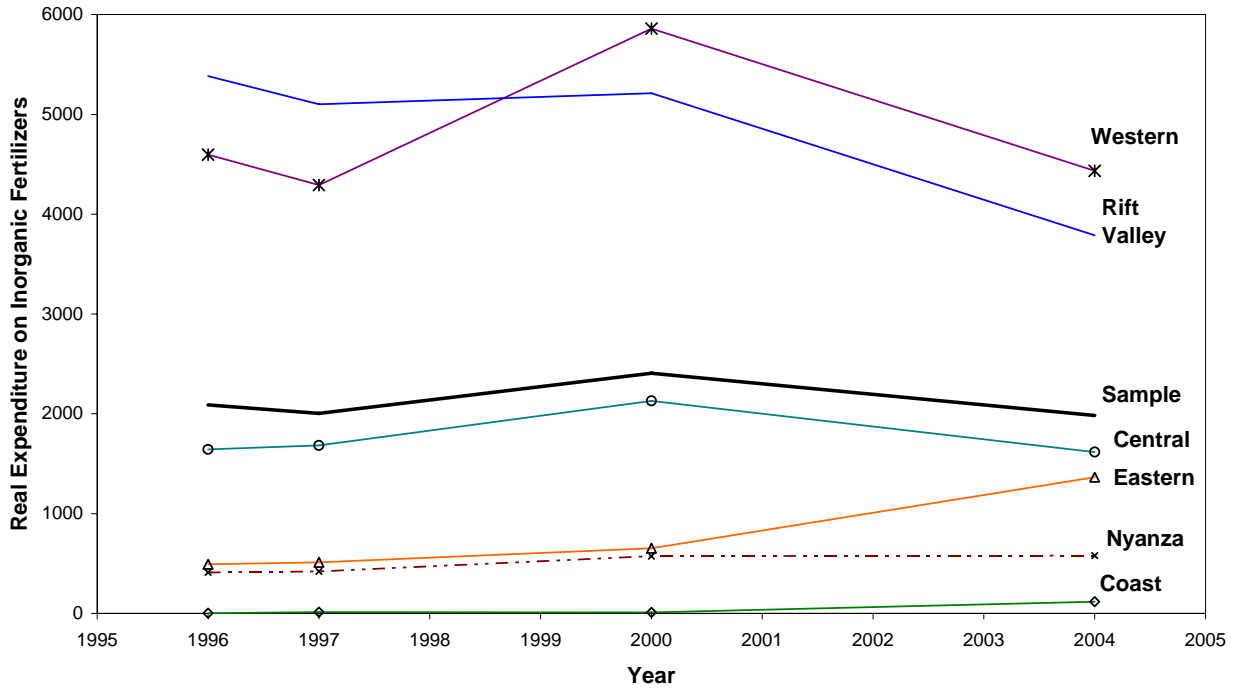


Figure 4a

Marginal Distribution of Yields by Sector, 1997

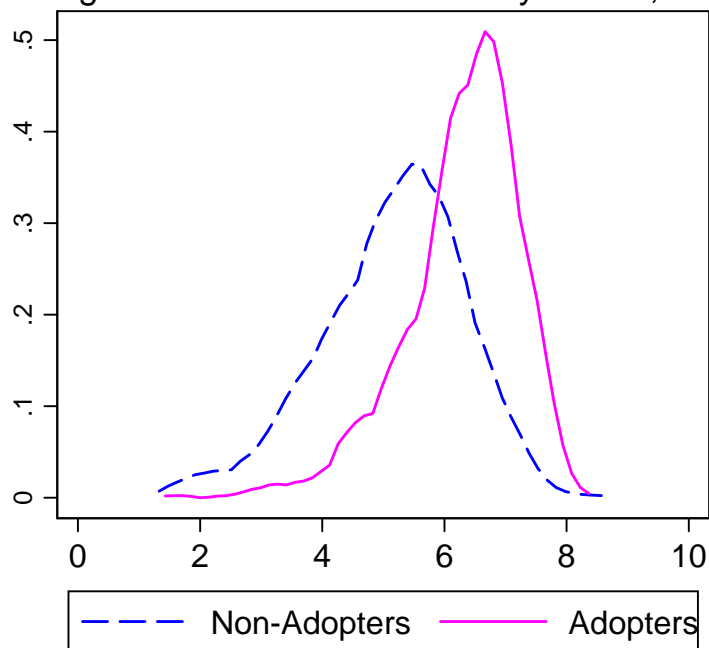


Figure 4b

Marginal Distribution of Yields by Sector, 2004

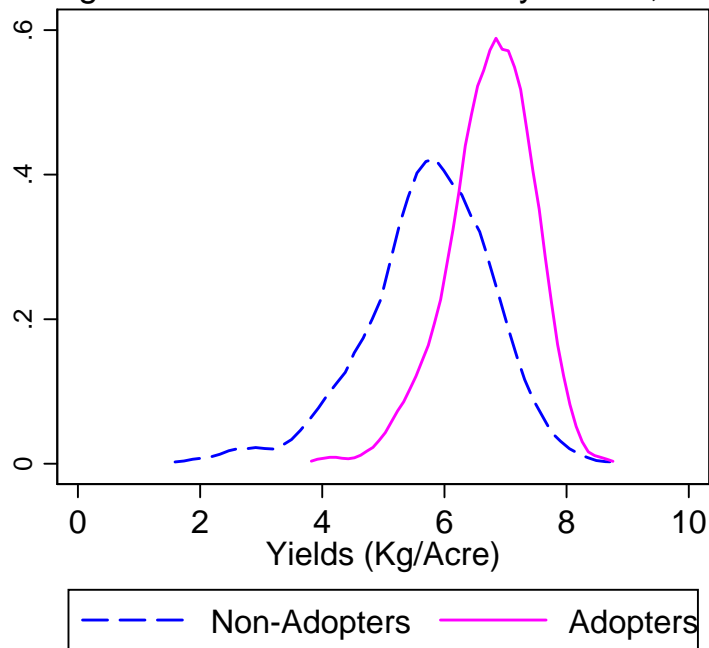


Figure 5a

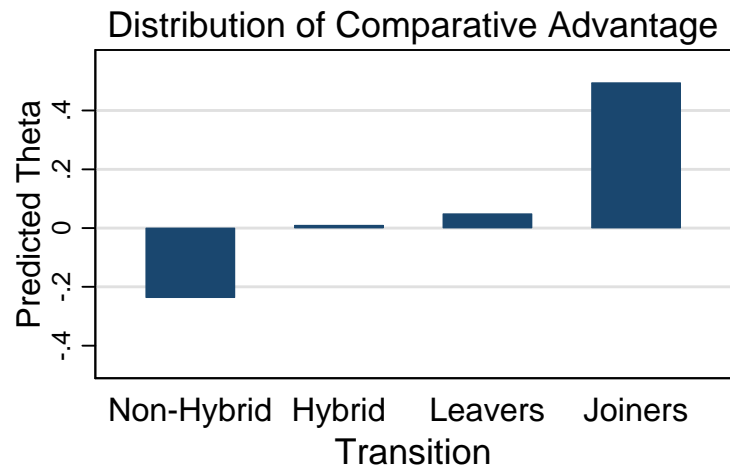


Figure 5b

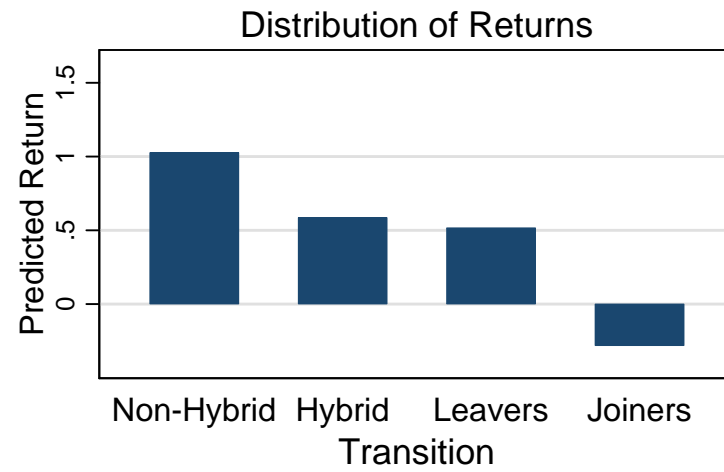


Figure 5c

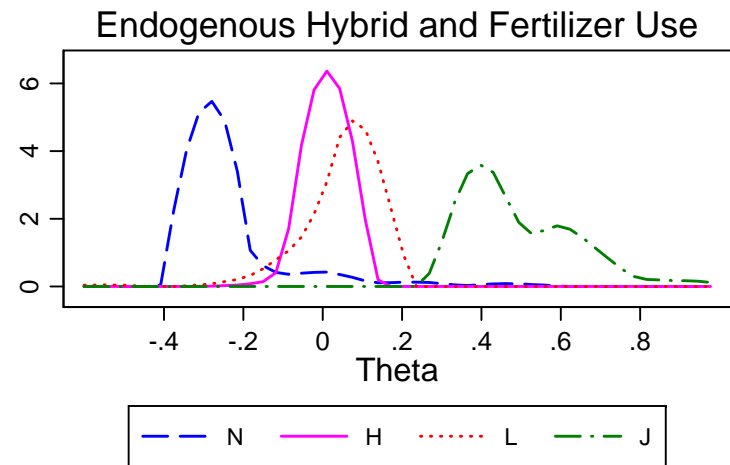
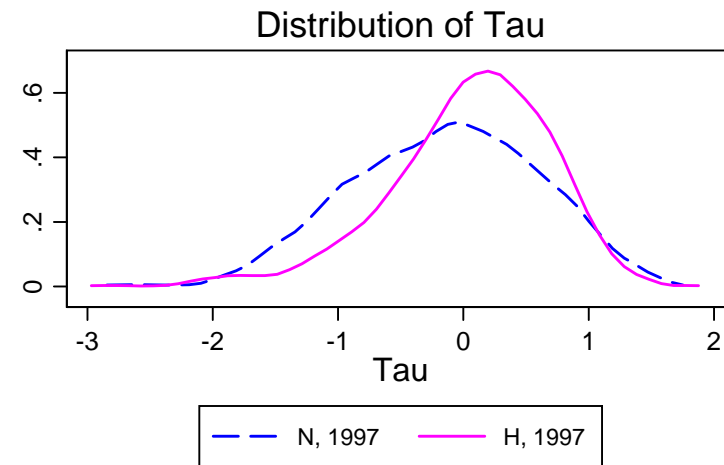


Figure 5d



**Table 1: Trends in Yields of Staples
Average Annual % Changes in Yields (Hg/Ha), by Decade**

	1961-1970	1971-1980	1981-1990	1991-2004
Kenya				
Maize	0.362	2.373	1.169	-1.198
Wheat	5.646	2.333	-3.078	0.984
India				
Maize	1.502	0.842	1.900	2.572
Wheat	4.876	2.514	3.343	1.235
Rice	0.954	1.714	3.310	0.838
Mexico				
Maize	2.057	4.267	-0.548	1.447
Wheat	4.586	3.204	-0.255	1.664
Zambia				
Maize	-0.267	10.403	1.571	-1.707

Notes: Surprisingly, in the 1960's the levels of yields of maize in Kenya, India and Mexico were comparable and very similar (all ranging between 10,000 and 12,000 Hg/Ha). The levels of yields in Zambia were slightly lower (about 8,000 Hg/Ha) whereas those in Uganda and Malawi were comparable to those in Kenya.

Source: FAOSTAT Online Database

Table 2a: Summary Statistics, by Sample Year

	1997 Sample	2004 Sample
Yield (Log Maize Harvest Per Acre)	5.907 (1.153)	6.350 (0.977)
Acres Planted	1.903 (3.217)	1.957 (2.685)
Total Seed Planted (Kg per Acre)	9.575 (7.801)	9.072 (6.863)
Total Purchased Hybrid Planted (Kg per Acre)	6.273 (6.926)	5.080 (5.260)
Hybrid (dummy)	0.658 (0.475)	0.604 (0.489)
Fertilizer (Kg DAP (di-ammonium phosphate) per Acre)	20.300 (38.444)	24.610 (34.001)
Fertilizer (Kg MAP (mono-ammonium phosphate) per Acre)	1.566 (10.165)	0.308 (4.538)
Fertilizer (Kg CAN (calcium ammonium nitrate) per Acre)	6.473 (24.727)	8.957 (21.702)
Total Fertilizer Expenditure (KShs per Acre)	1361.7 (2246.3)	1354.6 (1831.2)
Land Preparation Costs (Kshs per Acre)	960.88 (1237.1)	541.43 (1022.8)
Family Labor (Hours per Acre)	293.25 (347.49)	354.27 (352.68)
Hired Labor (KShs per Acre)	1766.0 (3346.4)	1427.4 (2130.3)
Main Season Rainfall (mm)	620.83 (256.43)	728.11 (293.29)
Distance to Closest Fertilizer Seller (km)	6.288 (9.774)	3.469 (5.964)
Household Size	7.109 (2.671)	8.409 (3.521)

Notes: Standard deviations in parentheses. KShs is Kenyan shillings (exchange rate over this period was KShs 75 = 1\$). All monetary variables are deflated.

Table 2b: Summary Statistics, by Hybrid/Non-Hybrid Use

	1997 Sample		2004 Sample	
	Hybrid	Non-Hybrid	Hybrid	Non-Hybrid
No. of Households	791	411	726	476
Yield (Log Maize Harvest Per Acre)	6.296 (0.934)	5.158 (1.167)	6.751 (0.692)	5.738 (1.030)
Total Maize Acres Cultivated	1.982 (3.557)	1.753 (2.428)	2.087 (3.029)	1.758 (2.042)
Total Seed Planted (Kg per Acre)	9.669 (6.569)	9.394 (9.750)	8.746 (4.156)	9.569 (9.608)
Fertilizer (Kg DAP per Acre)	28.755 (44.115)	4.028 (13.266)	37.148 (37.294)	5.488 (13.909)
Fertilizer (Kg CAN per Acre)	9.087 (29.715)	1.442 (7.152)	12.708 (24.961)	3.235 (13.622)
Land Preparation Costs (KShs/Acre)	1043.9 (1242.7)	801.08 (1211.7)	659.83 (1079.7)	360.83 (901.0)
Expenditure on Fertilizer (KShs/Acre)	1922.3 (2542.9)	282.64 (740.53)	1893.3 (1964.7)	533.09 (1211.4)
Inorganic Fertilizer Use (Dummy)	0.7421 (0.4378)	0.2311 (0.4221)	0.8994 (0.3009)	0.4055(0.4915)
Main Season Rainfall (mm)	651.70 (228.82)	561.44 (293.88)	825.41 (215.20)	579.69 (332.05)
Hired Labor (KShs/Acre)	1864.3 (2680.6)	1576.7 (4347.8)	1616.5 (2197.4)	1139.0 (1991.6)
Family Labor (Hours/Acre)	260.35 (264.13)	356.57 (461.71)	343.6 (336.1)	370.58 (376.33)
Distance to Closest Fertilizer Seller (km)	4.684 (7.993)	9.374 (11.93)	2.419 (2.420)	5.069 (8.760)
Household Size	7.162 (2.616)	7.007 (2.773)	8.457 (3.340)	8.336 (3.783)

Notes: Standard deviations in parentheses. The mean of yields is higher and the standard deviation of yields is lower in the hybrid sector.

**Table 2c: Break Down of Labor Costs
By Hybrid/Non-Hybrid Use**

	1997 Sample		2004 Sample	
	Hybrid	Non-Hybrid	Hybrid	Non-Hybrid
Hired Labor (KShs/Acre)				
Land Preparation	408.4 (1171)	514.2 (3178)	408.2 (869.1)	501.8 (1096)
Planting	159.8 (318.9)	105.4 (293.5)	143.4 (354.1)	76.67 (216.9)
Weeding	678.8 (1149)	651.8 (1484)	635.7 (1088)	403.1 (915.5)
Harvest	365.1 (642.5)	196.6 (766.1)	151.1 (303.5)	96.42 (340.5)
Post-Harvest Activities	236.5 (472.6)	98.65 (308.6)	241.1 576.7	47.37 (182.1)
Fertilizer Application	15.21 (84.67)	10.02 (156.4)	13.24 (79.32)	2.374 (28.04)
Other	0.501 (9.727)	0 (0)	23.84 (281.6)	11.26 (227.5)
Family Labor (Hrs/Acre)				
Land Preparation	47.95 (113.5)	102.4 (214.6)	51.93 (98.22)	97.79 (180.1)
Planting	29.12 (31.81)	35.23 (45.19)	38.27 (56.13)	39.55 (49.01)
Weeding	93.75 (107.1)	127.5 (151.6)	120.1 (151.3)	134.6 (160.2)
Harvest	42.45 (46.04)	49.54 (143.0)	58.77 (72.36)	49.60 (62.80)
Post-Harvest Activities	44.07 (55.77)	39.89 (70.00)	68.54 (91.76)	44.69 (59.98)
Fertilizer Application	3.171 (8.778)	1.935 (15.09)	3.928 (9.976)	1.510 (6.706)
Other	0.047 (1.313)	0 (0)	2.041 (13.62)	2.814 (24.73)

Notes: Standard deviations in parentheses. On average, family labor accounts for 75% of total labor use (the rest being hired labor). The main production activities that use labor are listed above. Land preparation accounts for 23% of total labor use on average, planting accounts for 10%, weeding accounts for 36% and harvesting and post harvesting activities account for 30%. Post harvesting activities include threshing, winnowing, bagging and storing the maize.

**Table 2d: Transitions Across Hybrid/Non-Hybrid Sectors
For the Sample Periods 1997, 2000 and 2004**

Transition in Terms of Technology Used (1997 2000 2004)	Fraction of Sample (%) (N=1202 Households)
N N N	20.38
N N H	2.83
H N H	6.07
N H H	4.91
H N N	5.99
H N H	3.16
H H N	7.15
H H H	49.50

Notes: This table shows all the possible three period transitions in my sample of farmers, and the fraction of my sample that experiences each of these transitions. The three periods correspond to 1997, 2000 and 2004. In the first column, the three letters represent the transition history with respect to technology, where “H” represents the use of hybrid and “N” represents the use of non-hybrid. For example, the transition “N N N” stands for farmers who used non-hybrid maize in all three periods – they make up about 20.4% of my sample.

The survey instrument asks about hybrid use in multiple sections of the questionnaire (since it is a rather large part of household decisions). We check and confirm the coherency of these responses in the field, which greatly reduces the likelihood that the observed switching behavior is appreciably affected by measurement error.

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

	OLS, Pooled	OLS, Pooled	OLS, Pooled	FE	FE
Hybrid	1.074 (0.040)	0.695 (0.039)	0.541 (0.041)	0.017 (0.070)	0.090 (0.065)
Acres (x1000)	-	-	0.035 (5.749)	-	-0.509 (0.140)
Seed Kg per Acre (x10)	-	-	0.184 (0.024)	-	0.179 (0.032)
Land Preparation Costs per Acre (x1000)	-	-	0.066 (0.016)	-	0.075 (0.023)
Fertilizer per Acre (x1000)	-	-	0.075 (0.009)	-	0.054 (0.012)
Hired Labor per Acre (x1000)	-	-	0.037 (0.006)	-	0.027 (0.008)
Family Labor per Acre (x1000)	-	-	0.374 (0.050)	-	0.467 (0.072)
Year = 2004	0.501 (0.038)	0.480 (0.035)	0.566 (0.041)	0.444 (0.032)	0.587 (0.044)
Constant	5.200 (0.038)	4.636 (0.080)	3.954 (0.113)	5.896 (0.051)	-2.383 (5.582)
Province Dummies	No	Yes	Yes	-	-
R-Squared	0.266	0.400	0.502	0.049	0.089

Notes: Standard errors are in parentheses. All regressions have 2404 observations (two periods). Covariates not reported include household size, controls for the age-sex composition of the household (henceforth this includes variables for the number of boys (aged <16 years), the number of girls, the number of men (aged 17-39), the number of women and the number of older men (aged >40 years)), the main season rainfall and the average long term main season rainfall. Results are almost identical if the sample is for three periods without family labor as a covariate. Hired labor is measured in KShs per acre and family labor is measured in hours per acre. The OLS and household fixed effects specifications run are:

$$y_{it} = \delta + \beta h_{it} + X'_{it}\gamma + \varepsilon_{it} \qquad y_{it} = \delta + \alpha_i + \beta h_{it} + X'_{it}\gamma + \varepsilon_{it}$$

Table 3b: CRE Model Reduced Forms and Structural Estimates
Dependent Variable is Yields (Log Maize Harvest Per Acre)

Reduced Form Estimates

	Without Covariates		With Covariates		With Covariates and Interactions of Covariates with Hybrid	
	<u>Yields, 1997</u>	<u>Yields, 2004</u>	<u>Yields, 1997</u>	<u>Yields, 2004</u>	<u>Yields, 1997</u>	<u>Yields, 2004</u>
Hybrid, 1997	0.674 (0.075)	0.538 (0.065)	0.579 (0.064)	0.415 (0.060)	0.467 (0.242)	0.501 (0.228)
Hybrid, 2004	0.809 (0.072)	0.723 (0.062)	0.411 (0.065)	0.563 (0.063)	1.214 (0.259)	0.630 (0.230)

Optimal Minimum Distance (OMD) Structural Estimates

	Without Covariates	With Covariates	With Covariates and Interactions of Covariates with Hybrid
β	0.0322 (0.0701)	0.1588 (0.0653)	-0.3039 (0.2522)
λ_1	0.5795 (0.0621)	0.4166 (0.0570)	0.5683 (0.2103)
λ_2	0.7332 (0.0684)	0.4062 (0.0622)	1.0447 (0.2351)
χ^2_1	44.63	0.193	460.5

Notes: Standard errors are in parentheses. Reduced forms are estimated without covariates, with covariates (acreage, real fertilizer expenditure, real land preparation costs, seed, labor, household size, household age-sex composition variables and rainfall variables) and with all covariates interacted with hybrid. The covariates are all treated as endogenous. Results are the same if three periods without family labor are used. Results are similar if all covariates except hybrid are treated as exogenous.

The reduced form for each t , the projection used to estimate the structural model by minimum distance and the structural model are respectively:

$$y_{it} = \delta_t + \gamma_1 h_{i1} + \gamma_2 h_{i2} + \gamma_3 h_{i3} + X_i' \pi_t + \epsilon_{it} \quad \alpha_i = \lambda_0 + \lambda_1 h_{i1} + \lambda_2 h_{i2} + v_i \quad y_{it} = \delta + \beta h_{it} + \alpha_i + u_{it}$$

Structural coefficients β and projection coefficients (λ 's) are reported. OMD are optimal (weighted by inverted reduced form variance-covariance matrix) minimum distance estimates. Equally weighted (using the identity matrix) and diagonally weighted (using only the diagonal elements from the OMD weight matrix) minimum distance results are similar.

Table 4: Heckit and Treatment Effect Estimates under Non-Random Assignment (ATE, TT, MTE, LATE)

Year	Heckman Two-Step Estimates: Selection Correction λ		Implied Treatment Effects		
	Hybrid Sector	Non-Hybrid Sector	ATE	TT	MTE Slope
1997	-0.854 (0.170)	1.659 (0.864)	2.391	0.917	-2.512 (0.880)
2004	-0.957 (0.181)	0.028 (0.152)	1.279	0.921	-0.985 (0.237)
IV (LATE) Estimates (Conditional on Covariates)					
First Stage: Effect of Distance			-0.288 (0.108)	-	
First Stage: Effect of Distance Interacted with Wealth Quintile (x 100)					
	Second Wealth Quintile (Coefficient on Interaction)		-	-0.221 (0.302)	
	Third Wealth Quintile (Coefficient on Interaction)		-	-0.057 (0.032)	
	Fourth Wealth Quintile (Coefficient on Interaction)		-	0.329 (0.288)	
	Fifth Wealth Quintile (Coefficient on Interaction)		-	0.507 (0.273)	
F Test p-value on Excluded Instruments			0.008	0.108	
Second Stage: Effect of Predicted Hybrid on Yields			2.768 (1.123)	1.536 (0.816)	

Notes: Standard errors are in parentheses. All regressions control for covariates. The first two upper panel columns show (for each year) the two-step selection corrected estimates of coefficients on the Inverse Mills ratio for hybrid and non-hybrid sectors from:

$$h_i = Z_i' \pi + u_i^s$$

$$y_i^H = X_i' \gamma^H + \lambda^H [\phi(Z_i' \hat{\pi}) / \Phi(Z_i' \hat{\pi})] \quad \text{and} \quad y_i^N = X_i' \gamma^N + \lambda^N [\phi(Z_i' \hat{\pi}) / (1 - \Phi(Z_i' \hat{\pi}))]$$

The third column reports the average treatment effect accounting for selection and the fourth column the treatment on the treated. Finally, the MTE slope is just the difference in the λ coefficients for the hybrid and non-hybrid selection terms (the difference between coefficients reported in the first two columns). ATE, TT, MTE are all evaluated at the mean X_i 's. The lower panel reports two set of IV estimates. The first column uses the distance to *closest* fertilizer supplier (not where fertilizer is purchased) as the excluded instrument. The second column uses the distance interacted with wealth quintiles as excluded instruments (controlling for the asset quintile dummies and the distance main effect in both stages).

All regressions control for the full set of covariates as per earlier tables (including household size and controls for the age-sex composition of the household as above).

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

Variable	Without Covariates		With Covariates	
	1997 Yield	2004 Yield	1997 Yield	2004 Yield
Hybrid Stayers	1.505 (0.066)	1.280 (0.056)	0.869 (0.073)	0.683 (0.063)
Leavers	0.809 (0.094)	0.648 (0.079)	0.537 (0.084)	0.370 (0.069)
Joiners	1.007 (0.114)	0.883 (0.096)	0.469 (0.101)	0.498 (0.084)
Acres (x100)	-	-	0.561 (0.782)	-0.744 (0.802)
Seed Kg per Acre (x10)	-	-	0.218 (0.035)	0.197 (0.032)
Land Preparation Costs per Acre (x1000)	-	-	0.066 (0.023)	0.058 (0.021)
Fertilizer per Acre (x1000)	-	-	0.063 (0.012)	0.061 (0.012)
Hired Labor per Acre (x1000)	-	-	0.028 (0.008)	0.057 (0.010)
Family Labor per Acre (x1000)	-	-	0.415 (0.075)	0.318 (0.064)

Notes: Standard errors are in parentheses. In each year, the following regression is run both with and without the covariates:

$$y_i = \delta + \mu_1 h_{i11} + \mu_2 h_{i10} + \mu_3 h_{i01} + X_i' \pi + u_i$$

where h_{i11} is the dummy indicating that farmer i is a hybrid stayer (plants hybrid in both periods), h_{i10} indicates he is a leaver (plants hybrid the first year, not the second), h_{i01} indicates he is a joiner (plants hybrid the second year, not the first). The coefficients reported are μ_1 , μ_2 and μ_3 . Covariates included that are not reported above are province dummies, household size, variables for the age-sex composition of the household (includes variables for the number of boys (aged <16 years), the number of girls, the number of men (aged 17-39), the number of women and the number of older men (aged >40 years)), the main season rainfall and the average long term main season rainfall.

Note that hired labor is measured in KShs per acre and family labor is measured in hours per acre.

**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
Acres (x 10)	0.144 (0.056)	-0.053 (0.149)	-0.381 (0.153)	-0.941 (0.379)
Seed Kg per Acre (x 10)	0.281 (0.035)	0.129 (0.036)	0.219 (0.047)	0.147 (0.063)
Land Preparation Costs per Acre (x 1000)	0.056 (0.018)	0.135 (0.031)	0.033 (0.024)	0.097 (0.060)
Fertilizer per Acre (x 1000)	0.064 (0.008)	0.143 (0.032)	0.040 (0.011)	0.081 (0.086)
Hired Labor per Acre (x1000)	0.047 (0.007)	0.026 (0.010)	0.054 (0.011)	0.035 (0.029)
Family Labor per Acre (x1000)	0.297 (0.064)	0.435 (0.081)	0.497 (0.094)	0.581 (0.177)
Year = 2004	0.568 (0.050)	0.595 (0.068)	0.467 (0.058)	0.689 (0.096)
Average Return (β) when Returns Vary by Observables (evaluated at mean X's)		0.480 (0.048)		0.091 (0.076)
Number of Observations	1517	887	1517	887

Notes: Standard errors in parentheses. The following specification is estimated separately for the sample of farmers who use hybrid and non-hybrid maize:

$$y_{it}^j = \delta^j + X_{it}\gamma^j + \epsilon_{it}^j \quad j \in H, N$$

The covariates included that are not reported are province dummies, household size, age-sex composition of the household (includes variables for the number of boys (aged <16 years), the number of girls, the number of men (aged 17-39), the number of women and the number of older men (aged >40 years)), the main season rainfall and the average long term main season rainfall.

The results are similar for the three period sample without family labor as a covariate.

Note that hired labor is measured in KShs per acre and family labor is measured in hours per acre.

**Table 7: Two Period Basic Comparative Advantage CRC Model Reduced Form Estimates
Dependent Variable is Yields (Log Maize Harvest Per Acre)**

	Without Covariates		With Endogenous Covariates		With Interactions with Hybrid	
	Yields, 1997	Yields, 2004	Yields, 1997	Yields, 2004	Yields, 1997	Yields, 2004
Hybrid, 1997	0.833 (0.121)	0.471 (0.099)	0.719 (0.103)	0.316 (0.088)	0.926 (0.252)	0.139 (0.092)
Hybrid, 2004	1.139 (0.122)	0.766 (0.103)	0.702 (0.110)	0.508 (0.092)	0.474 (0.122)	0.520 (0.222)
Hybrid 1997*Hybrid 2004	-0.458 (0.156)	-0.194 (0.132)	-0.358 (0.132)	-0.098 (0.110)	-0.084 (0.147)	-0.115 (0.117)
Acres (x10)	-	-	0.106 (0.067)	-0.006 (0.098)	-0.220 (0.302)	-0.799 (0.300)
Seed Kg per Acre (x 10)	-	-	0.230 (0.052)	0.433 (0.060)	0.211 (0.062)	0.200 (0.086)
Land Preparation Cost per Acre (x 1000)	-	-	0.124 (0.025)	0.133 (0.039)	0.033 (0.051)	0.353 (0.077)
Fertilizer per Acre (x 1000)	-	-	0.079 (0.018)	0.042 (0.014)	0.281 (0.073)	0.106 (0.035)
Hired Labor per Acre (x 1000)	-	-	0.025 (0.014)	0.053 (0.010)	0.008 (0.014)	0.049 (0.019)
Family Labor per Acre (x 1000)	-	-	0.399 (0.115)	0.186 (0.071)	0.676 (0.187)	0.198 (0.128)
R-Squared	0.285	0.232	0.454	0.441	0.486	0.489

Notes: Standard errors are in parentheses. Reduced forms are estimated with standard covariates and then with interactions of all the covariates with hybrid. Covariates not reported include main season rainfall, household size and age-sex composition of the household (includes variables for the number of boys (aged <16 years), the number of girls, the number of men (aged 17-39), the number of women and the number of older men (aged>40 years)). Note that hired labor is measured in KShs per acre and family labor is measured in hours per acre. Where the covariates are interacted with hybrid, only main effects of the covariates are reported. Reduced forms are for the case where all covariates are exogenous: only hybrid is correlated with the comparative advantage, θ_i . See Table 8 for the projection and structural estimates. The reduced form equations run are as follows for 1997 and 2004 respectively:

$$y_{i1} = \delta_1 + \gamma_1 h_{i1} + \gamma_2 h_{i2} + \gamma_3 h_{i1} h_{i2} + \xi_{i1} \quad y_{i2} = \delta_2 + \gamma_4 h_{i1} + \gamma_5 h_{i2} + \gamma_6 h_{i1} h_{i2} + \xi_{i2}$$

Table 8a: Two Period Basic Comparative Advantage CRC Model OMD Structural Estimates

	With Only Hybrid Endogenous					
	Full Sample			Without HIV Districts		
	Without Covariates	With Covariates	With Interactions with Hybrid	Without Covariates	With Covariates	With Interactions with Hybrid
λ_1	0.648 (0.093)	0.565 (0.087)	0.456 (0.090)	0.471 (0.099)	0.305 (0.089)	0.139 (0.092)
λ_2	1.007 (0.112)	0.665 (0.104)	0.473 (0.116)	1.139 (0.122)	0.710 (0.112)	0.466 (0.123)
λ_3	1.636 (4.854)	-1.690 (4.316)	-0.485 (0.199)	-4.800 (9.173)	-0.936 (0.308)	-0.497 (0.257)
β	-0.543 (1.874)	1.023 (1.480)	3.534 (24.05)	2.287 (4.222)	0.623 (0.100)	0.790 (0.169)
ϕ	-0.794 (0.411)	-1.317 (1.262)	-17.82 (137.4)	-1.010 (0.228)	-1.518 (0.310)	-2.196 (1.142)
χ^2_1	40.089	11.25	139.5	175.5	114.1	305.2

Notes: Standard errors are in parentheses. The reduced forms for these estimates are reported in Table 7 (for the case where only the hybrid decision is endogenous, i.e. correlated with the θ_i 's). The projection used in this model is:

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

The structural coefficients reported are the average return to hybrid (β), the comparative advantage coefficient (ϕ), and the projection coefficients (λ 's). OMD is optimal weighted (the weight matrix is the inverted reduced form variance-covariance matrix) minimum distance. Results from diagonally weighted (the weight matrix is the OMD weight matrix with the off diagonal elements set to zero) and equally weighted minimum distance are similar. The χ^2 statistic on the overidentification test is the value of the OMD minimand.

Results are reported for two samples: the full sample and the sample without two districts where HIV is prevalent.

In addition, minimum distance results for three different specifications are reported: without covariates, with covariates and with covariates and interactions of the covariates with the hybrid decision (reported in Table 7). All the specifications with covariates assume that all covariates are exogenous: only hybrid is correlated with the comparative advantage, θ_i .

Covariates include acreage, land preparation costs, fertilizer, hired labor, family labor, main season rainfall, household size and age-sex composition of the household (includes variables for the number of boys (aged <16 years), the number of girls, the number of men (aged 17-39), the number of women and the number of older men (aged >40 years)).

Table 8b: Joint Sector Comparative Advantage CRC Model OMD Structural Estimates

With Joint Sector Fertilizer-Hybrid Decision				
	Full Sample		Without HIV Districts	
	With Covariates	With Interactions with Hybrid	With Covariates	With Interactions with Hybrid
β	0.639 (0.095)	1.148 (0.813)	0.420 (0.051)	0.901 (0.175)
ϕ	-1.602 (1.684)	-3.133 (4.003)	-1.687 (0.554)	-2.051 (1.282)

Table 8c: Comparative Advantage CRC Model OMD Estimates: Both Hybrid and Fertilizer Endogenous

With Both Fertilizer and Hybrid as Endogenous				
Projection: $\theta_i = \lambda_0 + \lambda_1 h_{i1} + \lambda_2 h_{i2} + \lambda_3 h_{i1} h_{i2} + \lambda_4 h_{i1} f_{i1} + \lambda_5 h_{i2} f_{i1} + \lambda_6 h_{i1} h_{i2} f_{i1} + \lambda_7 h_{i1} f_{i2} + \lambda_8 h_{i2} f_{i2} + \lambda_9 h_{i1} h_{i2} f_{i2} + \lambda_{10} f_{i1} + \lambda_{11} f_{i2} + \nu_i$				
	Full Sample		Without HIV Districts	
	With Covariates	With Interactions with Hybrid	With Covariates	With Interactions with Hybrid
β	0.088 (0.096)	0.915 (0.417)	0.603 (0.060)	0.686 (0.174)
ϕ	-0.449 (0.176)	-3.772 (2.707)	-1.788 (0.277)	-2.118 (0.641)

Notes: Standard errors are in parentheses. OMD is optimal weighted (the weight matrix is the inverted reduced form variance-covariance matrix) minimum distance. Results from diagonally weighted (the weight matrix is the OMD weight matrix with the off diagonal elements set to zero) and equally weighted minimum distance are similar. The structural coefficients reported are average return to hybrid (β), the comparative advantage coefficient (ϕ). Results are reported for two samples: the full sample and the sample without two HIV districts and are for two periods of data. In addition, minimum distance results for two different specifications are reported: with covariates and with covariates and interactions of the covariates with the hybrid decision. All the specifications with covariates assume that all covariates other than hybrid and/or fertilizer are exogenous. Covariates include acreage, land preparation costs, fertilizer, hired labor, family labor, main season rainfall, household size and age-sex composition of the household (includes variables for the number of boys (aged <16 years), the number of girls, the number of men (aged 17-39), the number of women and the number of older men (aged >40 years)).

Table 9: Correlates of Estimated θ_i 's
Dependent Variable is the Estimated θ_i 's

	(1)	(2)	(3)	(4)	(5)
Distance to Closest Fertilizer Seller (x100)	-0.301 (0.122)	-0.289 (0.121)	-0.285 (0.121)	-0.285 (0.121)	-0.315 (0.063)
Distance to Motorable Road (x100)	-0.904 (0.503)	-0.887 (0.501)	-0.901 (0.502)	-0.898 (0.501)	-0.978 (0.285)
Distance to Matatu Stop (x100)	0.032 (0.298)	-0.034 (0.298)	-0.016 (0.298)	-0.028 (0.299)	-0.021 (0.165)
Distance to Extension Services (x100)	-0.130 (0.155)	-0.063 (0.155)	-0.063 (0.155)	-0.061 (0.155)	0.002 (0.091)
Tried to Get Credit (x10)	-	-0.138 (0.153)	-	-	-
Tried but Didn't Receive Credit (x10)	-	-	0.027 (0.347)	-	-
Received Credit (x10)	-	-	-	-0.047 (0.154)	-0.164 (0.144)
Dummies for Household Head Education (P value on Joint Significance)	No	Yes (0.002)	Yes (0.002)	Yes (0.000)	Yes (0.000)
Province Dummies	Yes	Yes	Yes	Yes	No

Notes: Standard errors are in parentheses. This table reports the correlations of the predicted θ_i 's with various observables. The predicted θ_i 's used are from the two period case with endogenous hybrid and fertilizer use (Table 8c). The observables are the average of the 1997 and 2004 values. Columns (1) through (4) show the results while controlling for province dummies. Since a lot of the variation in infrastructure may be aggregate, column (5) also shows the results without province dummies, but just for the "Received Credit" variable on the credit side (the results are no different for the other credit variables).