NERICAs: Leading the African Green Revolution?

A comparative econometric study of the barriers to adoption of New Rice Crops (NERICAs) in West Africa

Jonathan Phillips

15th September 2008

Abstract

If NERICAs are to fulfil expectations of catalysing a Green Revolution in Africa, understanding their potential for diffusion and the barriers to adoption is critical. Applying program evaluation techniques to two datasets for Côte d'Ivoire and Nigeria I derive estimates of current and long-run adoption rates that confirm NERICAs have potential coverage comparable to other innovations in historic Green Revolutions. Achieving this potential requires a better understanding of the distinct dynamics of awareness and adoption. By identifying the barriers to each I show that there are 'quick wins' for intervention that can boost NERICA adoption. More generally, adoption depends on both social learning and structural constraints. The analysis also clarifies a number of methodological issues and suggests non-parametric estimators may be more reliable than their parametric counterparts in studies of new technology adoption.

Contents

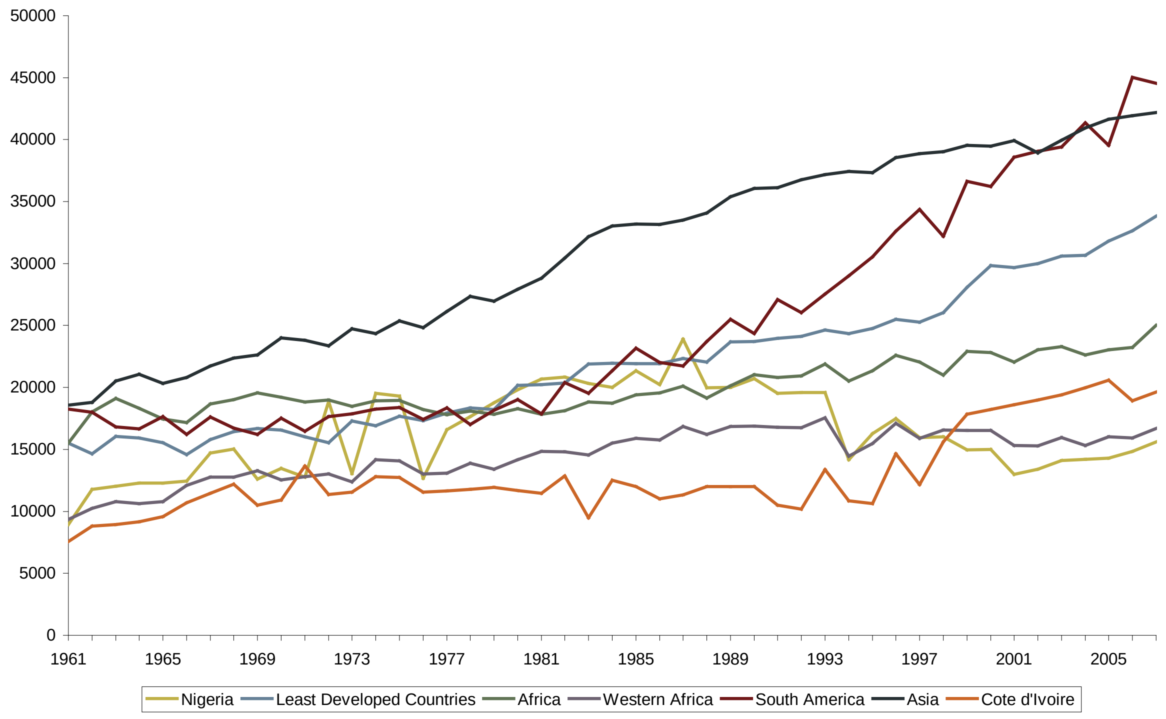
List of tables

List of figures

1 Introduction

The hardships created by low agricultural productivity are well recognised in West Africa: Low incomes, food insecurity, soil infertility and erosion, a lack of surplus for productive investment and even macroeconomic constraints such as the need to use scarce foreign exchange on food imports.[1] The solutions have been less apparent. The Green Revolution that occurred in Asia from the 1960s has proved elusive in Sub-Saharan Africa. Figure 1 shows that in rice productivity, Africa has performed worst out of the Least Developed Countries, and West Africa has been the laggard within Africa. Nigeria is no more productive now than it was in 1968 and Côte d'Ivoire has made limited progress only in the last decade.

Figure 1 Rice productivity by region



Source: FAOSTAT (2008)

[1] World Development Report (2008).

One promising innovation which has received both academic and political backing is the dissemination of high-yielding hybrid varieties of rice by the Africa Rice Centre (WARDA).[2] Boasting high yields, fast growth, more sustainable soil impacts, resistance to drought and disease, a high protein content, and marketable qualities, the range of NERICAs combine the resilience of African varieties with the potential of Asian varieties, making them well-adapted to the pressures of low-input agriculture and a diverse ecological environment. The optimism is conspicuous; at the 2006 Africa Rice Congress, NERICAs were identified as the vanguard of "technology development and dissemination", one of four pillars of an anticipated Green Revolution[3].

In this analysis, I take for granted the benefits that widespread NERICA adoption could bring.[4] Kijimi et al (2006) show that average yields could double. In Benin incomes rose by US$277 per hectare for men and US$337 per hectare for women.[5] There was a 6% increase in school attendance rates, a 14% increase in the gender parity index and a 36kcal per equivalent adult increase in calorie intake.[6] Mosley (2002) attributes these pro-poor properties to the scale-neutrality, labour-intensity, risk-reducing, price-reducing and off-farm linkage characteristic of hybrid seeds.

Of course, the magnitude of benefits depends on complementary improvements in soil management, local markets and government policies.[7] Yet the potential for households to benefit is clear. The focus of this paper is instead on *how many* households could benefit, i.e. on the breadth rather than depth of NERICAs' impact. More precisely, the focus is on whether NERICAs are likely to be adopted broadly enough to catalyse a Green Revolution in West Africa.

[2] Academic backing has come from both the scientific and development disciplines, see Africa Rice Congress (2006). Political backing has come from the New Partnership for African Development (NEPAD), see Nwanze et al. (undated).

[3] WARDA (2006). Also see Kaneda (2006), Scola (2007) and Africa Rice Congress (2006).

[4] The treatment effects I discuss are inevitably narrower in scope than a full cost-benefit analysis since they focus only on a single dimension of effects.

[5] Agboh-Noameshi et al (2006).

[6] Adekambi et al (2006).

[7] Goldman and Smith (1995) and Holmén (2003).

The counterpart to measuring the potential for adoption is to understand the strength of any barriers inhibiting adoption. To understand NERICAs potential I make use of two surveys carried out by WARDA in Côte d'Ivoire and Nigeria.

Section 2 cites the NERICA adoption question in the broader literature on innovation processes and the barriers to new technology adoption in agriculture. The dynamic, multi-stage and reversible nature of adoption is stressed and the contrast between social and structural theories of adoption is highlighted. Monty Jones, whose research laid the groundwork for NERICAs has linked adoption to key barriers: "If we don't develop the infrastructure, there's no way we'll attain the Green Revolution. How do you bring the NERICAs to farmers? How do you get farmers to know the seeds exist?"[8]

It is on this final information barrier of 'awareness' that the analysis centres. Being unaware of NERICAs is an absolute constraint to their adoption and introduces considerable bias into observed adoption rates. A failure to consider this constraint can – and has – led to a misunderstanding of new technologies' potential for adoption. Just as important, it can lead to a misdiagnosis of the other barriers to adoption. The source of the problem and econometric solutions grounded in the program evaluation literature are described in Section 3.

The results for both current adoption rates and significant barriers in Côte d'Ivoire and Nigeria are reported in Section 4. In Section 5 I use time-series techniques to predict the long-run potential adoption of NERICAs throughout the populations and make comparisons with countries that have experienced successful Green Revolutions.

In the conclusion, the contributions of this paper are summarised. In methodological terms, the paper resolves common conclusions, adapts econometric techniques to the new technology adoption context and stresses the limits of parametric techniques. The results

[8] New York Times, "In Africa, Prosperity from Seeds falls Short", http://www.nytimes.com/2007/10/10/world/africa/10rice.html?_r=2&hp&oref=slogin&oref=slogin

establish that the long-run potential of NERICAs is comparable to that of innovations in other Green Revolutions. By correcting adoption estimates for biases, strategic recommendations for WARDA that overcome the barriers to adoption can be made. These transcend the theoretical impasse between social and structural theories, highlighting the interaction in their dynamics and exploiting differences in the factors conducive to awareness and adoption. While the conclusions are country-specific, the approach and lessons learned will be critical if NERICAS are to achieve their potential throughout West Africa.

2 Innovation adoption processes

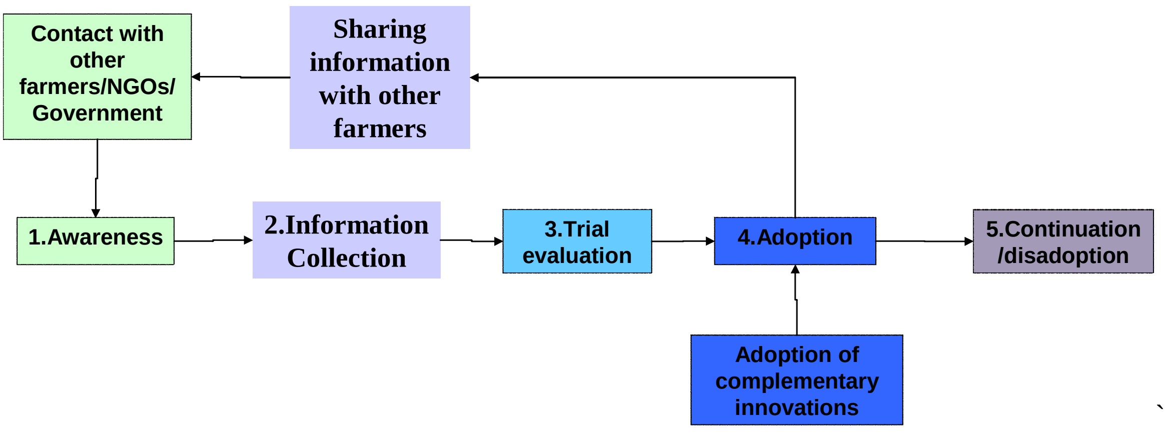
2.1 The centrality of information

Perspectives on adoption processes can be classified along a number of dichotomies:[9]

1. Economic structures and incentives vs. sociological influences
2. Static vs. dynamic adoption
3. Single stage vs. multi-stage adoption

In accommodating incomplete awareness of NERICAs, this study goes some way to bridging these dichotomies. The existence of incomplete information places emphasis on the role of social networks, communication and perceptions. In turn, this forces us to think of adoption as a dynamic process of "social learning" that occurs through the diffusion of information, building on previous experiments and anticipating long-run gains.[10] Finally, numerous studies have documented the complexity of the adoption process shown in Figure 2.

Figure 2 Stages of the adoption process[11]



Feder et al's (1985) comprehensive survey defines adoption as the decision to use an innovation in long-run equilibrium *given* full information about its potential.[12] Since then,

[9] Skinner and Staiger (2004), Dimara and Skuras (2004).
[10] Foster and Rosenzweig (1995).
[11] Adapted from Pannell et al. (2006).
[12] Feder et al (1985), p.256.

much emphasis has been placed on rolling back this informational assumption and in this study I confine the definition of adoption to the growing of one or more NERICA varieties.

Saha et al (1994) argue that the probability of adoption (and its intensity) is boosted by information because an increased understanding of potential benefits reduces outcome risk. Similarly, where there is considerable heterogeneity in the population, more of the information regarding the determinants of other farmers' success is likely to be *unobservable*. It therefore cannot be used by the farmer to judge the extent to which his own adoption would produce similar gains, slowing down the adoption process. The role of social learning is argued to be particularly important for the adoption of new varieties of rice by Munshi (2004), who compares diffusion patterns with wheat and concludes that greater heterogeneity in, and sensitivity to, growing conditions for rice make the determinants of successful adoption unobservable and usable information scarce.

Possessing information is not, then, a binary variable, and its acquisition takes many forms. For instance, Dimara and Skuras (2004) distinguish passive learning from active information gathering, where farmers search out information on an innovation.[13] However, in this analysis I focus on the initial binary stage – awareness, defined as a positive response to the survey question "have you heard of NERICAs?" – because the innovation is a nascent one (awareness is around 10% in the Côte d'Ivoire sample) and awareness is likely to be the most immediate hurdle to adoption. Nevertheless, the complexities of heterogeneity and unobservability will remain central.

2.2 Hypotheses on the barriers to adoption

In the latter stages of Figure 2, the adoption of NERICAs should be seen simultaneously as an innovation process (involving learning and adaptation to local conditions), a production decision (reflecting potential profitability and risk) and a consumption decision (for

[13] The distinction was also made by Feder and Slade (1984).

households which consume part of their crop).[14] It is therefore unsurprising that in practice NERICAs have been used to complement, rather than substitute for, existing varieties.[15]

Further, adoption rarely occurs in isolation. Instead, farmers engage in a "multistage procedure of organisational change" to optimise their farming system in response to new technologies.[16] Given the properties of NERICAs, and in particular their complementarity with fertilisers, a number of key choice variables are likely to be simultaneously determined.[17] Goldman and Smith (1995) provide evidence for Northern Nigeria (and India) that any Green Revolution relies on coincident changes in fertilizers, power sources, land use patterns and labour market adaptation.

These considerations emphasize the wide range of factors that may act as barriers inhibiting adoption. Doss (2006) urges micro-level studies to highlight policy-relevant variables, facilitating targeted interventions.

— Kijimi et al (2006) stress the importance of training and extension services to aid local adaptation. The authors also note that prior rice-growing experience generates good management practices which increase the benefits from NERICA adoption.

— While seeds are not a particularly lumpy investment,[18] credit and cash constraints can be important because adoption often entails complementary investments. Ideally, of course, it is the *availability* of credit and cashflow that is relevant. Our data forces us to settle for the use of credit in the current period.

— Doss (2006) notes the ability to hire additional labour may be a prerequisite to adoption. In product markets, expected crop prices and price-volatility are just as central. The

[14] Pannell et al. (2006) stresses the role that 'trialability' plays in adoption – divisible technologies such as seed varieties are likely to be more amenable to adoption than entire farming systems which cannot be trialed.

[15] WARDA (2008), Compendium, pp.123-124.

[16] Dimara and Skuras (2004).

[17] Indeed, an entire module of the NERICA Compendium is dedicated to appropriate changes in farm management. WARDA (2008), Module 6.

availability of complementary inputs such as fertiliser, and of seeds themselves may also inhibit adoption.

— Conley and Udry (2003) show that farmers adjust their activities in line with the successful experimentation of others, so social networks are important to adoption. Bandiera and Rasul (2002) argue that while there are positive spillovers from sharing best practices, there may also be free-riding in experimentation.

— Asfaw and Admassie (2003) note that human capital improves “the allocative and technical efficiency of producers by exposing them to a more systematic and dynamic production system and enhances their ability to choose the optimal bundle of input and output mix”.[19] Indeed, the belief that a lack of education can be a barrier to adoption is so widespread that it has been termed the “Schultz hypothesis”.[20]

— Reviewing the evidence on farm size, Feder et al (1985) conclude that there is a positive relationship with adoption and that this is likely to be driven by the existence of high transaction costs, including information costs and access to credit.[21] An alternative explanation is provided by Feder and Slade (1984) who argue that only large-scale farmers have the resources to invest in information accumulation.

— Mosley (1982) highlights the many sources of technological, financial and policy risk that farmers face, and the lack of formal and informal institutions in rural areas to insure against these risks. Risk and risk-preferences are not captured in our data.

— Goldstein and Udry (2005) show farm tenure arrangements are connected to investment in new techniques, but no such information has been recorded in the surveys.

[18] Feder et al (1985), p.263, suggest that high-yielding varieties tend to be scale-neutral.
[19] Asfaw and Admassie (2003), p.216.
[20] Ibid., p.217.
[21] Feder et al (1985), p.261 and p.272.

— Finally, individual plots may simply not be suited to NERICAs. Unfortunately, accommodating plot-level variability is not possible with our datasets.

Table 1 relates these hypotheses to the variables in our datasets.

Table 1 Barriers to adoption and associated variables

Constraints	Côte d'Ivoire	Nigeria
Credit/cash availability	Log of total agricultural cash income in 1999	Use of credit from any source
Labour markets	Household size in 2000	Number in household
Product markets		Minimum distance to market
Extension support for adaptation	i) Contact with extension agencies; ii) Participation in PVS trials before 2000; iii) Participation in on-farm trials before 2000.	i) Attended field-day demonstration ii) Varietal trials on household land
Social networks		i) Membership of an organisation ii) Information from other farmers iii) Information from radio/TV
Education	Years of formal schooling	Highest formal school
Farm size	Log of maximum farm area cultivated in past five years	Total size of all plots
Input availability		i) Fertilizer and herbicide use ii) Village electricity iii) Bicycle/Motorcycle/Car ownership
Risk preferences	-	-
Farm tenure	-	-
Plot-specific conditions	-	-

In summary, this framework allows us to emphasize two sets of dynamics. First, by *incorporating* incomplete information it allows us to discover the importance of information to innovation diffusion. Second, by *controlling* for information differences it allows us to gain insight into the relative importance of other barriers. These reflect two different policy focuses and perceptions of how innovation diffusion occurs. The former focuses on dynamic aspects of social learning and has implications for civil society, extension services and Government's spatial policy. The latter focuses on the structural constraints to adoption and has implications for market and institutional reforms. This paper seeks to show that combining the approaches in empirical work makes the results of each more informative.

3 Methodology: Estimating adoption rates

There are two figures that emerge naturally from any study of technology adoption: the full-sample adoption rate,[22] and the adoption rate in the ‘exposed’ sub-sample that is aware of the technology.[23] With incomplete exposure, the former is smaller than the latter. Neither are robust estimates because each is afflicted by a particular bias.

3.1.1 Non-Exposure Bias

In their report on the Nigerian survey, Spencer et al (2006) conclude that:

> “Considerable uptake of NERICA1 is observed (30% of farmers in Ekiti are estimated to have cultivated NERICA1 in 2005, and 42% and 19% respectively in PVS and near-PVS villages in Kaduna). These uptake rates are higher than have reported in other West Africa studies.”

These full sample adoption rates are misleading and unsuitable for comparison *despite* the sophisticated random stratified sampling technique used in the survey.

To understand why, recall that adoption is not possible without awareness of NERICAs. This ‘treatment’ is required before the farmer’s response (their decision to adopt or not) can be properly observed. Untreated agents *cannot* adopt, even if they would have chosen to on being made aware of NERICAs; the observed adoption response for these individuals is censored and the full-sample adoption rate is biased downwards. For binary adoption decisions, “zeros may be generated by at least two distinct processes”;[24] one reflecting a

22 $\frac{N^a}{N}$, where N^a is the number of adopters and N is the sample size.

23 $\frac{N^a}{N^e}$, where N^e is the number exposed.

24 Dimara and Skuras (2004), p.188.

decision not to exercise the option and the other arising from a lack of awareness that proscribes the *opportunity* for adoption.

For example, in the Côte d'Ivoire survey, only 11% of the sample had knowledge of NERICAs. The maximum possible full sample adoption rate would have been 11% (it was in fact 4%). Comparison with the 30% for Ekiti (based on 66% exposure) is clearly misleading. So the adoption rate in the full sample is likely to be an *under*estimate wherever exposure is incomplete.

3.1.2 Selection Bias

An obvious solution to non-exposure bias is to examine adoption rates among the exposed sub-sample. For Côte d'Ivoire, this amounts to 38% and for the Nigerian states, 48%.

While more reasonable than the previous comparison, these estimates implicitly assume that the exposed sub-sample is representative of the population. In practice, *selection* biases affect who is aware of new technologies. For NERICAs, awareness among farmers has been spread by word-of-mouth and through "informal channels",[25] which are unlikely to be randomized.

One mechanism through which self-selection could operate is that farmers with the most to gain from adoption actively search for information about such technologies. Alternatively, farmers more closely connected to farming institutions and social networks may be most likely to be exposed to information, and these farmers may also happen to be on average the most predisposed to adoption.

A quite different mechanism arises where information is actively diffused by WARDA and other agencies. Often, and reflecting the competitive research and funding environment in which such institutions operate, there are strong imperatives to either focus on the poorest segment of the population, or on the most progressive, in order to demonstrate the maximum impact of interventions.

Note that selection bias is not selection *into* the survey sample but *within* the sample, in the complex and informal process by which awareness of NERICAs diffuses. Since the treated group is then unrepresentative, their disposition to adopt cannot be extrapolated to the wider population, and in particular the untreated group.

The dilemma that these two biases pose in combination is that while using the full sample introduces non-exposure bias that leads to underestimation, restricting our sample to the treated introduces selection bias, which may generate under- or overestimation.[26]

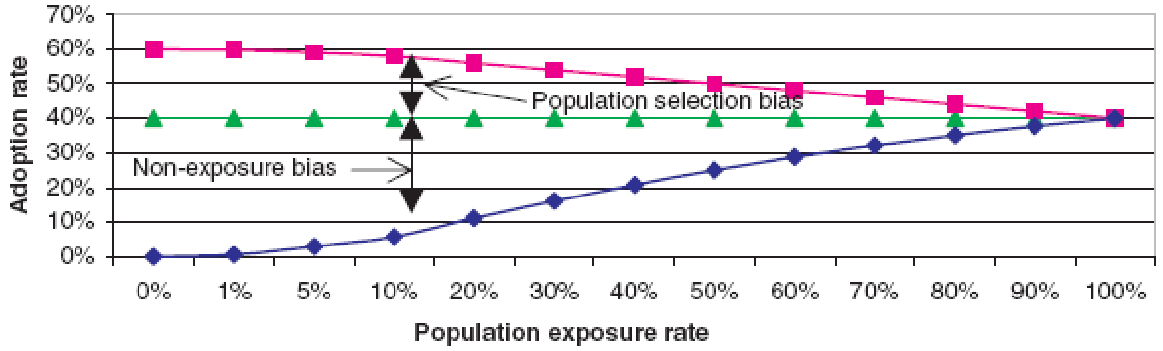
In fact, Diagne (2006) shows that the typical full sample adoption rate, such as that calculated in Spencer et al (2006), is actually a *joint* estimate of the likelihood of exposure and subsequent adoption. Figure 3 summarises the two biases (in the case of positive selection bias) and shows that their magnitude is inversely related to the level of exposure. As the whole sample is made aware of NERICAs, selection no longer occurs and everyone who might want to adopt is able to.[27]

[25] Diagne 2006, p.209. That ‘treatment’ occurs naturally in the population and is not a controlled application of treatment by a third party is, if anything, likely to accentuate self-selection bias since there are no institutional controls that might seek to randomize receipt of treatment.

[26] Overestimation is likely in the situation described, but there is no reason that selection bias could also lead to underestimation if farmers less likely to know about NERICAs are more likely to adopt once they become aware.

[27] Thus, non-exposure bias and selection bias should be seen as two sides of the same coin: non-exposure only matters if selection biases exist and selection biases depend on incomplete exposure.

Figure 3 Potential biases in surveys



Source: WARDA (2003), Annual report, p.39.

The same problem afflicts the identification of barriers to adoption, since the circumstances faced by the exposed non-adopters may not be representative. The basic probit model adopted by Spencer et al. (2006) is unreliable because of its failure to control for awareness. Again, Diagne (2006) stresses that such estimates identify the *joint* barriers to exposure and adoption.

3.2 A counterfactuals framework

These biases arise from observing only exposed farmers' adoption responses. A counterfactual framework allows us to examine all possible responses.[28] Formally, for each individual in the sample:

d_i the binary treatment variable takes the value 1 with treatment and 0 in its absence;

y_{0i} the binary outcome variable (adoption) in the absence of treatment;

y_{1i} the outcome variable in the presence of treatment;

x_i characteristics influencing the likelihood of adoption;

z_i other characteristics influencing the likelihood of treatment and including x_i.

e_i unobservable characteristics influencing the likelihood of adoption.

[28] Referred to as the 'Rubin Causal Model' in statistics.

Further, we separate the outcome variable into an average component, u_1 and u_0, and an individual-specific component, v_1 and v_0:

$$y_{0i} = u_0 + v_0$$

$$y_{1i} = u_1 + v_1$$

Ideally, $(y_{1i} - y_{0i})$ would provide the impact of treatment on each individual. Indeed, while information about y_i only allows us to establish an *associative relationship* between treatment and response, y_{1i} and y_{0i} allow us to infer a *causal relationship* based on the counterfactual.[29] However, in a cross-sectional sample, y_{1i} and y_{0i} are mutually exclusive states – each person is *either* treated or untreated – so the counterfactual is unobservable.[30] The observed outcome y_i is a function of both potential outcomes and treatment status:

$$y_i = d_i \cdot y_{1i} + (1 - d_i) y_{0i} = y_{0i} + d_i (y_{1i} - y_{0i})$$

An interesting characteristic of new technology adoption is that $y_{0i}=0$ for all individuals, reflecting the impossibility of adoption without knowledge of the innovation. Yet, the problem remains since we do indeed observe $y_{0i}=0$ for the untreated, and cannot therefore observe y_{1i} for all individuals.

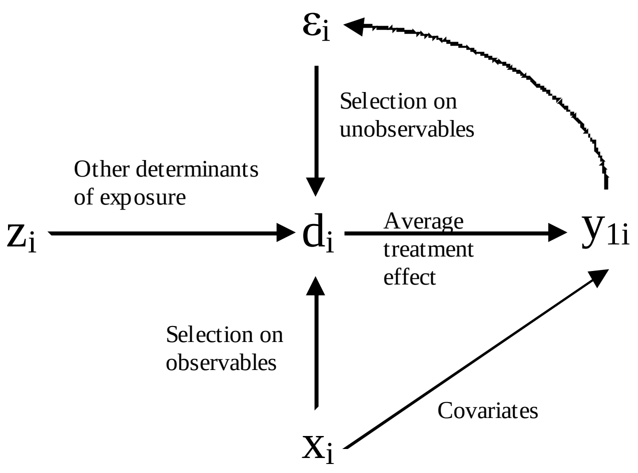
3.3 Calculating average treatment effects

Given the sample data available, the next best course of action is to obtain an estimate for the *average* impact of treatment by comparison between groups. However, in making comparisons *across* individuals we must be particularly careful that we are comparing like

[29] Lee (2005), p.9.
[30] Heckman and Vytlacil (2001), p.3. Specifically, we only observe y_{1i} when $d_i=1$ and y_{0i} when $d_i=0$.

with like and not merely capturing differences in *xi*, *zi* or ε_i between individuals.[31] Figure 4 illustrates the complex causation possibilities and the challenge of inference.

Figure 4 Causation diagram



3.3.1 Parameters of interest

When averaging across individuals, the question immediately arises as to "which average?" The candidates are shown in Table 2.

[31] Lee (2005), p.21. Even in time-series analysis, assessing the change from y_{0i} to y_{1i} over time as the agent is treated is difficult because other factors influencing the outcome variable are likely to change over time and may be difficult to control for.

Table 2 Potential measures of the impact of treatment[32]

	Measure	Description	Positives	Negatives
Average Treatment Effect (ATE)	$E(y_1 - y_0)=u_1-u_0$ $=u_1$	The ATE answers the question, on average, what is the effect on the probability of adoption of treating a random individual selected from the population?	Represents potential adoption rate under full exposure	Includes agents who may be impossible to treat
Average Treatment Effect on the Treated (ATT)	$E(y_1 - y_0 \vert d=1)$ $= u_1-u_0 + E(v_1-v_0 \vert d=1)$ $= u_1+E(v_1 \vert d=1)$	Answers the question, on average, what was the probability of adoption among those exposed?	Intuitive as the actual adoption rate among the sample that were able to adopt	Embodies selection biases
Average Treatment Effect on the Untreated (ATU)	$E(y_1 - y_0 \vert d=0)$ $= u_1-u_0 + E(v_1-v_0 \vert d=0)$ $= u_1-E(v_1 \vert d=0)$	Answers the question, on average, what proportion of those who have not yet been exposed would adopt if knowledge were extended to them?	Represents the potential for further adoption	Embodies selection biases

Note that the ATE is simply the weighted average of the ATT and ATU, where the weights reflect the probability of exposure to treatment.

$$ATE = ATT \times P(d=1) + ATU \times (1 - P(d=1))$$

There are three circumstances where ATE is equal to ATT and ATU. The first is full exposure. The second is where the impact of treatment is homogenous, regardless of individual characteristics.[33] The third is where treatment assignment is independent of outcomes, for example under randomization, so that selection bias is inconsequential; those that select into treatment are no more or less likely to adopt. Algebraically, if $d \coprod (y_{1i}, y_{0i})$,[34] the adoption rate among the exposed is a valid estimator:

$$E(y \vert d=1) - E(y \vert d=1) = E(y_1 \vert d=1) - E(y_0 \vert d=0) = E(y_1) - E(y_0) = E(y_1)$$

In general, and in the case of NERICAs, these conditions do not hold. In addition to non-random treatment, one challenge in developing NERICAs has been the high sensitivity of

[32] The measures here make use of the fact that $y_0=0$.
[33] Algebraically, $E(v_1-v_0|d=1)=0$ so the final term in the ATT equation in Table 1.1 drops out and ATE=ATT.
[34] $\coprod$ means independence throughout.

yields to local conditions in West Africa. Thus, we anticipate that the parameters will differ, and which we are interested in depends on the empirical question of interest.

Arguably it is the ATE which is of most interest since this allows us to calculate the potential adoption rate across the whole population. It is plausible that awareness could become near-universal, particularly if supported by interventions.[35]

3.4 The failure of ordinary estimation techniques

Saha et al (1994) propose the most explicit way to represent selection bias is by defining separate structural equations for treatment and adoption.[36]

$$\text{(i)}\quad y_i = \alpha_0 + \alpha_d d_i + \alpha_x x_i + e_i$$

$$\text{(ii)}\quad d_i = \beta_0 + \beta_z z_i + \mu_i$$

The reduced form equation is:

$$y_i = \gamma_0 + \alpha_d \beta_z d_i + \alpha_x x_i + \alpha_d z_i + \varphi_i$$

$$\text{where } \gamma_0 = \alpha_0 + \alpha_d \beta_0 \text{ and } \varphi_i = e_i + \mu_i$$

The implication of selection bias is that the error terms e_i and μ_i are correlated by a measure ρ as a result of d_i occurring in (i). A probit or logit model of (i), as carried out by Spencer et al. (2006), may give unreliable results because there is correlation between d_i and e_i, preventing the separate influence of each on *y* being assessed.

[35] The attention to the ATE rather than the ATT is unusual for this reason; with most programs the feasible policy question is of moderate extension or contraction of treatment, while there are few limits to potential growth in awareness. If it emerged that boosting awareness were actually a very expensive and slow process, greater attention may need to be paid to alternative parameters. Imbens and Wooldridge (2008), p.11.

[36] Saha et al (1994) also suggest that a third equation, defining the *intensity* of adoption for those who have chosen to adopt is also appropriate. This reflects the further selection bias and truncation that occurs because we only observe the intensity that adopters select and not the intensity that non-adopters would select, if they chose to adopt. See Section 4 an application of this approach.

If the error terms are normally distributed, then the expectation of a truncated bivariate normal distribution shows that:[37]

$$E(y=1|d=1)=\Phi(\alpha_x'x)+\frac{\rho\phi(\beta_z'z)}{1-\Phi(\beta_z'z)}$$

where Φ is the cumulative density function and ϕ the probability density function of the normal distribution. The standard probit model estimates $E(y=1)=\Phi(\alpha_x'x)$[38] and therefore fails to evaluate the second term which represents the selection bias. The ATE predictions and estimates of α_x – the determinants of adoption– would be biased.

3.5 Econometric solutions

3.5.1 Conditional Independence

Individual-specific outcomes can be separated into observable ($g_i(x)$) and unobservable (e_i) components:

$$v_0=g_0(x)+e_0 \quad E(e_0|x,z)$$

$$v_1=g_1(x)+e_1 \quad E(e_1|x,z)$$

Controlling for 'selection on observables' simply requires us to compare treated and untreated individuals *with the same x*. If this eliminates all the selection bias then, *conditional on x*, treatment is independent of outcomes, a situation which is variously termed "conditional independence", "ignorability of treatment", "selection on observables" or "unconfoundedness".[39]

[37] Saha et al (1994), p.840.
[38] Greene, (1993), *Econometric Analysis*, p.637.
[39] Rosenbaum and Rubin (1983), Diagne and Demont (2007), Barnow, Cain and Goldberger (1980), Imbens (2004). An extra condition is the 'common support' requirement to ensure that there are values of *x* shared by both the treated and untreated groups: $0<P(d=1|x)<1$.

$$d \amalg (y_{1i}, y_{0i}) | x_i$$ [40]

In this case we can attribute any systematic differences in outcomes between the treated and untreated groups to known variables.[41] To calculate treatment effects, the unknown counterfactual for the untreated group can be inferred from the known outcomes for the treated group (Table 2).[42] This approach underlies many of the estimators in the following section.

Table 3 Counterfactuals and Conditional Independence[43]

	Known outcomes	Unknown counterfactuals	Inference under conditional independence
Treated	$E(y_1 \| d=1, x)$	$E(y_0 \| d=1, x)$	$E(y_0 \| d=1, x) = E(y_0 \| d=0, x)$
Untreated	$E(y_0 \| d=0, x)$	$E(y_1 \| d=0, x)$	$E(y_1 \| d=0, x) = E(y_1 \| d=1, x)$

3.5.2 Additional Assumptions

An additional requirement that we make throughout is the Stable Unit Treatment Value Assumption (SUTVA).[44] This requires that treatment of one individual does not affect the outcome for others.[45] This is a simplification since Holloway et al (2002) identify positive neighbourhood effects from the adoption of high-yielding rice in Bangladesh.

A priority of our analysis is allowing for heterogeneous effects across individuals, $(v_1 \neq 0)$.[46] A key challenge for the transfer of the Green Revolution to Africa has been the diversity of

[40] Wooldridge (2002, p.607) shows that the weaker assumption of conditional *mean* independence is actually sufficient for the identification of average treatment effects: $E(y_0|x,d) = E(y_0|x)$, $E(y_1|x,d) = E(y_1|x)$

[41] An assumption commonly paired with conditional independence is that there is overlap in the values of the covariates for the treated and untreated groups, facilitating comparison.

[42] In the new technology adoption context where $y_0=0$, only the second inference for the untreated is required.

[43] Note, since $y_0=0$, only the second inference is used in this paper.

[44] Rubin (1978).

[45] Technically, observations are independently and identically distributed.

[46] Imbens and Wooldridge (2008) emphasize the accommodation of heterogeneity as a central contribution of the modern literature.[46]

ecological and economic conditions faced by farmers and it is likely that behavioural responses to treatment will vary greatly.

Finally, the following derivations make use of the fact that $y_0=0$ to simplify the estimators. Throughout, I define x as the set of covariates determining adoption and x^{ex} as the subset of x which is *exogenous* in the structural equation for adoption. z represents the determinants of exposure to treatment and contains x^{ex} as a subset. z^{IV} represents the members of z which are not in x^{ex}.

3.5.3 Alternative Estimators

Table 4 Alternative estimators

		ATE Estimator	Additional Assumptions	Source
Assuming conditional independence and overlap				
A	**Semi-parametric Estimator 1**	$A\hat{T}E = \frac{1}{n}\sum_{i=1}^{n}\left[E(y\vert x, d=1)\right]$		Wooldridge (2002), p.609.
B	**Semi-parametric estimator 2**	$A\hat{T}E = \frac{1}{N}\sum_{i=1}^{N}(\frac{y_i}{\hat{p}(z)})$	$P(d=1\vert x,z) = P(d=1\vert z)$ $P(y_1=1\vert x,z) = P(y_1\vert x)$ $p(z) > 0$	Diagne and Demont (2007), p.204.
C	**Parametric Estimator**	$E(y\vert d,x) = \alpha d + \beta_x d\left[x - \overline{x}\right]$	$E(v_i\vert x)$ is a linear function of *x*.	Wooldridge (2002), p.613.
D	**Propensity Score (parametric)**	$E(y\vert d,x) = \alpha d + \beta_i d\left[p(z) - \overline{p}(z)\right]$	$0 < p(z) < 1$ $E(y_0\vert p(z))$ and $E(y_1\vert p(z))$ are linear in *p(z)*	Wooldridge (2002), p.618.
E	**Propensity Score (semi-parametric)**	$A\hat{T}E = \frac{1}{n}\sum_{i=1}^{n}\left[E(y\vert p(x), d=1)\right]$	$0 < p(z) < 1$	Lee (2005), p.94.
Not assuming conditional independence				
F	**Instrumental Variables**	$y_i = \alpha d_i + \beta_x d(x - \overline{x}) + de_1$	Instrument relevance: $L(d\vert z,x\rangle \neq L(d\vert x)$ Instrument validity: $L(v_0\vert x,z) = L(v_0\vert x)$ Individual gains are linear in *x* $v_i = g_i(x) + e_i$ Mean of the error term is independent of *x* and *z*: $E\left[d(e_1)\vert x,z\right] = E\left[d(e_1)\right]$	Wooldridge (2002), p.626.

A full derivation of each estimator is provided in Appendix 1. However, a number of key features are worth highlighting. Models (A)-(E) use conditional independence to derive consistent estimators by deploying the outcome for the treated group in place of the hypothetical outcome y_1 once account has been taken of differences in x:[47]

$$ATE(x) = E(y_1 - y_0 | x) = E(y_1 | x) - E(y_0 | x)$$

$$= E(y_1 | d = 1, x) - E(y_0 | d = 0, x)$$

$$= E(y | d = 1, x))$$

$$ATE = E(ATE(x))$$

(A) is intuitive, taking the outcome for the treated at particular values of x and averaging across the full sample. The counterfactual outcome for the untreated is implicitly the same as for the treated *once* we have removed any differences in *x* between the groups.

(B) is derived in a similar way to (A), but makes use of our definition of X as a subset of Z to define the propensity score, $p(z) = \Pr(d = 1 | z)$, for receiving treatment. This can be estimated with a flexible logit or probit model.

(C) includes an interaction term between treatment and the covariates to take account of heterogeneity in responses to treatment. The ATE is estimated by the parameter α in this and subsequent estimators. Using $y_0=0$ allows us to exclude a constant and linear covariates term because there is no variation in the pre-treatment outcome; selection mechanisms work on the basis of post-treatment outcomes. However, a difficulty posed by $y_0=0$ is that a cell of the permutation matrix is empty (where d=0 and y=1). This 'quasicomplete separation'[48] of the treatment variable prevents probit estimation and requires us to use a linear probability model as recommended by Caudill (1998).

[47] The second line follows by conditional independence, the third by $y_0=0$ and the fourth is possible using the overlap assumption. Note that conditional independence allows us to write the ATE in terms of *y*, which is observable, rather than y_1, which is not.

[48] Zorn (2005).

(D) is a uni-dimensional matching technique following Rosenbaum and Rubin (1983). Conceptually, individuals with the same probability of treatment – measured by the propensity score – but whose treatment status differs provide an unbiased comparison for estimating the ATE. From the counterfactual perspective, untreated individuals are assigned an outcome based on individuals in the treated group with the same propensity score.

(E) is the semi-parametric equivalent of (D). It is estimated using the command **psmatch2** in Stata.

(F) is the only estimator not to assume conditional independence and make allowance for our inability to control for all determinants of treatment. Instead, instrumental variables provide exogenous variation in treatment status, mimicking random assignment and permitting comparison between the treated and untreated.[49] However, the IV estimator imposes strict assumptions:

i. Instruments must be relevant,[50] so that *z* affects the probability of treatment. $L(d|z,x) \neq L(d|x)$.[51] Weak instruments may give less efficient estimates than a standard OLS.

ii. Instruments must be valid, so that *z* only impacts the outcome variable *y* through treatment and not directly (see Figure 4).[52] $L(v_0|x,z) = L(v_0|x)$.

iii. $E[d(e_1)|x,z] = E[d(e_1)]$ so that the mean of the error term does not depend on *x* and *z*, allowing them to be used as instruments. Since we do not assume $e_1=0$, we are permitting heterogeneity in the *effects* of treatment on the basis of unobservables. Unfortunately, the assumption requires that unobservables are uncorrelated with treatment, and therefore cannot be the basis on which *selection* operates. Heckman

[49] Blundell and Costa Dias (2002).
[50] Following Lee (2005), p.130. This is the inclusion restriction.
[51] L(x) represents the linear projection of x.

and Vytlacil (2005) clarify the distinction to show that while “heterogeneity in responses” is permitted, “heterogeneity in choices” is not.[53] I argue in Appendix 1 that for the low threshold of ‘awareness’, ‘heterogeneity in choices’ is less prominent than in other settings because farmers cannot align their information-acquisition with the individual-specific gains *before* they are aware of NERICAs’ specific properties. This may be one of the more plausible settings where IV estimators can be deployed.

3.6 Sources of confusion

One source of confusion in the literature is the distinction between different selection mechanisms and levels of heterogeneity in the sample. Table 4 shows how the two are connected and which models they are linked to. Selection only introduces bias where it is undertaken on the basis of outcomes, and therefore heterogeneity is a requisite for selection. Imbens and Angrist (1994) show that in the most general case of ‘active selection on unobservables’ the ATE is not identified and the Local ATE (defined only for agents who move from the untreated to treated group) is all that can be estimated.

[52] This is the exclusion restriction and states that z does not belong in the structural equation for y.
[53] Heckman and Vytlacil (2005), p.669.

Table 5 Heterogeneity and Selection Mechanisms

Adoption Outcomes		Selection Mechanisms	Estimator
Homogeneous effects: $E(y_1\text{-}y_0)=\alpha$	$\longrightarrow$	Randomized treatment: $d \coprod (y_{1i}, y_{0i})$	Difference in sample means between treated and untreated
Heterogeneity in pre-treatment outcome: $y_0 \neq 0$	$\longrightarrow$	Selection on initial status: $d \coprod (y_{0i} \vert x_i)$	(A)-(E) including a constant and covariate vector in the regressions for (C) and (D)
Homogeneity conditional on observables: $E(y_1\text{-}y_0 \vert x)=\alpha$	$\longrightarrow$	Selection on Observables : $d \coprod (y_{1i}, y_{0i}) \vert x_i$	(A)-(E)
Heterogeneity in responses (on unobservables): $E(y_1\text{-}y_0 \vert x,e)=\alpha$, $E[d(e_1) \vert x, z] = E[d(e_1)]$	$\longrightarrow$	Passive selection on Unobservables: $d \coprod (y_{1i}, y_{0i}) \vert x_i, e_i$ and $E[d(e_1) \vert x, z] = E[d(e_1)]$	(F)
Heterogeneity in choices (on unobservables): $E[d(e_1) \vert x, z] \neq E[d(e_1)]$	$\longrightarrow$	Active selection on Unobservables: $d \coprod (y_{1i}, y_{0i}) \vert x_i, e_i$	ATE not identified.

A second source of confusion involves the circumstances under which the *x* and *z* variables should be used in the multiple stages of the various models.[54] Diange and Demont (2007) include all the determinants of treatment, *z*, in the first stage exposure model. Yet, selection bias only arises where it affects adoption outcomes, so achieving 'balance' between the characteristics of the treated and untreated groups only requires controlling for x^{ex}. Note endogenous regressors are omitted to ensure a valid model. Yet, if the instruments z^{IV} are valid, the only cost to their inclusion is in the degrees of freedom. In this paper I nevertheless include *z* in the first-stage model for the following reason: The role of z^{IV} in (F) is to produce exogenous variation in treatment that helps balance *unobservables* between the two groups. To the extent that we are rarely certain whether selection on observables (conditional independence) holds, the inclusion of z^{IV} can only help improve the balance between the treated and untreated for the non-IV-based methods as well.

[54] For instance, in Diagne and Demont (2007) conditional independence is stated in terms of *z*, not *x*, which is an unnecessarily strict assumption.

For second-stage estimates of *adoption*, Heckman and Vytlacil (2005) show that the inclusion of endogenous variables and the full vector of *x* does not prohibit inference.[55] While endogenous variables correlated with error terms are permitted, a replacement assumption requires that the covariates would have been the same in the untreated counterfactual. In general, Lee (2005) demonstrates that this covers only variables determined pre-treatment. Our data includes some variables determined after treatment,[56] and that may indeed respond to treatment. These include fertiliser use, area devoted to rice, and practicing upland rice cultivation. These variables are included because they are of considerable interest. The cost is that our results must be interpreted as the effect of treatment *net* of any change in covariates that are induced by treatment.[57]

3.7 Application to the WARDA Surveys

3.7.1 Description of surveys, sampling and data

The population of interest to our study is rice farmers.[58] The surveys exhibit considerable overlap on socioeconomic, treatment and outcome variables. However, in Côte d'Ivoire extensive village level data exists on issues such as crop knowledge. In Nigeria, there is much greater detail about farmers' assets and techniques, such as tractor ownership and fertiliser use.

The sampling technique was similar between studies and suggests a non-random sample. In Côte d'Ivoire, village-level selection ensured "key sites" where WARDA had been carrying out extension activities were included.[59] These sites constituted 32 of the 50 villages in the survey. 30 rice farmers were randomly selected in each village to give a sample of 1,500. Only 11% of the full sample had knowledge of NERICAs.

[55] Heckman and Vytlacil, (2005), p.677.
[56] The precise date when individuals became aware of NERICAs is unknown.
[57] Lee (2005), p.46.

In Nigeria, Ekiti and Kaduna were selected based on WARDA's extension activities. Six villages where participatory varietal selection (PVS) had been carried out were randomly selected and for each Local Government Area in which they were located a non-PVS (but upland rice-growing) village was randomly selected. This first stage of stratification was complemented by a second stage stratification between males (75%) and females (25%). Eight upland rice farmers were selected in each village to give a sample size of 192.[60]

A useful way to gain an understanding of the data is to examine the difference in covariate means for the treated and untreated groups. The significance test we use in Table 5 is a two-sample t-test where the samples have unequal variances. The larger the difference between groups, the more biased is the adoption rate among the exposed relative to the ATE.[61]

Table 6 Testing for differences between treated and untreated groups[62]

Mean difference (treated group minus untreated group)			
Variable	**Côte d'Ivoire**	**Variable**	**Nigeria**
Number of traditional varieties known	3.52***	Total size of all plots (no year defined, not logged)	1.45
Number of NARS upland varieties known	-0.06	Max rice area share in 2005 (%)	0.14**
Number of Upland WARDA intraspecific varieties	0.67***	Rice sale 2004 % (not binary)	1.36
Log of farm size (maximum total area cultivated in the past 5 years)	0.13**	Number in household	-0.14
Idle (uncultivated) area share in 1999	0.02	Age	2.08
Rice area share in 1999	-0.04**	Equals unity if number of years lived in village is equal to age	
Commercialized crops (including rice) area share in 1999		Years lived in village	2.43
Log of total agricultural cash income in 1999	0.28**	Dummy that's unity where earns >10% income in paid full-time employment OR casual employment	
Income from rice in 1999 (thousands CFA)	7.24*	Highest formal school	0.52
Grow rice partially for sale in 2000	0.02	Practiced upland rice cultivation in 2003 or 2004	-0.02

[58] While it is possible that other farmers could be induced to switch to rice farming by NERICAs, at their current stage of diffusion the priority is to assess the potential among existing rice farmers.
[59] Diagne (2006), p.212.
[60] Spencer et al. (2006, p.24) raise concerns about the reliability of village participation in PVS in Ekiti.
[61] Imbens and Wooldridge (2008), p.24.
[62] *=significant at 10%, **=at 5%, ***=at 1%.

Household Size	-0.58	Varietal trials on household land (ever)	0.12
Age	0.28	Attended field-day demonstration (ever)	0.20**
Was born in the same village	-0.03	Sex	-0.13*
Years of residence in the village	0.42	Yoruba	-0.15*
Has a secondary occupation	0.01	Hausa	0.02
Years of formal schooling	1.31***	Kadara	0.05
Practice upland rice cultivation	0.24***	Kind of school attended (koranic vs. formal)	0.36**
Participation in PVS trials before 2000	0.1**	Agricultural Development Projects zone	0.08
Participation in on-farm trials before 2000	0.16***	Owns bicycle	0.15*
Female	0.04	Owns motorcycle	0.01
Bete	0.09**	Owns car	0.04
Senoufo	-0.3***	Uses manual spray	0.13*
Yakouba	0.35***	Uses tractor	0.10
Forest	0.42***	Uses ox-drawn equipment	0.01
Contact with WARDA	0.16***	Credit from any source	-0.03
Contact with ANADER	-0.05*	Uses fertilizer	0.15**
Contact with CIDT/GVC	-0.04**	Uses herbicide	0.10
Contact with SATMACI/SODERIZ	0.01	Minimum distance to market	-1.22
		Village electricity	0.14*
		Membership of organization	0.29***
Instrumental Variables			
Number of NERICA varieties known in the village	1.42***	Variety trial/dem in village	0.26**
Number of traditional varieties known in the village	4.36***	Reseach/Extension information	-0.23
Number of NARS upland varieties known in the village	-0.4**	Information from other farmers	0.08
Number of upland WARDA intraspecific varieties known in the village	0.75***	Info from other sources	0.07
Village participatory varietal selection	0.55***	Radio/TV information	0.00
Village contact with ANADER	-0.02	Owns radio	0.03
Village contact with CIDT/GVC	-0.26***		
Village contact with SATMACI/SODERIZ	0.12**		

Selection bias is plausible because the covariates are ‘unbalanced’ between treated and untreated groups. Only the percentage of farm devoted to rice produces conflicting results between countries. Consistent with the theoretical predictions, the treated group knows more varieties, has more assets, more farmland, more education, is more likely to be a member of social networks, and to have previous contact with WARDA. The lower levels of significance in the Nigerian data may reflect either weaker selection bias or the much smaller sample size.

3.7.2 Instrumental and Endogenous Variables

Nichols (2007) stresses “IV estimators are generally only as good as the excluded instruments”. In the Côte d’Ivoire case, the use of village-level variables is very effective at fulfilling the IV assumptions. On the one hand, given the importance of social networks, there is likely to be substantial correlation between village awareness of NERICAs and an individual farmer’s awareness. On the other hand, village awareness on its own is unlikely to change farmers’ adoption outcomes *other than* through increased individual awareness.

Unfortunately, this data was not collected in Nigeria. The one village-level variable available is the conduct of varietal trials in the village. There is a risk that this could boost adoption through related support to individual farmers from extension agencies. However, we have controlled for individuals’ participation in varietal trials and field demonstrations so arguably this variable captures only the information portion of extension services. Other instruments for awareness include research/extension information, information from other farmers, ownership of a radio, and radio/tv information. In Section 2 I argued that some forms of information could support adoption beyond increasing awareness so it is particularly important to test these instruments.

Endogenous variables which are *affected* by awareness or adoption include the proportion of income from rice, since we know NERICAs boost yields. Knowledge of other varieties, access to credit and contact with extension agencies may also be endogenous. An alternative source of endogeneity arises where innovations are *simultaneously* adopted as a package. Changes in fertilizer use and land management are therefore excluded from the first-stage model to avoid any bias.

3.8 Testing key assumptions

Appendix 2 reports a comprehensive set of tests for the key assumptions in the models. However, the central assumption of conditional independence in (A)-(E) cannot be formally tested and since there are reasons to believe that conditional independence may not hold, the results must be treated with caution.[63] The instrumental variables model (F) avoids this difficulty, but relies on the relevance and validity of the instruments. Tests suggest these broadly hold, but validity may be violated in Cote d'Ivoire under certain specifications despite the use of innovative instruments.

63 Imbens (2003).

4 Results

4.1 Adoption rates

Table 7 records the adoption rates derived from the various estimators of the treatment effects. It should also be recalled that these adoption rates are *net* of any changes in the covariates that might have indirect effects on adoption outcomes.[64]

Table 8 Estimated adoption rates according to different parameters and estimators (%)

Estimator		Côte d'Ivoire	Nigeria
Semi-parametric (A)	**ATE**	21%	46%
	ATT	36%	48%
	ATU	19%	41%
Semi-parametric (B)	**ATE**	20%	45%
	ATT	38%	48%
	ATU	18%	37%
Parametric (C)	**ATE**	*54%*	46%
	ATT	36%	48%
	ATU	*56%*	46%
Propensity Score (D)	**ATE**	*39%*	45%
	ATT	37%	48%
	ATU	*39%*	39%
Propensity Score (E)	**ATE**	19%	44%
	ATT	37%	48%
	ATU	17%	39%
Instrumental Variables (F)	**ATE**	*73%*	*56%*
	ATT	40%	51%
	ATU	*77%*	*68%*

[64] For instance, if awareness encourages fertiliser use, which in turn reinforces the incentive to grow NERICAs, then the actual impact of treatment is greater than reported here.

4.1.1 A problem with the parametric results

The most striking aspect of the results are the italicized anomalies, particularly for Côte d'Ivoire. These are associated with parametric estimators (C), (E) and (F),[65] and imply *negative* selection bias since ATE>ATT. This is contrary to most of the theory presented and also with the earlier comparison of group means which strongly suggested that farmers with the most information, assets and experience were more likely to be in the treated group.

The results reveal an underlying difficulty with the use of parametric estimators in the context of new technology adoption. This arises from the need on the one hand to make strict functional form assumptions to fit the models, and on the other hand the 'quasicomplete separation' noted in Section 3 due to the fact that the outcome variable is binary and y_0=0. This prohibits probit or tobit estimation because the cell (*d=0,y=1*) in the permutation matrix is empty and coefficients are consequently infinite.

Therefore this paper followed Caudill (1998) in applying a linear probability model. However, inspection of the results reveals that the exclusion of observations from the untreated group has no impact on the coefficient estimates in the parametric models. Clearly, the characteristics of the untreated group should be central to the coefficient estimates because it is the comparison of these characteristics which allows us to correct for selection bias. This insensitivity arises because in the linear probability model (applied to model (D) for illustration):

$$\alpha = \frac{\sum y_i d_i \sum I_i^2 - \sum y_i I_i \sum d_i I_i}{\sum d_i^2 \sum I_i^2 - (\sum d_i I_i)^2}, \text{[66] where } I_i = d_i[p(x)_i - \overline{p}(x)]$$

Typically, the coefficient would be sensitive to the dropping of observations where $d_i=0$ because even though no impact is made on terms such as $\sum y_i d_i$ and $\sum d_i^2$, I_i values would also be removed, altering the first terms in both the numerator and denominator.

[65] The difficulty also arises with the parametric estimator of the propensity score estimator (E).
[66] Adapted from Gujarati (2003), p.208.

However, our derivation under the condition that $y_0=0$ means that I_i is an interaction term between the demeaned propensity score and d_i. Hence, wherever $d_i=0$, there is *also* no variation in I_i. Consequently, the estimator α is invariant to the characteristics of the untreated group. It is unsurprising, then, that the ATE for (D) is close to the ATT; only information from the treated group is used in estimation.

The cause of the problem is that the controls we make between groups are only on the basis of heterogeneity *in responses*, and not in pre-treatment outcomes. This reflects the situation where $y_0=0$, since pre-treatment outcomes are by definition uniformly zero. But heterogeneity *in responses* is only observed for the treated group (for the untreated there is no variation in y_i), and this is the only information the estimator takes into account; it is simply not known how the untreated group with particular characteristics would behave once they are treated.

Yet, this inference is quite straightforward in the non-parametric approaches; we identify treated individuals with comparable *x* characteristics and attribute their outcome to the untreated. In the parametric approach, the equivalent 'matching' is achieved by controlling for heterogeneity in responses associated with differences in *x*. But this control function – because it is an interaction term – only reflects differences in the propensity to be treated (differences in *x*) *within* the treated group.[67]

Parametric methods then use *linearity* assumptions ($E(y_1|p(z))$ in the propensity score case) to extrapolate the control function to the untreated group and thereby ostensibly correct selection bias. It is assumed that changes in the propensity score have a linear effect on adoption outcomes across *both* the treated and untreated groups. Thus, the relationship between $p(x)$ and y_i within the treated group is extended to the untreated to infer what the impact of treatment would be on the untreated. Rather than directly matching treated and untreated individuals on the basis of *x* (or $p(x)$) (as non-parametric methods do), parametric approaches examine variation in the treated group and use linearity to assume that the same relationship holds across all individuals, allowing the ATE to be calculated.

So the parametric models may be very sensitive to linearity assumptions. This is exacerbated in the $y_0=0$ case because the standard covariate term is redundant and the linearity must be extrapolated solely from the interaction term relating to the treated group; the linearity is forced to do a lot more of the 'work' and reliance on it is inversely proportional to the level of exposure to treatment.

With a binary outcome variable, linearity is inevitably a more onerous condition and more difficult to assess. The linear probability model implies that individuals with propensity scores of, say, 0.45 and 0.50 are as similar as those with values of 0.01 and 0.06, when in practice the latter two are likely to be very different.[68] This is particularly important since with positive selection bias the treated group will have an above-average propensity score anyway; we cannot rely on 'local' linearity near the mean.[69] Formal tests of functional form are rejected in all the parametric models, [70] and Appendix A2.2 reports a simple test suggesting that many of the covariates should not be represented linearly.

Moreover, the small size of the treated group for Cote d'Ivoire (11% of the sample) risks introducing substantial errors from extrapolation if the coefficient on the interaction term is imprecise. Finally, it is possible that the *slope* of the propensity score's impact on adoption is intrinsically different between the treated and untreated group. This would clearly prevent robust linear inference and could arise due to the endogeneity issues raised earlier. For example, if treatment influences farm management or social network behaviours, the

[67] In some respects this is obvious; the only information we have about the relationship between x characteristics and adoption choices is for the treated group and so there is a great reliance on this information to make much broader conclusions about the potential behaviour of the untreated.

[68] Imbens and Wooldridge (2008), p.29. A standard problem with the linear probability model – that fitted probabilities are not constrained to the range {0,1} – does not seem to pose any difficulties in this case.

[69] For instance, for Cote d'Ivoire, the average of I_i is 0.026 despite the demeaning, because the interaction term focuses on the treated group only while the mean $p(x)$ is taken across the full sample.

[70] Undertaken using the Breusch-Pagan heteroskedasticity test and Ramsey's RESET test for functional form.

apparent sensitivity of adoption to these characteristics may be reduced. While I have tried to avoid this endogeneity, the complexity and interdependence of these decisions make it unlikely that the possible sources have been exhausted.

Of course, probit or logit models may prove a better distributional fit, but as show earlier these *cannot* be applied in the case of quasicomplete separation, which again derives from $y_0=0$. At root, then, the weakness of the models is their dependence on parametric assumptions of linearity. Hence, ATE estimates for the parametric models are unlikely to be robust.[71] The reliability of Diagne's (2006) parametric estimates in this context remains unclear.[72]

In summary, use of the linear probability model may be inappropriate where (i) both the treatment and outcome variables are binary, (ii) $y_0=0$ because we are dealing with a new technology, and (iii) there is incomplete exposure.[73]

4.1.2 Non-parametric results

Focussing on the non-parametric estimators for Côte d'Ivoire the ATT range is 36-38%. The estimated adoption rate given full exposure – the ATE – is 19-21%.[74] Recall that these figures must be caveated by their reliance on conditional independence and their ability only to account for selection on observables.

Since ATT>ATE, positive selection bias means a focus on the exposed sample adoption rate would over-estimate the population adoption rate by a factor of two; those most likely to adopt are over-represented in the treated group. Figure 5 illustrates the selection bias graphically. The close correspondence between the ATE and the ATU reflects the low

[71] This is ironic given that the behavioural assumption for the IV estimator of heterogeneity in responses but not choices is perhaps most plausible in the new technology adoption setting. Yet, it is precisely this setting that entails $y_0=0$ and therefore the failure of the IV technique.

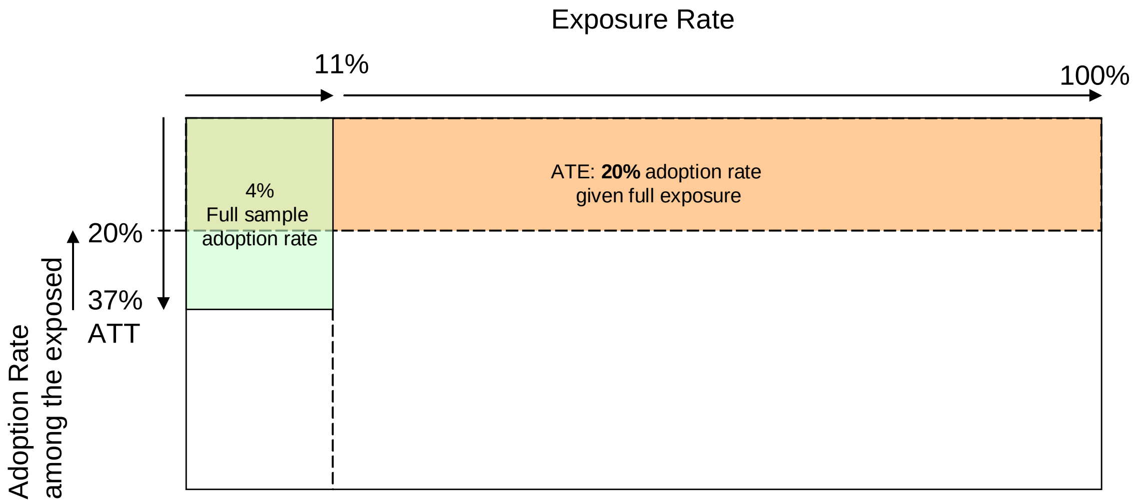
[72] While Diagne (2006) uses the number of NERICA varieties adopted as the dependent variable, similar problems are likely to arise for the large proportion of the sample who are not exposed and so adopt zero varieties. Further, Diagne does not appear to employ a poisson model for second-stage estimation, as would be appropriate for this count data.

[73] The parametric method is less problematic where $y_0 \neq 0$ because I_i includes a non-interacted covariate term which uses information from the full sample. Nevertheless, linearity is still assumed.

[74] This result is broadly consistent with the estimate provided by Diagne and Demont (2007) of 18%.

exposure rate of the population; with only 11% of the population exposed, the overall population much more closely resembles the untreated group.

Figure 5 The adoption rate parameters for Côte d'Ivoire

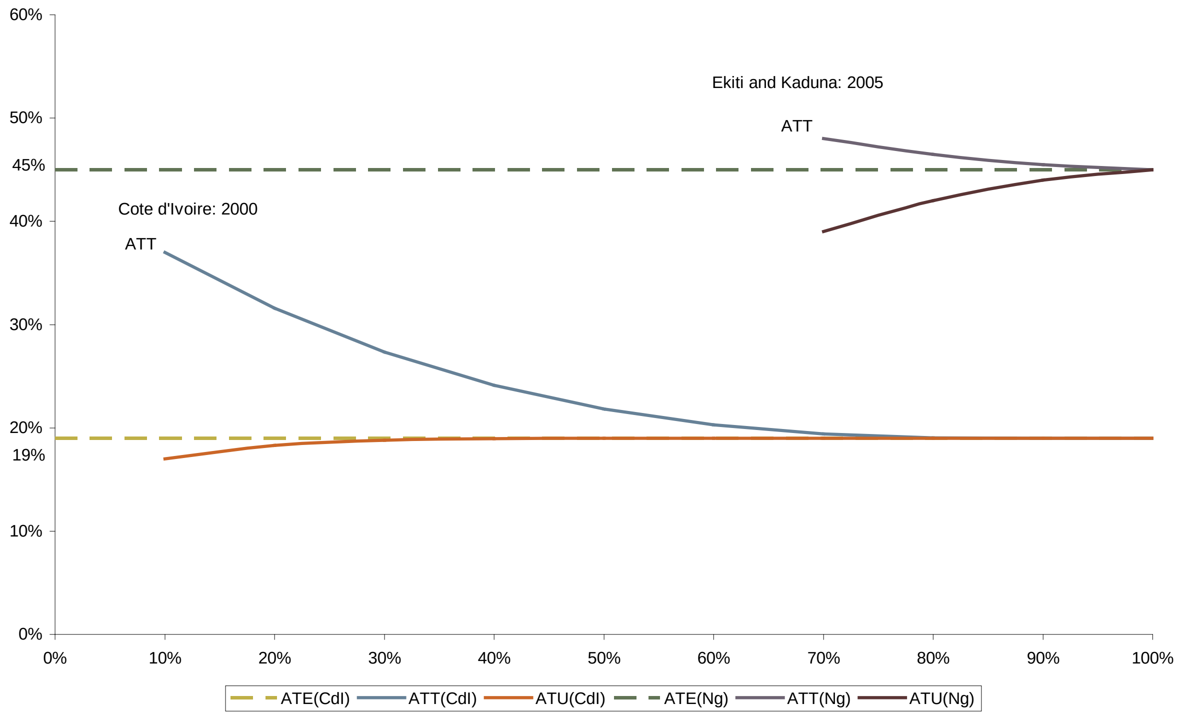


For Nigeria, the parametric estimators are closer to their non-parametric counterparts. However, this is likely to reflect the weak selection bias that is present, and since the parametric estimators still predict the wrong direction of selection bias they are ignored here. The non-parametric ATT estimates are 48% and the ATE 45-46%. There is evidence for marginal selection bias, although much less than in Côte d'Ivoire (it is possible this is a consequence of the small sample size). In this case the ATE is closer to the ATT, reflecting the higher exposure rate (71%). Compared to Spencer et al (2006), the ATE estimate is considerably higher having corrected for non-exposure bias.

Figure 6 illustrates how the adoption rates are likely to change as exposure increases throughout the population.[75] The differential starting positions of the two cases are clear; Côte d'Ivoire begins from 11% exposure while Nigeria begins from 71% exposure. For this reason, the ATT converges faster for Côte d'Ivoire and the ATU converges faster for Nigeria.

Comparatively, the ATT is higher in Nigeria than Côte d'Ivoire, indicating that actual adoption has been higher. Further, compounded by the greater selection bias in Côte d'Ivoire, the ATE is considerably lower. However, differences in survey timing are likely to account for much of this difference (see Section 5). Other WARDA surveys show slightly higher ATE adoption rates; 50% in Benin in 2004 and 58% in Guinea in 2001.[76]

Figure 6 Changes in adoption rates as exposure increases



Appendix 3 reports the results where the outcome variable is the *intensity* of adoption.

4.2 Determinants of Exposure

Before examining the determinants of adoption, it is instructive to consider the determinants of exposure to treatment – equivalently, the factors driving selection – as estimated by the propensity score.

[75] The propensity score estimates are used for illustration and the assumption is made that convergence to full exposure is achieved along a logistic path.
[76] Somado et al. (2008), p.125.

Table 9 The determinants of exposure[77]

Côte d'Ivoire		Nigeria	
Constant	-6.18	Constant	4.05
Log of farm size (max. total area cultivated in the past 5 years)	0.074 (0.023)	Total size of all plots	-0.009 (0.019)
Idle (uncultivated) area share in 1999	0.129 (0.642)	Max rice area share in 2005 (%)	1.351 (0.566)**
Rice area share in 1999	-0.989 (0.727)	Rice sale 2004 % (not binary)	-0.002 (0.006)
Commercialized crops (including rice) area share in 1999	0.617 (0.508)	Number in household	-0.016 (0.017)
Log of total agricultural cash income in 1999	0.036 (0.120)	Age	0.018 (0.014)
Income from rice in 1999 (thousands CFA)	0.001 (0.003)	Was born in the same village	0.005 (0.010)
Household Size	0.033 (0.024)	Years lived in village	0.005 (0.081)
Age	-0.008 (0.012)	Secondary employment (positive where >10% income)	-0.104 (0.821)
Was born in the same village	0.095 (0.325)	Highest formal school	0.208 (0.325)
Years of residence in the village	0.008 (0.010)	Practiced upland rice cultivation in 2003 or 2004	0.423 (0.276)
Has a secondary occupation	0.498 (0.223)**	Varietal trials on household land in the past	-0.199 (0.297)
Years of formal schooling	0.020 (0.029)	Female	-0.199 (0.297)
Practice upland rice cultivation	2.00 (0.597)***	Yoruba	-6.403 (1.467)***
Female	0.444 (0.249)*	Hausa	-5.805 (1.404)***
Bete	-1.594 (0.563)***	Kadara	-5.794 (1.457)***
Senoufo	0.103 (1.01)	Kind of school attended (koranic vs. formal)	0.352 (0.214)*
Yakouba	-0.216 (0.611)	Owns bicycle	0.176 (0.280)
Forest	-1.877 (1.528)	Owns motorcycle	-0.114 (0.275)
Number of NERICA varieties known in the village	1.10 (0.290)***	Owns car	0.269 (0.367)
Number of traditional varieties known in the village	-0.003 (0.014)	Minimum distance to market	0.000 (0.015)
Number of NARS upland varieties known in the village	0.510 (0.135)***	Village electricity	0.343 (0.336)
Number of upland WARDA intraspecific varieties known in the village	0.356 (0.165)**	Variety trial/demonstration in village	0.343 (0.336)

[77] *=significant at 10%, **=at 5%, ***=at 1%.

Village participatory varietal selection	0.975 (0.419)**	Reseach/Extension information	-0.486 (0.286)*
Village contact with ANADER	0.231 (0.267)	Information from other farmers	0.010 (0.289)
Village contact with CIDT/GVC	-0.679 (0.680)	Info from other sources	0.199 (0.528)
Village contact with SATMACI/SODERIZ	2.398 (0.685)***	Membership of an Organisation	0.515 (0.255)**
		Radio/TV information	0.280 (0.425)
		Owns radio	0.059 (0.558)

In Côte d'Ivoire, the most significant determinants of exposure are that farmers practice upland rice cultivation, have a secondary occupation and report Bete as their ethnicity. Sex differences are evident at the 10% level. At the village level, high levels of crop variety knowledge and contact with extension agencies all increase the probability of exposure. Nearly all of these variables seem to capture elements of farmers' social networks.

In Nigeria, the share of land devoted to rice, ethnicity and membership of an organisation are all significant at the 5% level. Kind of school attended and access to research information are also significant at 10%. Again, these all plausibly reflect the density of social networks and information flows. The only counter-intuitive result is that research information appears to reduce the probability of hearing about NERICAs.

The findings refine some of the implications from the comparison of sample means. Education, membership of social networks and contact with extension agencies are indeed important means by which information about NERICAs is transmitted. However, there is no significant evidence that farm size or assets boost exposure. The difference arises because the propensity score holds all other variables constant; so *given* their participation in social networks, farmers with larger farms are no more likely to adopt. This is important because it implies farm size is not necessarily a *barrier* to exposure. By encouraging small farmers to join social networks, WARDA can directly increase awareness of NERICAs.

4.3 Barriers to Adoption

The barriers to adoption are identified after controlling for exposure, showing what the barriers would be even if the population was fully exposed. Given that the parametric estimators do not adequately control for exposure, barriers are identified from the non-parametric estimators.

ATE(x) adoption rates from estimator (A) are predicted for subgroups of interest in the sample and the difference in means between groups is tested using a t-test (Tables 11 and 12).

The analysis highlights how adoption rates between groups can vary; in Côte d'Ivoire farmers who have had contact with WARDA exhibit a 52% adoption rate compared to 17% among those without contact. Participating in PVS trials also helps. Practicing upland rice cultivation has an unsurprisingly positive effect. Having a secondary occupation boosts uptake from 18 to 24%. Being older and living outside the forest zone also encourages adoption. Surprisingly, devoting more land to commercial crops and having a higher income from agriculture reduce the likelihood of adoption. This suggests credit and liquidity constraints are not significant.

Of course, it is an open question whether groups with higher adoption rates should be targeted to maximize impact, or if groups with lower adoption rates should be targeted to ensure equal access to the gains from adoption.

Table 11 Testing differences in ATEs between Ivorian sub-groups

	Mean	Standard Error	t-test of mean equality
Above average farm size	0.195	0.141	
Below average farm size	0.221	0.017	Not rejected
Bete ethnicity	0.251	0.026	
Non-Bete ethnicity	0.196	0.012	Rejected at 10%
Male	0.207	0.013	
Female	0.204	0.019	Not rejected
Contact with WARDA	0.52	0.045	
No Contact with WARDA	0.175	0.011	Rejected at 1%

Inside forest zone	0.187	0.014	
Outside forest zone	0.228	0.017	Rejected at 10%
Secondary occupation	0.236	0.017	
No secondary occupation	0.182	0.014	Rejected at 5%
Above average commercial crops share	0.186	0.016	
Below average commercial crops share	0.227	0.015	Rejected at 10%
Above average agricultural cash income	0.170	0.014	
Below average agricultural cash income	0.245	0.017	Rejected at 1%
Above average age	0.243	0.017	
Below average age	0.178	0.014	Rejected at 1%
Above average years of schooling	0.212	0.014	
Below average years of schooling	0.194	0.018	Not rejected
Practice upland rice cultivation	0.251	0.014	
Don't practice upland rice cultivation	0.096	0.015	Rejected at 1%
Participation in PVS Trials before 2000	0.299	0.011	
No participation	0.200	0.040	Rejected at 5%

In Nigeria, motorcycle ownership increases adoption from 38 to 60% and bicycle ownership is nearly as effective. Being a member of an organisation and using herbicide have a similar effect. Most important, however, appears to be the presence of electricity in the village, which increases adoption from 28% to 54%. Indeed, a farmer owning a bicycle, participating in local organisations, using herbicide and having access to electricity is four times more likely to adopt NERICAs than a farmer without those characteristics.[78] Complementary policy interventions in these areas will be needed to maximize NERICA adoption.

Table 12 Testing differences in ATEs between Nigerian sub-groups

	Mean	Standard Error	t-test of mean equality
Above average farm size	0.484	0.024	
Below average farm size	0.448	0.036	Not rejected
Owns a bicycle	0.517	0.028	
Doesn't own a bicycle	0.404	0.026	Rejected at 1%
Owns a motorcycle	0.601	0.030	
Doesn't own a motorcycle	0.377	0.023	Rejected at 1%
Male	0.478	0.022	
Female	0.400	0.041	Not rejected
Credit use	0.481	0.026	
No credit	0.44	0.03	Not rejected

[78] WARDA's (2004) 'Strategic framework for rice sector revitalization in Nigeria' pays little attention to these barriers (with the exception of herbicide use, p.17).

Membership in organisation	0.501	0.026	
No membership	0.396	0.029	Rejected at 1%
Fertiliser use	0.46	0.023	
No fertiliser use	0.461	0.039	Not rejected
Herbicide use	0.489	0.023	
No herbicide use	0.379	0.034	Rejected at 1%
Ekiti state	0.476	0.028	
Kaduna state	0.446	0.028	Not rejected
Village electricity	0.535	0.022	
No village electricity	0.281	0.031	Rejected at 1%
Above average age	0.465	0.031	
Below average age	0.456	0.025	Not rejected
Practice upland rice cultivation	0.46	0.020	
Don’t practice upland rice cultivation	0.62	0.088	Not rejected
Above average distance from market	0.500	0.028	
Below average distance from market	0.428	0.027	Rejected at 10%
Kadara ethnicity	0.437	0.051	
Non-Kadara ethnicity	0.465	0.021	Not rejected

In summary, the evidence supports only a minority of the potential barriers identified in Section 2. Farm size, education and credit/cash availability do not appear significant. Asset and input ownership, membership of social networks and contact with extension agencies appear to be the most important barriers. The latter two are particularly striking, because they are significant *beyond* their role in promoting awareness, confirming the hypotheses of Conley and Udry (2003) and Kijimi et al (2006) in Section 2. The Schultz hypothesis of a similar dual role for education is rejected, but the general conclusion that social and structural dynamics play complementary and overlapping roles in explaining adoption feeds back to help reconcile the theoretical debate.

4.4 Identifying target groups

While I have separated the determinants of exposure and adoption to highlight the different dynamics of each, in practice WARDA is concerned with achieving both. ‘Quick wins’ can be targeted by identifying “adopter types” who do not face strong barriers to adoption but have yet to be exposed to knowledge of NERICAs. A focus on disseminating information to this target group could be a cheap and effective means of boosting NERICA adoption. Of course, some characteristics are conducive to both awareness and adoption. The interesting characteristics are those which are conducive to adoption but which have a broadly neutral or negative impact on the probability of exposure.

The evidence suggests that in Côte d'Ivoire such farmers are older, grow fewer commercial crops, have less agricultural cash income and live outside the forest zone. Given the low level of current exposure, WARDA may wish to focus on these – presumably more rural – farmers outside the forest zone to maximize adoption in the short-run.

In Nigeria, these farmers seem to own bicycles or motorcycles. One novel policy in Nigeria could be to provide information through bicycle or motorcycle repair shops, since there is a striking relationship between ownership and adoption even though owners are no more likely to have heard of NERICAs.

5 NERICAs' long-run potential

5.1 Modelling adoption over time

The preceding methods provide spot estimates of the potential adoption rate under full exposure *at the point in time when the surveys were carried out*. Yet, even with full exposure, adoption rates are not static. As Section 2 stressed, farmers do not make a once-and-for-all decision about adoption. Moreover, the barriers to adoption are unlikely to be insurmountable; a farmer may need to earn a sufficient surplus or have access to adequate support only once to achieve successful adoption. Once this threshold has been crossed, the barriers to *continued* cultivation are likely to be lower than to first-time adoption. Simply by stochastic processes then, we would expect adoption rates to converge to their potential only slowly over time.

The aim in this section is to understand the *long-run potential* for NERICA adoption. Following Diagne (2006) I exploit the inclusion of an additional year's recall data in the surveys to estimate the likely diffusion trajectory of NERICAs. Assuming adoption follows an autoregressive (AR1) process, we have:

$$\alpha = \alpha_0 + \gamma y_{t-1}$$

This decomposes the treatment effect into two components: α_0 reflects the treatment effect on someone who is aware of NERICAs but has not adopted prior to year *t*, and γy_{t-1} captures the treatment effect in year *t-1*. Instead of considering a binary outcome variable we focus on the number of NERICA varieties that have been adopted. This allows us to model adoption as a poisson process under the assumption that the adoption of each variety of NERICA is binomial independent.[79]

Substituting *α* into equation (F) gives:

$$y_i = \alpha_0 d_i + \gamma d_i y_{t-1} + \beta_x d(x - \bar{x}) + de_1$$

which can be estimated with the IV model.[80] The starting point of our series is then given by the probability of first-time or repeat adoption on or before *t,* $1 - e^{-\alpha_0 - \gamma y_{t-1}}$. The end point is given by the long-run population adoption rate, $1 - e^{\frac{-\alpha}{1-\gamma}}$. The trajectory between the two points follows a formula for count variables derived from Freeland and McCabe (2004).[81]

$$P(y_{t0} + h > 0 | y_{t0} \geq 0) = 1 - e^{-\alpha(\frac{1-\gamma^h}{1-\gamma})} \text{ where h=1,2, …}$$

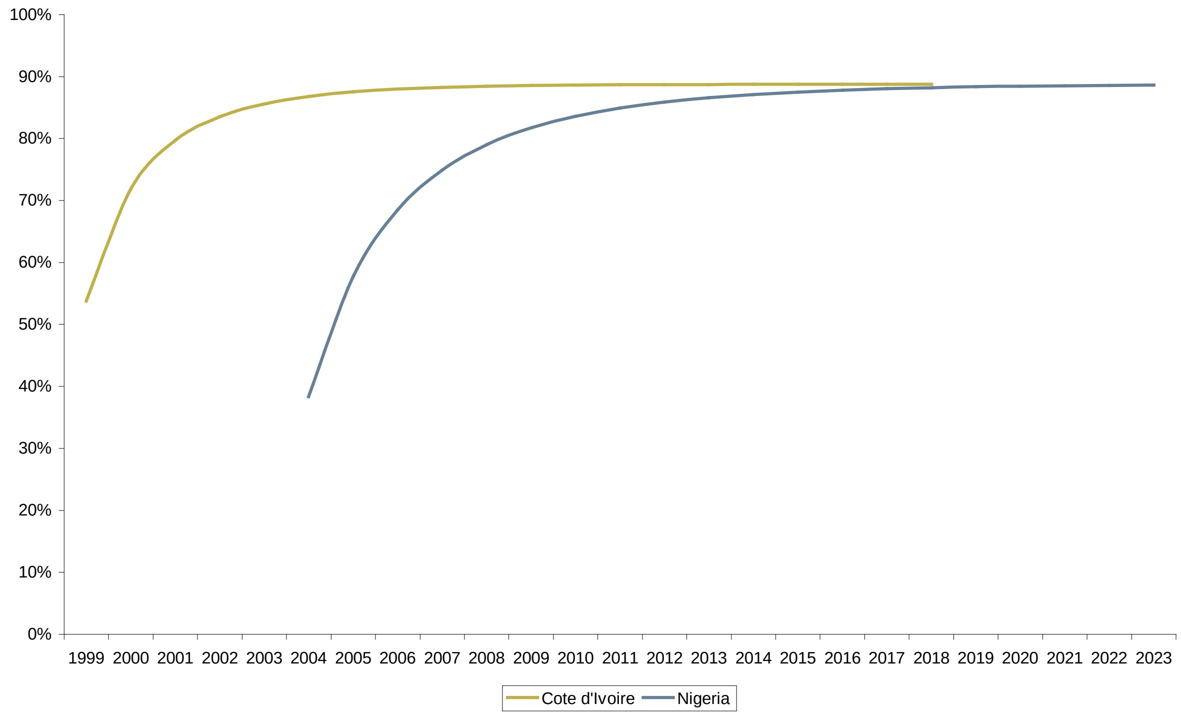
Figure 7 graphs the results. Clearly Côte d'Ivoire has a head-start on Nigeria, but the two converge at similar rates to the identical long-term population rate of 89%. This suggests the magnitude and persistence of barriers to adoption are comparable in both countries.

[79] Binomial independence is a simplifying assumption. Arguably it will not hold in this context because experience with one hybrid variety is likely to increase the probability of adopting a second hybrid variety. However, the number of NERICA varieties is small and the assumption is likely to facilitate a useful first approximation.

[80] Unfortunately, the more robust non-parametric estimators are not directly amenable to the inclusion of an autoregressive term. The results should be treated with appropriate caution. One avenue for research is the combination of non-parametric estimators for long-run adoption potential with robust estimates of the autoregressive parameter.

[81] Note that the equation appearing in Diagne (2006) appears to confuse the parameters. Moreover, the term $(1-\alpha^h)^{yto}$ in Diagne's equation (5) is redundant and actually introduces bias into the starting point of the trajectory.

Figure 7 **Projected adoption curves over time for Côte d'Ivoire and Nigeria under full exposure**

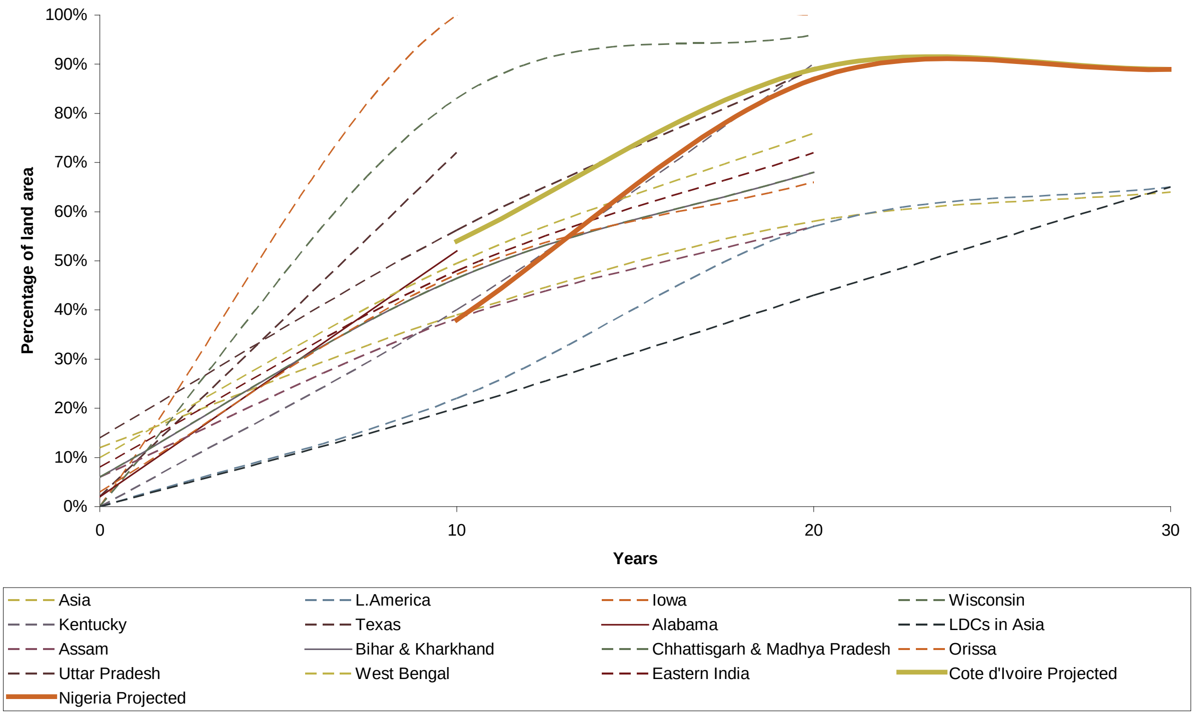


It would be interesting to combine the two analyses – of adoption rate among the treated as exposure increases and of adoption rates given full exposure – to provide projections for the likely path of actual adoption rates. Unfortunately, there is no time-series data for awareness of NERICAs and the rate of diffusion of information remains unknown. Since this would be of use to WARDA in planning their support provision, there is a strong argument for recording awareness in longitudinal studies.

5.2 Comparison with other Green Revolutions

The time-series estimates made above can be compared with the historical rates at which high-yielding varieties were adopted elsewhere. Two features are of interest; the peak or equilibrium population adoption rate, and the speed with which this is achieved. Speed is important since the success of Green Revolutions has been founded not just on their breadth, but on achieving *simultaneous* adoption that has provided impetus to the rapid development of upstream and downstream markets. It is of course this speed which offers the greatest poverty-alleviating prospects for NERICAs as well.

Figure 8 Comparison of adoption trajectories



Sources: Evenson and Gollin (2003), Grilliches (1957), National Bank for Agriculture and Rural Development (undated).

The peak adoption rate compares favourably, matching leading states in India such as Utter Pradesh. The speed of adoption is also somewhat faster than most cases, lagging only US states such as Wisconsin. This is particularly striking given that the start of our series seem to match the Indian cases only a decade after initial adoption. NERICAs have achieved these adoption levels in considerably less than a decade. In short, the evidence points to the scope for NERICAs to contribute to a Green Revolution.

6 Conclusions

The analysis provides five principal findings.

1. **The potential adoption rates for NERICAs in Côte d'Ivoire and Nigeria are at a level (89%) that could stimulate a Green Revolution.** It should be stressed that NERICAs on their own could not constitute a Green Revolution; nor is this level of adoption likely to be achieved automatically. Yet, with complementary innovations in other areas such as fertilizer application, farm management and in labour markets, there is a unique opportunity to boost West African agricultural productivity.

2. **Simple adoption rate estimates can be misleading where awareness is not universal.** Non-exposure and selection biases were strong in Côte d'Ivoire and specific program evaluation techniques are required to robustly estimate meaningful adoption rates. In the case of new technology adoption, non-parametric estimators appear more reliable than their parametric counterparts. In Côte d'Ivoire, the adoption rate in 2000 if the population had been fully aware of NERICAs was 19-21% and in Nigeria in 2005 it was 45-46%.

3. **Similarly, barriers to adoption may be misidentified.** With appropriate controls, the key barriers to adoption in Côte d'Ivoire are identified as having a secondary occupation and contact with WARDA. This suggests WARDA's support role should go well beyond publicising NERICAs. In Nigeria, organisational membership, herbicide use, village electricity and bicycle ownership are the key barriers.

4. **Both social learning and structural constraints determine adoption patterns.** The diffusion of awareness and adoption share important conduits; organisational membership and extension services matter beyond their role in raising awareness, confirming key hypotheses. Yet, they also follow somewhat different dynamics. Social networks – in the form of education, organisational membership and contact with

extension agencies – are the principal determinants of awareness while structural factors such as farm organization, asset ownership and technical support matter more for adoption. The implication for the theoretical debate is that social and structural theories of adoption should be considered complementary and be embedded in the broader framework I have adopted. A fruitful area for future research would be the temporal and spatial interaction of the two processes which our data does not allow us to investigate.

5. **Key target groups offer 'quick wins' for boosting adoption.** Certain characteristics seem to make farmers amenable to adoption, but less aware of new technologies. In Côte d'Ivoire WARDA may wish to focus on older, poorer and less commercial farmers in rural areas outside the forest zone. In Nigeria, providing information to bicycle or motorcycle owners may be effective.

In addition, this paper has made a number of methodological contributions that can be generalized to other studies of new technology adoption. Alternative selection mechanisms posited by the literature have been linked to different forms of heterogeneity in the population, and to consistent estimators in each case. Simplified estimators have been provided for the new technology adoption scenario where adoption is impossible without awareness. However, the use of parametric models may be inappropriate where (i) both the treatment and outcome variables are binary, (ii) $y_0=0$ because we are dealing with a new technology and, (iii) there is incomplete exposure.

Appendix 1 Derivation of estimators

A1.1 (A) Semi-parametric Estimator 1

The following proof follows Wooldridge (2002).[82] Taking conditional expectations over the equation linking observed to potential outcomes:

$$y_i = d_i \cdot y_{1i} + (1-d_i)y_{0i} = y_{0i} + d_i(y_{1i} - y_{0i})$$

$$E(y|x,d) = E(y_0|x,d) + d\left[E(y_1|x,d) - E(y_0|x,d)\right]$$

In the NERICA adoption case, we can set $y_0=0$:

$$E(y|x,d) = dE(y_1|x,d)$$

and applying the conditional independence condition:

$$E(y|x,d) = dE(y_1|x)$$

Next, separate out the cases of *d=1* and *d=0* and calculate the difference to obtain an expression for the ATE conditional on x, ATE(x):

$$E(y|x,d=1) = E(y_1|x),\ E(y|x,d=0) = 0$$

$$ATE(x) = E(y_1|x) - E(y_0|x) = E(y|x,d=1) - E(y|x,d=0) = E(y|x,d=1)$$

[82] Wooldridge (2002), p.609.

There is a particular ATE(x) for each possible set of values that x can take. A consistent non-parametric estimate for the unconditional ATE can then be obtained from the sample by simply averaging the ATE(x) across the full sample:

$$\text{(A)} \quad A\hat{T}E = \frac{1}{n}\sum_{i=1}^{n}\left[E(y|x, d=1)\right]$$

For example, we obtain the difference in *y* between treated and untreated individuals who own a radio, then do the same for those who don't own a radio and take a weighted average by the population in each group. Obtaining ATT performs the same averaging process, but for the treated sub-sample where *d=1*:

$$A\hat{T}T = \frac{1}{\sum_{i=1}^{n} d_i}\left\{\sum_{i=1}^{n} d_i\left[E(y|x, d=1)\right]\right\}$$

Vitally, by employing the conditional independence assumption, these expressions are only in terms of *y* and not the unobservable y_1, so estimation is straightforward. While it is possible to use fully non-parametric methods (Härdle and Linton, 1994) to estimate $E(y|x, d=1)$, little is lost by employing flexible parametric methods. This would include a standard OLS regression where *y* is continuous and a logit or probit model where *y* is binary.

A1.2 (B) Semi-parametric Estimator 2

A particularly simple form of the estimator is provided by Diagne and Demont (2007).

$$y = dy_1 + (1-d)y_0 = dy_1$$

Taking expectations of the outcome equation and applying conditional independence:

$$E(y|x, z) = E(dy_1|x, z) = E(d|x, z) \cdot E(y_1|x, z)$$

Our notation allows *z* to represent the determinants of exposure (which *x* is redundant in explaining), $P(d=1|x,z)=P(d=1|z)$, and *x* to represent the determinants of adoption (which *z* is redundant in explaining), $P(y_1=1|x,z)=P(y_1|x)$. The first condition is automatically fulfilled since *x* is a subset of *z*, and the second condition is simply the instrumental variables validity condition that the instruments do not appear in the structural equation for adoption. With this construction, the equation can be rewritten as:

$$E(y|x,z)=E(d|z)\cdot E(y_1|x)$$

The first term simply refers to the probability of exposure and can be replaced by a propensity score *p(z)* estimated by a logit or probit model together with the assumption that *p(z)*>0.

$$E(y|x,z)=p(z)\cdot E(y_1|x)$$

The following simple estimators then follow:

$$ATE(x)=\frac{E(y|x,z)}{p(z)} \quad ATE=E(\frac{y}{p(z)}) \quad \text{(B)} \quad A\hat{T}E=\frac{1}{N}\sum_{i=1}^{N}(\frac{y_i}{\hat{p}(z_i)})$$

A1.3 (C) Parametric Estimator

Alternatively, a full parametric model can be developed. This has the advantage of mitigating the ‘dimension problem’, which causes difficulties for non-parametric estimation where there are many dimensions of *x* and therefore few individuals who share *x*-characteristics. Including the individual-specific gains, v_i, in the equation for the observed outcome:

$$y=dy_1=d(u_1+v_1)=du_1+dv_1$$

Taking conditional expectations:

$$E(y|d,x)=\alpha d+dE(v_1|x)$$

and expressing $E(v_i|x)$ as a linear function of *x* to be estimated, we obtain the regression:

$$\text{(C)} \quad E(y|d,x) = \alpha d + \beta_x d\left[x - \bar{x}\right]$$

By capturing the impact of treatment once we have controlled for *x*, $\alpha = u_1 = E(y_1)$ is our estimator of the ATE.

Conceptually, the process is similar to the non-parametric approach, simply partialling out the observable influences on treatment and outcomes to obtain the average effect of treatment. Imbens and Wooldridge (2008) show for the more general case where $y_0 \neq 0$ that the ATE estimator can be decomposed into the difference in sample means and a correction for the difference in covariates, weighted by the population in each sub-group:[83]

$$\alpha = \bar{y}_1 - \bar{y}_0 - (\frac{N_0}{N_0 + N_1} \cdot \hat{\beta}_1 + \frac{N1}{N0 + N1} \cdot \hat{\beta}_0)'(\bar{x}_1 - \bar{x}_0)$$

The interesting feature of regression (C) is that it includes an interaction term between the treatment variable, d_i, and the demeaned values of *x*. This reflects the possibility of *heterogeneous effects* from treatment; depending on their characteristics *x*, farmers are likely to react differently to treatment in deciding whether (or how much) to adopt.

The final equation is different to that estimated by Wooldridge (2002) because of the condition that $y_0 = 0$. Specifically, the constant and covariate terms $\gamma_0 + \beta_x x$ are missing. The intuition for this is that these terms represent “levels” differences in the starting point for the outcome variable y_0. Since this is constrained to be zero, there is no variation in the initial level of adoption outcome. The only variation in treatment effects arises from the heterogeneous impact once individuals have become aware of NERICAs. That is to say that

[83] Imbens and Wooldridge (2008), p.24.

while there is heterogeneity in *responses* to treatment, there is no heterogeneity in the pre-treatment outcome (see discussion in Section 3.6).[84]

When the above outcome is a binary measure of adoption, a problem is posed for standard logit and probit models by the fact that $y=0$ whenever $d=0$. Zorn (2005) describes the problem as 'quasicomplete separation', reflecting the fact that a single cell ($d=0, y=1$) in the permutation matrix is empty (see the example for Côte d'Ivoire in Table x).[85] Intuitively, as separation increases, the ability to predict y given d increases and in the limit becomes perfect.[86] Consequently, parameters and standard errors are infinite. While there are remedial measures available, they involve the addition of artificial data or are valid only under strict data criteria.[87] Where such estimates are required in this paper we have adopted the linear probability model as a simple alternative suggested by Caudill (1988).[88] The difficulties this creates for estimation are discussed in the main text.

Table A.1 Permutation matrix for treatment and outcome dummy variables

	y=0	**y=1**
d=0	1369	0
d=1	87	53

[84] Note that this is not saying that differences in the covariates across the population are immaterial. These pre-determined factors have a strong impact on adoption outcomes, but only *once* treatment has taken place. They do not affect the actual starting point that we wish to measure; all agents start from zero adoption when they are untreated. In other circumstances this need not be the case, but it is a characteristic of the analysis of the adoption of new technologies.

[85] Zorn (2005), p.5.

[86] For complete separation, the reason is obvious: whenever $d=0$, $y=0$ and whenever $d=1$, $y=1$, so there is no variance in y left to be explained by the other covariates. The problem may seem less serious when we have quasicomplete separation such that whenever $d=0$, $y=0$, but when $d=1$, y can take any value. Zorn (2005) notes that while other covariates may be relatively unaffected, the parameter and standard error on d will still be infinite. Since d is a critically important variable in our model, the problem is a severe one.

[87] It is unclear how the estimators in Diagne (2006) have been corrected to accommodate this problem.

[88] Caudill (1988), 'An advantage of the Linear Probability Model over Probit or Logit;', *Oxford Bulletin of Economics and Statistics*, 50(4), November, pp.425-427.

A1.4 (D) Propensity Score Estimator (parametric)

There is a large literature on the use of ‘matching’ to estimate the ATE. Pairs of individuals with similar x-characteristics from the treated and untreated group are identified and, provided we believe conditional independence holds, then the difference in the outcome variable between the two represents the ATE(x). Averaging across all the pairings of *x* gives the ATE. One difficulty with the approach is that with many dimensions of *x*, there may be few similar cases both in the treated and untreated groups; described as the lack of a ‘common support’.

Rosenbaum and Rubin (1983) show that a one-dimensional matching estimator that combines all the *x* and *z* variables to create a ‘propensity score’ $p(x) = \Pr(d = 1|x)$ that controls for treatment is equivalent to matching under the assumptions that (i) $0 < p(x) < 1$ and (ii) $E(y_0|p(x))$ and $E(y_1|p(x))$ are linear in p(x). The propensity score can be estimated by a logit or probit model and then replaces the control function x in the parametric regression:

$$\text{(D)} \quad E(y|d,x) = \alpha d + \beta_i d\left[p(x) - \overline{p}(x)\right]$$

A1.5 (E) Propensity Score Estimator (Semi-parametric)

Rosenbaum and Rubin (1983) prove that the conditional independence assumption can be directly expressed in terms of the propensity score rather than the covariates *x*;

$$d \coprod (y_0, y_1)|p(x)$$

The estimator is then derived in an analogous way to (A) to get the following expression which is averaged over the full sample to obtain the ATE.

$$E(y|p(x), d = 1) = E(y_1|p(x))$$

$$\text{(E)} \quad A\hat{T}E = \frac{1}{n}\sum_{i=1}^{n}\left[E(y|p(x), d = 1)\right]$$

A1.6 (F) Instrumental Variables Estimator

The above techniques are contingent on the validity of the conditional independence assumption. But it many cases we may not be able to control for the factors that determine the selection mechanism so that conditional independence fails. Our discussion of the data in Section 2.2 showed that there are key determinants of adoption (and therefore potential sources of selection bias) – for example, plot-specific conditions and risk preferences – that we cannot control for due to data limitations. Hence, there is a justification for implementing a model that facilitates robust estimation in the presence of unobservables. Moreover, it has been argued that the conditional independence assumption should only be made where we have "extensive knowledge of the process determining treatment status".[89] This is unlikely since these processes are case-specific and have not received sufficient research attention. Evenson and Mwabu (1998) provide empirical evidence for the strong role of unobservables such as farm management techniques in determining yields.

Where unobservable factors cause heterogeneous gains from treatment, selection is determined by those gains and is bound-up in e_i, the residuals term. The correlation this produces between d_i and e_i prevents us using ordinary OLS regression estimation to obtain consistent estimators of the ATE.

As Blundell and Costa Dias (2002) argue,[90] the introduction of instrumental variables, z_i, can help resolve the problem by producing exogenous variation in treatment that does not affect outcome variables (other than through the impact on treatment), mimicking the random assignment of treatment in a social experiment. This enables us to control for any biases in selection and derive consistent estimators. An interesting interpretation of IV's is given by Moffitt (1996) as "representing a comparison in a different dimension" to the treatment

[89] Angrist (undated).
[90] p.15

variable.[91] Thus, in the binary case, rather than comparing the outcomes in the treated and untreated groups to derive the ATE we compare the outcomes in the group with *z=1* with the outcomes in the group with *z=0*.

The use of instrumental variables requires us to fulfil the two standard assumptions:

i. Be relevant (the inclusion restriction),[92] so that *z* actually affects the probability of treatment. COR(z,d)≠0 or $L(d|z,x\rangle \neq L(d|x)$.[93]

ii. Be valid (the exclusion restriction), so that *z* only impacts the outcome variable *y* indirectly through the treatment variable *d* and not directly (see figure 4). That is to say that *z* does not belong in the structural equation for *y*. COR(Z,ε)=0 or $L(v_0|x,z) = L(v_0|x)$.

Following Wooldridge (2002), we expand the expression for our observed outcome to include the unobservable influences on individual-specific outcomes, e_i. This is possible if we recall that $v_i = g_i(x) + e_i$ and assume that $E(v_1|z,x) = E(v_1|x)$, which mean that the expectation of individual-specific outcomes are independent of the instrumental variables.

$$y = dy_1 = du_1 + d(g_1(x)) + d(e_1)$$

Making the functional form assumption that $g_1(x)$ is linear, we obtain the regression:

$$\text{(F)} \quad y_i = \alpha d_i + \beta_x d(x - \overline{x}) + de_1$$ [94]

where the final term is the error and $\alpha = u_1$ is the estimator of the ATE. This regression still cannot be estimated by OLS because *d* and the error are correlated. However, it can be

[91] Moffitt (1996), p.3.
[92] Following Lee (2005), p.130.
[93] L(x) represents the linear projection of x.
[94] A useful Stata command for this sort of model is **treatreg**, but the command only facilitates a single endogenous variable so it cannot be applied in this case.

estimated by instrumental variables techniques using instruments for *d* and $d(x-\bar{x})$ if we assume that $E\left[d(e_1)\middle|x,z\right]=E\left[d(e_1)\right]$. Then the mean of the error term does not depend on *x* and *z*, permitting any instruments that are functions of *x* and *z* to be used as instruments. The assumption guarantees that there is zero covariance between treatment and the unobservable individual-specific treatment effect e_1. Since we do not assume $e_1=0$, we are permitting heterogeneity in the *effects* of treatment, i.e. there can be heterogeneity across individuals in the gains from adoption on the basis of unobservables and not merely the covariates *x*. This was a primary aim of our analysis. Unfortunately, the assumption requires that these unobservable individual-specific components are uncorrelated with treatment, and therefore cannot be the basis on which selection mechanisms operate. Heckman and Vytlacil (2005) clarify the distinction to show that while “heterogeneity in responses” is permitted, “heterogeneity in choices” is not.[95]

While this appears a major restriction to the analysis because we would expect individuals to possess and act on more information than is available to the researcher, in our own context the assumption may not be too limiting, for the following reason.

Our information threshold for treatment is very low – simply having heard of NERICAs’ existence is sufficient – and selection mechanisms determining this awareness are arguably based on a limited number of factors affecting the *general* gains from adopting new technologies. It is less likely that selection is based on the detailed farmer-specific gains from adoption because awareness necessarily precedes any detailed understanding of the characteristics of NERICAs or the conditions under which they grow most effectively. Farmers themselves are unlikely to be aware of the relationship between their own characteristics (which are unobservable to ourselves) and the potential outcomes from adopting NERICAs. This information, as Section 2 argued, is acquired by more *innovation-specific* cost-benefit analysis about the dedication of resources to information-acquisition. At

[95] Heckman and Vytlacil (2005), p.669.

the initial awareness level, the factors likely to bring a farmer into awareness of NERICAs probably constitute more general characteristics such as farm size, use of existing technologies, contact with extension agencies and involvement in social networks that determine their general inquisitiveness about new technologies.

Much of the treatment effect literature assumes that biases are created by self-selection into treatment based on the subsequent gains, what I term "*active* selection". However, the mechanisms of selection may be broader, including for example involvement in social networks where information tends to be disseminated; this may be termed "*passive* selection". If such involvement did not have any bearing on the gains from adoption, it would not bias the treatment effect estimates obtained. But passive selection need not be neutral; if those more deeply involved in social networks also happen to have larger farms or higher incomes, then they may experience greater gains from adoption and be more likely to adopt *once treated*. In such cases, the *motivation* of gains is not what drives selection into treatment (i.e. membership of a social network), but biases in the estimation of treatment effects can still result. Of course, if there are gains to be made from membership of a social network we would not expect people to pass-up these opportunities. However, these gains are likely to be apparent only in a quite general form, for example more rapid awareness of new technologies. Farmers are unlikely to be able to join a social network because they anticipate a *specific* technology which will be of great benefit to themselves *before* they are aware of that technology.[96] As such, the drivers of selection may be more generic and are more likely to be observable to the researcher.[97]

[96] Moreover, it remains the case that unobservable characteristics affect subsequent adoption decisions via the decision to acquire further information and reduce the perceived risk around the innovation. But the sole requirement for the validity of the methodology is that unobservables do not affect the likelihood of *treatment* in the form of awareness of NERICAs.

[97] Heckman (1997) makes a compelling critique that most applications of IV for treatment effects are extremely sensitive to assumptions about the way people process information. This example identifies an important class of circumstances relating to the initial awareness of innovations where these assumptions about information processing hold. Indeed, it may be valuable for future studies to distinguish the generic information-processing related to the full spectrum of new innovation availability from innovation-specific information over which individuals make much more calculated information-acquisition decisions and use a much greater wealth of information about their personal characteristics and circumstances. In the former case, IV methods may still be able to estimate the ATE, while in the former Heckman is correct that the only plausible estimator is the local average treatment effect (LATE), which requires careful interpretation.

Of course, it should not be assumed that data has been collected on every aspect of behaviour even if it is readily observable – it was noted that information on important and observable factors is missing from our dataset. We should, then, remain sceptical that the IV approach is robust in this paper. Yet, as a general lesson for future studies, to the extent that passive selection is common in the adoption context, then the assumptions in the IV model are more likely to be met.

Assuming the assumptions are met, Wooldridge (2002) illustrates the following consistent estimation procedure:

1. Estimate $\hat{G}_i = P(d_i = 1 | x_i, z_i)$, the fitted values of the propensity score, using a probit model. The consistency of the method does not depend on the correct specification of this model;

2. Use $\hat{G}_i$ and $\hat{G}_i(x - \bar{x})$ as efficient instruments for d_i and $d_i(x - \bar{x})$ in the IV estimation of (F);

The coefficient α in (F) is then the ATE.

One difficulty with the above process is that error terms are likely to be heteroskedastic.[98] For this reason, we have reported robust standard errors in Section 4.

There are two principal limitations of IV estimators. First, there is no guarantee that IV estimators are superior to OLS estimators; if the IV is weak, it may not be a worthwhile approach.[99] Tests for the strength of the instruments are discussed in the main text. Second, in this as in other studies based on observational data, good instrumental variables are elusive. In experiments and randomized trials, a natural and effective instrument is treatment

[98] Wooldridge (2002), p.628.
[99] Moffitt (1996).

assignment or eligibility for treatment, but in our case these are not available.[100] Close examination of proposed instruments is therefore essential.

[100] Moffitt (1996). Wooldridge (2002), p.622 also stresses that random assignment of eligibility does not guarantee fulfilment of the IV conditions, which still need to be examined.

Appendix 2 Testing the assumptions

A2.1 Conditional independence

Imbens (2003) shows that a formal test of conditional independence is not possible. It would be straightforward to test whether *observed* outcomes were independent of treatment, $E(y|x,d)=E(y|x)$, by a simple endogeneity test, eg. the Durbin-Wu-Hausman test. But the conditional independence assumption refers to the set of *hypothetical* outcomes, $E(y_1|x,d)=E(y_1|x)$. Since y_1 is partially unobserved, this is untestable. The data can never reject this assumption.

Nevertheless, indirect tests have been suggested.[101] One which can be applied to our data is to test whether variables known to be invariant to treatment are revealed as such by a simple regression on the treatment variable and other covariates. If the null hypothesis of no effect is not rejected, we have greater reason to believe that conditional independence holds and that treatment status is effectively exogenous given covariates.

The test was carried out for the following variables which arguably should be invariant to treatment: household size, age, sex and village number. The only variable that rejected the null hypothesis of no relationship is sex for Côte d'Ivoire, which was significant at the 1% level. Given this mixed evidence and general scepticism about the plausibility of conditional independence, I also utilise an IV estimator that avoids the assumption.

A2.2 Linear regression validity

The fully parametric models we employ, (C), (D) and (F), make strong linearity assumptions about the effect of covariates on outcomes. Imbens and Wooldridge (2008) suggest that even if this is true locally it may not be true globally. They present a rule-of-thumb based on the normalized difference in covariate means and argue that above 0.25, linear specifications may be inappropriate.

$$\Delta x = \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{S_1^2 + S_0^2}}$$

The threshold is passed for 30% of the Côte d'Ivoire variables and 10% of the Nigerian variables, suggesting the parametric methods may be less appropriate for application to Côte d'Ivoire.

A2.3 Relevance of Instruments

The strength of the correlation between the instruments *z* and endogenous variables determines the efficiency of IV estimates. The standard test for weak identification is the rank test for a matrix. In the presence of multiple instruments, this assesses whether the instruments may all be highly correlated with only one of the endogenous variables, but very weakly correlated with another endogenous variable, in which case the identifying power of the IV approach would be limited. The Stata command **ivreg2** reports an LM test under the null hypothesis of under-identification, i.e. of less than full rank. Rejecting this null hypothesis supports model identification.[102]

A2.4 Validity of Instruments

Tests for whether the instrument is correlated with the error term generally rely on the model being over-identified (having more instruments than endogenous variables). Wooldridge's (2002) method prohibits such a test because the use of fitted values from a first-stage probit

[101] Imbens and Wooldridge (2008), p.43.
[102] Hall et al (1996) argue that a rejection of the null hypothesis does not guarantee that weak identification is not a problem.

means the model is exactly identified. However, Nichols (2007) suggests a solution. Noting that if $E(z'\varepsilon)=0$ and z^* includes squares and cross-products in z, then typically $E(z^{*\prime}\varepsilon)=0$. The additional instruments achieve over-identification. A variation on this approach is adopted here, with the z^{IV} variables being used directly as instruments in addition to their indirect use in the calculation of $\hat{G}_i$. **Ivreg2** reports the Sargan-Hansen test, with the null hypothesis that all the instruments are uncorrelated with the errors. In this case, z is correctly excluded from the structural equation explaining adoption and the instruments are valid.[103]

Table 6 contains the results for the two instrumental variable tests. Rejection of the validity assumption in Côte d'Ivoire is a cause for concern, but appears to be highly sensitive to the extra instruments included to achieve identification. Since these are only indicative tests, the methodology is retained but it may be appropriate to treat the IV estimates with caution.

Table 7 Instrumental variables tests

Assumption	Test	P-value: Côte d'Ivoire	P-value: Nigeria
Relevance	Anderson canonical correlation LM test	0.0004 Assumption supported	0.0000 Assumption supported
Validity	Sargan Statistic	Exactly identified	Exactly identified
	Sargan Statistic (using additional instruments)	0.0286 Assumption rejected	0.3002 Assumption supported

[103] Again, interpretation of the test requires caution because Baum et al (2007) demonstrate that a failure to reject the null hypothesis is neither a necessary nor sufficient condition for validity.

Appendix 3 Results for the intensity of adoption

An extension of the analysis considered the impact of treatment on the *intensity* of adoption in Nigeria using plot-level data on crop variety. The same methodology was employed but using a continuous outcome variable relating to the percentage of rice area devoted to NERICAs. Notably, this mitigates the parametric difficulties described earlier. Estimates (A)-(E) consistently estimated the ATE and ATT at 23-24%, suggesting those first aware of NERICAs were no more or less intensive adopters. The IV estimator (F) estimated the ATE at 30% and the ATT at 26%.

However, in the same way that there can be non-exposure bias, there can also be non-adoption bias too; farmers who are aware of NERICAs but are prevented from adopting due to the barriers identified in the following section are included, creating a large number of zero-entries and biasing downwards the estimates; we do not observe their intensity of adoption *if* they had chosen to adopt.[104] If we believe that the decision to adopt and how much to adopt are separable decisions, an alternative method is to condition the estimators on adoption status as well as treatment status. As an indication, method (A) modified in this way gives an ATE estimate of 56%. Thus, if farmers have been both exposed *and* chosen to adopt, over half of the area of their rice land is used to grow NERICAs.

[104] The estimates are not as misleading as those vulnerable to non-exposure bias because it is easier to construe non-adoption as a choice while non-exposure is an irresolvable constraint. An estimate which shows the average of adoption rates *allowing* the choice not to adopt (24%) can be equally as informative as an estimate of the impact which only considers those who do adopt (56%); which is useful depends on the context.

Bibliography

1. Abdulai, A., Monnin, P., and Gerber, J. (2008), 'Joint estimation of information acquisition and adoption of new technologies under uncertainty', *Journal of International Development*, 20, pp.437-451.
2. Adekambi S.A., Diagne, A. and Biaou, G. (2006), 'Impact de l'adoption des variétés NERICAs sur la scolarisation des enfants au Bénin'. Mimeo, Centre du riz pour l'Afrique (ADRAO), Cotonou, Bénin.
3. Africa Rice Congress (2006), 'Beyond the First Generation of NERICAs in Africa: Paradigms and Partnerships for the Next Decade'.
4. Agboh-Noameshie A, Kinkingninhoun, F.M.M. and Diagne, A. (2007), 'Gendered Impact of NERICA adoption on farmers' production and income in Central Benin'. Paper presented at the 2nd African Association of Agricultural Economists Conference, Accra, Ghana, 20-22 August.
5. Angrist, J.D. (undated), 'Treatment effects', MIT. Available at http://econ-www.mit.edu/files/32.
6. Asfaw, A. and Admassie, A. (2003), 'The role of education on the adoption of chemical fertiliser under different socioeconomic environments in Ethiopia', *Agricultural Economics*, 30, pp.215-228.
7. Bandiera, O. and Rasul, I. (2002), 'Social Networks and Technology Adoption in Northern Mozambique', *CEPR Discussion Paper*, No.3341, April.
8. Barnow, B.S., Cain, G.G., and Goldberger, A.S. (1980), 'Issues in the Analysis of Selectivity Bias' in Stormsdorfer, G. and Farkas, G., *Evaluation Studies Review Analysis*, Vol.5.
9. Besley, T. and Case, A. (1993), 'Modelling Technology Adoption in Developing Countries', *American Economic Review Papers and Proceedings*, 83:2, pp.396-402.
10. Blundell, R. and Costa Dias, M. (2002), 'Alternative Approaches to Evaluation in Empirical Microeconomics', *Institute for Fiscal Studies Working Paper,* CWP10/12, March.
11. Caudill, S.B. (1988), 'An Advantage of the Linear Probability Model over Probit or Logit', *Oxford Bulletin of Economics and Statistics*, Department of Economics, University of Oxford, 50:4, November, pp.425-427.
12. Conley, T.G. and Udry, C.R. (2003). 'Learning about a New Technology: Pineapple in Ghana'. Available at www.econ.yale.edu/growth_pdf/cdp817revised.pdf.
13. Diagne, A. and Demont, M. (2007), 'Taking a new look at empirical models of adoption: average treatment effect estimation of adoption rates and their determinants', *Agricultural Economics*, 37, pp.201-210.
14. Diagne, A. (2004), 'Rate and determinants of adoption of a technology with partial diffusion: The case of diffusion and adoption of NERICA rice varieties in Côte d'Ivoire, September, WARDA.
15. Diagne, A. (2006), 'Diffusion and Adoption of NERICA Rice Varieties in Côte d'Ivoire', *The Developing Economies*, XLIV-2, June, pp.208-231.
16. Dimara, E. and Skuras, D., (2004) 'Adoption of agricultural innovations as a two-stage partial observability process', *Agricultural Economics*, 28, pp.187-196.

17. Doss, C.R. (2006), ‘Analyzing Technology Adoption using Microstudies: Limitations, Challenges and Opportunities for Improvement’, *Agricultural Economics*, 34, pp.207-219.

18. Evenson, R.E. and Gollin, D. (2003), ‘Assessing the Impact of the Green Revolution’, *Science*, 300:758.

19. Evenson, R.E. and Mwabu, G. (1998), ‘The Effects of Agricultural Extensionon Farm Yields in Kenya’, *Yale University Economic Growth Centre Discussion Paper,* No.798.

20. FAOSTAT (2008), Available at http://faostat.fao.org.

21. Feder, G. and Slade, R. (1984), ‘The Acquisition of Information and the Adoption of New Technology’, *American Journal of Agricultural Economics*, 66, pp.312-330.

22. Feder, G., Just, R.E. and Zilberman, D. (1985), ‘Adoption of Agricultural Innovations in Developing Countries: A Survey’, *Economic Development and Cultural Change*, 33:2, January, pp.255-298.

23. Foster, A.D. and Rosenzweig, M. (1995), ‘Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture’, *Journal of Political Economy*, 103:6, December, pp.1176-1209.

24. Freeland, R.K. and McCabe, B.P.M. (2004), ‘Forecasting discrete valued low count time series’, *International Journal of Forecasting*, 20, pp.427-434.

25. Goldman, A. and Smith, J. (1995), ‘Agricultural revolutions in India and Northern Nigeria: Exploring the Nature of Green Revolutions’, *World Development*, 23:2, pp.243-263.

26. Goldstein, M. and Udry, C. (2005), ‘The Profits of Power: Land Rights and Agricultural Investment in Ghana’, *Yale University Economic Growth Centre Discussion Paper*, No.929.

27. Greene, W.H. (2003), *Econometric Analysis*, 5th Edition, Prentice Hall, New York.

28. Griliches, Z. (1957), ‘Hybrid Corn: An Exploration in the Economics of Technological Change’ in *Technology, Education and Productivity: Early Papers with Notes to Subsequent Literature (1988)*, New York, Basil Blackwell, pp. 27-52.

29. Gujarati, D.N. (2003), *Basic Econometrics*, New York, McGraw-Hill.

30. Hall, A.R., Rudebusch, G.D. and Wilcox, D.W., (1996), ‘Judging Instrument Relevance in Instrumental Variables Estimation’, *International Economic Review*, 37:2, pp. 283-298.

31. Härdle, W. and Linton, O. (1994), ‘Applied Nonparametric Methods’, *Cowles Foundation Discussion Papers* 1069, Cowles Foundation, Yale University.

32. Heckman, J. (1997), ‘Instrumental Variables: A Study of Implicit Behavioural Assumptions used in making Program Evaluations’, *Journal of Human Resources*, 32:3, Summer, pp.441-462.

33. Heckman, J.J. and Vytlacil, E. (2001), ‘Policy-Relevant Treatment Effects’, *American Economic Review Papers and Proceedings*, 91:2, May, pp.107-111.

34. Heckman, J.J. and Vytlacil, E. (2005), ‘Structural Equations, Treatment Effects and Econometric Policy Evaluation’, *Econometrica*, 73:3, pp.669-738.

35. Holloway, G., Shankar, B. and Rahman, S. (2002), ‘Bayesian Spatial Probit Estimation: A primer and an application to HYV rice adoption’, *Agricultural Economics*, 27, pp.383-402.

36. Holmén, H. (2003), ‘A Green Revolution for Africa: Does it need to be so controversial?’ *International Centre for Economic Research Working Paper*, No.4, January.

37. Imbens, G.M. and Wooldridge, J.M. (2008), 'Recent Developments in the Econometrics of Policy Evaluation', *NBER Working Paper*, 14251,

38. Imbens, G.W. (2003), 'Semiparametric Estimation of Average Treatment Effects under Exogeneity: A Review', *NBER Technical Working Paper*, No.294, October.

39. Imbens, G.W. (2004), 'Nonparametric estimation of average treatment effects under exogeneity: a review', *Review of Economics and Statistics*, 86:1, February, pp.4-29.

40. Imbens, G.W. and Angrist, J.D. (1994), 'Identification and Estimation of Local Average Treatment Effects', *Econometrica*, 62:2, March, pp.467-475.

41. Kaneda, C (2006), 'NERICAs: Their Development, Dissemination and Expectation for African Green Revolution', *Japanese Journal of Tropical Agriculture,* 50:5, pp.293-299.

42. Kijimi, Y., Sserunkuuma, D. And Otsuka, K. (2006), 'How Revolutionary is the Nerica Revolution? Evidence from Uganda', *The Developing Economies*, XLIV-2, June, pp.252-267.

43. Lee, M-J. (2005), *Micro-Econometrics for Policy, Program and Treatment Effects*, Oxford University Press, New York.

44. Maddala, G.S., (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge, Cambridge University Press.

45. Moffitt, R.A. (1996), 'Identification of Causal Effects using Instrumental Variables: Comment', *Journal of the American Statistical Association*, 91:434, June, pp.462-465.

46. Mosley, P. (2002), 'The African Green Revolution as a Pro-Poor Policy Instrument', *Journal of International Development*, 14, pp.695-724.

47. Munshi, K., (2004) 'Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution', *Journal of Development Economics*, 73, pp.185-213.

48. National Bank for Agriculture and Rural Development of India (undated), '9. Adoption of modern varieties and input uses'. Available at *www.nabard.org/publications/dynamic/adoption_eng.pdf.*

49. New York Times, "In Africa, Prosperity from Seeds falls Short", http://www.nytimes.com/2007/10/10/world/africa/10rice.html?_r=2&hp&oref=slogin&oref=slogin

50. Nichols, A. (2007), 'Intro to Stata'. Available at http://www.google.co.uk/url?sa=t&source=web&ct=res&cd=1&url=http%3A%2F%2Fpped.org%2Fstata%2FIntroSt.pdf&ei=15DLSLLFBoHm1ga35pCvDQ&usg=AFQjCNGQdprps_9Ex-8r0O_XkAAu3W41Ew&sig2=9ECHknvBiSyx-27MeAP_yA.

51. Norman, J.C. and Kebe, B. (2007), 'African smallholder farmers: rice production and sustainable livelihoods', *International Rice Commission.*

52. Nwanze, K.F., Mohapatra, S. and Kouka, P-J. (undated), 'In Search of the Right Solutions for Africa's Development', Africa Rice Center. Available at http://nabc.cals.cornell.edu/pubs/nabc_16/talks/Nwanze.pdf

53. Pannell, D.J., Marshall, G.R., Barr, N., Curtis, A., Vanclay, F. and Wilkinson, R. (2006),'Understanding and promoting adoption of conservation practices by rural landholders', *Australian Journal of Experimental Agriculture*, 46, pp.1407-1424.

54. Rosenbaum, P.R. and Rubin, D.B. (1983), 'The Central Role of the Propensity Score in Observational Studies for Causal Effects', *Biometrika*, 70:1, pp.41-55.

55. Rubin, D.B. (1978), 'Bayesian inference for causal effects: the role of randomization', *Annals of Statistics*, 7, pp.34-58.

56. Saha, A., Love, A.H. and Schwart, R. (1994), 'Adoption of Emerging Technology under Output Uncertainty', *American Journal of Agricultural Economics*, 76:4, pp.836-846.

57. Scola, N. (2007), 'NERICAs and Africa's Green Revolution', http://www.nancyscola.com/2007/10/nericas_and_africas_green_revo.html.

58. Skinner, J. and Staiger, D., (2004) 'Technology Adoption from Hybrid Corn to Beta Blockers', November.

59. Somado, E.A., Guei, R.G. and Keya, S.O. (2008), 'NERICA: the New Rice for Africa – a Compendium', WARDA.

60. Spencer, D., Dorward, A., Abalu, G., Philip, D. and Ogungbile, D. (2006), 'Evaluation of Adoption of NERICA and other Improved Upland Rice Varieties following Varietal Promotion Activities in Nigeria: A Study for the Gatsby and Rockefeller Foundations: Final Report', January.

61. WARDA (2002), 'Annual Report 2001-02'.

62. WARDA (2003), 'Annual report 2002-03'.

63. WARDA (2006), 'Experts Underline Four Prerequisites for Rice-Based Green Revolution in Africa Policy, Capacity Building, Technologies and Infrastructure', August. Available at http://www.warda.org/warda/newsrel-congress1-aug06.asp.

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

65. Wooldridge, J.M. (2006), *Introductory Economics: A Modern Approach*, Thomson South-Western.

66. World Development Report (2008), *Agriculture for Development*, International Bank for Reconstruction and Development.

67. Zorn, C. (2005), 'A Solution to Separation in Binary Response Models', *Political Analysis*, 13:2.

Oxera
Park Central
40/41 Park End Street
Oxford OX1 1JD
United Kingdom

Tel: +44 (0) 1865 253 000
Fax: +44 (0) 1865 251 172

www.oxera.com