THE IMPORTANCE OF LEARNING IN THE ADOPTION OF HIGH-YIELDING VARIETY SEEDS

LISA A. CAMERON

To date, due to the lack of panel data, most micro-level empirical studies of technology adoption have used cross-sectional data. These studies cannot examine the dynamic processes of adoption such as learning. This article uses panel data to study the adoption of a new high-yielding variety seed. First, it establishes that learning is an important variable in the adoption process. Second, it establishes that cross-sectional estimates of a dynamic process are biased but that the extent of this bias may be small. Third, it illustrates the econometric methods needed to estimate a dynamic model when controlling for unobserved household heterogeneity.

Key words: learning, panel data, technology adoption.

High-yielding variety seeds have played a vital role in enabling poor farmers in the lesser-developed regions of the world to increase dramatically the size of their harvests and hence to increase their living standards. Therefore, the process by which these seeds are adopted has long been an area of research for development economists. Adoption is essentially a dynamic process which involves learning about the new technology over time. Although the dynamic aspect of adoption has been amply recognized in the theoretical literature (O'Mara; Lindner, Fischer, and Pardey; Fischer and Lindner; Lindner and Fischer) due to the scarcity of micro-level data across time (panel data), almost all previous micro-level empirical studies of adoption have used cross-sectional data and have thus been unable to explore the dynamic nature of the process.[1] Exceptions to this are Besley and Case (1993b), and Foster and Rosenzweig, who used panel data and established the importance of learning. Besley and Case (1993b) model the farmers as being uncertain about the profitability of the new seed relative to the old seed. They simulate the subgame-perfect number of plots to be sown to the new seed (given that farmers learn about the new seed's profitability through experience) and compare this with the pattern found in their data. In contrast, Foster and Rosenzweig model the optimal input use as being unknown and stochastic. Farmers learn about the optimal combination through their experience and the experience of their neighbors. Foster and Rosenzweig test their model on a three-year panel of data from twenty-five villages in India. They conclude that learning from own experience and learning from neighbors' experience are both determinants of adoption. The finding that learning is an important determinant of adoption is in contrast to earlier work by McGuirk and Mundlak which suggested that adoption was constrained by insufficient irrigation and fertilizer, not by insufficient information. These contrasting findings have divergent policy implications. Evidence of the importance of learning in the adoption of new technologies provides support for policy initiatives such as educational support facilities for the technologies as opposed to irrigation and input subsidy schemes.

This article's objectives are threefold. It aims to provide further evidence of the importance of learning using empirical methods different from Besley and Case (1993b) and a learning model and data set different from Foster and Rosenzweig. In addition to this, it examines the extent of the bias in cross-sectional estimates that ignore dynamic processes such as learning.[2] Finally, it illustrates the

Lisa A. Cameron is a lecturer in the Department of Economics, University of Melbourne, Australia.

The author is thankful to Angus Deaton, Christina Paxson, and Tim Besley for offering many helpful comments. Any remaining errors are the author's.

[1] For a comprehensive survey of the literature on the adoption of agricultural technology see Feder, Just, and Zilberman. For examples of studies that have examined the dynamic pattern of adoption at the aggregate level using time-series data see Griliches and Dixon.

[2] Rahm and Huffman, Lindner, and Besley and Case (1993a) have commented on the problems inherent in cross-sectional models that seek to explain a dynamic process.

Amer. J. Agr. Econ. 81 (February 1999): 83–94

econometric methods needed to estimate a dynamic model and control for unobserved household heterogeneity (an issue previously addressed in Arellano and Bond).

This study uses ICRISAT panel data on thirty-one households in the village of Kanzara in Maharashtra, India, to study the dynamic process of learning in the adoption of a new high-yielding variety (HYV) cotton seed. The data covers the period 1975–84; the new seed was introduced in 1980. The remainder of the article is structured in the following way. First, a simple theoretical model of learning is developed. Farmers are modeled as being uncertain about the profitability of the new seed relative to the old seed and, they learn about this through their own experience with the seeds.[3] Second, it is established that if the theoretical model is true, then cross-sectional estimates of the structural variables are biased. Third, the nontrivial difficulties of testing a model which has a lagged dependent variable when controlling for household heterogeneity are discussed. This section also discusses the consequences of the possible endogeneity of some of the explanatory variables. Fourth, the learning model is estimated on the panel data and the panel estimates are compared with estimates from the cross-sectional data (without dynamic terms) to quantify the extent of the cross-sectional bias. Conclusions are then drawn in the final section.

A Simple Model of Learning

A very simple model of learning is developed and tested in this article. The model can be viewed as the first step in determining the importance of learning in the adoption process and a launching pad for testing more complex and realistic learning models in future work. The adoption decision is modeled as the decision between planting a plot of land to the new HYV cotton seed and planting it to the traditional cotton seed.[4] Assume that farmers aim to maximize expected profits in each period.[5] The farmer is uncertain of the profitability of the new seed relative to the old seed and learns about this over time from his own experience with the new seed. The profitability of household i's plot j in period t depends on farm characteristics such as farm size, x_{it}, and plot characteristics such as soil type, w_{ijt}. When forming expectations of the new seed's profitability relative to the old seed, the farmer therefore takes into account the above factors and augments this with his or her stock of seed-specific knowledge, z_{it}.

We can therefore write

$$\text{(1)} \qquad \mathrm{E}(\pi_{ijt}^{hyv} - \pi_{ijt}^{o}) = f(x_{it}, w_{ijt}, z_{it})$$

where $\mathrm{E}(\pi_{ijt}^{hyv} - \pi_{ijt}^{o})$ is household i's expectation of the profit differential between the new seed and the old seed on plot j in year t.

If the variable y_{ijt} reflects the adoption decision and equals 1 if the new seed was sown by household i in plot j in period t and otherwise equals zero, we can write

$$\text{(2)} \qquad \begin{aligned} y_{ijt} &= 1 \quad \text{if } \mathrm{E}(\pi_{ijt}^{hyv} - \pi_{ijt}^{o}) > 0 \\ &= 0 \quad \text{if } \mathrm{E}(\pi_{ijt}^{hyv} - \pi_{ijt}^{o}) \leq 0. \end{aligned}$$

Hence the planting decision is determined by farm and plot characteristics and the farmer's knowledge of the new seed:

$$\text{(3)} \qquad y_{ijt} = g(x_{it}, w_{ijt}, z_{it}).$$

In order to estimate the adoption decision it will thus be necessary to obtain an empirical measure (z_{it}') summarizing the knowledge gained from previous experience. The average of all profit differentials that the farmer has experienced in previous years is used, that is, the difference between the profitability per acre of the HYV seed and the traditional seed averaged over all previous periods in which the new seed was used:[6]

[3] The data and theoretical model are those of Besley and Case (1993b), but the empirical technique that establishes the importance of learning will differ. The choice of model was in part determined by the need for comparability with previous cross-sectional studies. An obvious alternative model is Foster and Rosenzweig's model of learning about optimal input choices. However, the main advantage of that model is that it allows for the estimation of learning from neighbor's experience. The ICRISAT data preclude such testing because they only cover one village and lack information on the situation of plots within the village. Hence, there is no compelling reason for preferring Foster and Rosenzweig's model over the one chosen.

[4] The choice to plant the land to cotton in the first place is ignored in this article. See footnote 17 for a further discussion of this point.

[5] This assumption ignores the role of strategic experimentation in the adoption process. It does not allow farmers to forego current profits in order to learn about a new seed and so to be in a position to possibly more than recoup this loss of profits in succeeding periods. Allowing for strategic experimentation complicates the problem significantly and is an area for further research. See Besley and Case (1993b) for an exposition that explicitly models such experimentation.

[6] Note that although the stock of knowledge, z_{it}, is assumed not to decrease over time, z_{it}' is the product of that knowledge and thus may increase or decrease over time.

(4) $z'_{it} = \sum_{n=1}^{t-1} [y_{i,t-n}(\pi^{hyv}_{i,t-n} - \pi^{o}_{i,t-n})/N_{it}]$

where N_{it} = the number of years that household i had planted the new seed and $y_{it} = 1$ if $y_{ijt} = 1$ for any j.

One can think of the household updating its knowledge based on the new observation but still placing some weight on observations from earlier periods. The farmer's knowledge is thus updated after every season in response to the differential profitability experienced in the previous period. If the new seed is not used in a period then no learning takes place in that period; thus, the farmer's stock of knowledge remains constant as does the summary measure, z'_{it}.[7] The knowledge from the farmer's experience in past periods is combined with plot and household specific characteristics to produce an expectation of the profit differential and an adoption decision for each plot.

If we assume that the relationship between the variables in equation (3) is linear, then we can write

$$(5) \quad y_{ijt} = b_0 + b_1 x_{it} + b_2 w_{ijt} + b_3 \sum_{n=1}^{t-1} [y_{i,t-n}(\pi^{hyv}_{i,t-n} - \pi^{o}_{i,t-n})/N_{it}] + e_{ijt}$$

where e_{ijt} is an error term arising from differences between the true knowledge variable z_{it} and the simplistic empirical summary measure z'_{it}. Equation (5) is the relationship to be estimated. In practice e_{ijt} will also reflect the effect of any unobservable characteristics in the $\mathbf{x}_{it}$ and $\mathbf{w}_{ijt}$ vectors in the empirical estimation. It is modeled as being the sum of a household specific effect and a random component.

Limitations of Cross-Sectional Data

Since cross-sectional data do not include leads and lags of variables, it is impossible to estimate a dynamic model as shown in equation (5). Instead, most cross-sectional studies involve regressing a measure of technology adoption on contemporaneous household and plot characteristics as shown in equation (6):

$$(6) \quad y_{ij} = q_o + q_1 x_i + q_2 w_{ij} + e_{ij}.$$

If the true model involves the dynamic learning term z'_{it}, then such a cross-sectional model will only yield unbiased estimates of the underlying coefficients, b_0, b_1, and b_2, when learning from own experience does not take place in the period of estimation (the average profit differential term will have a coefficient of zero because it imparts no new information and so does not affect the adoption decision) or when the learning term is orthogonal to the explanatory variables. Learning will no longer be taking place if the seed has been available long enough for all learning to have occurred. If learning is still taking place and the learning term is correlated with the explanatory variables, the cross-sectional estimates will suffer from omitted variable bias. Cross-sectional regressions can only be safely used to investigate the relationship between final seed usage and household and other characteristics once the adoption process is complete. Biased estimates can lead to incorrect policy implications being drawn from the estimated relationship between adoption and various household and plot characteristics. For instance, bias in the coefficient on soil type could lead to heavy introduction of the new seed in areas that are only marginally profitable. The extent of the bias in the cross-sectional estimates will be determined by the strength of the correlation between the learning term and the other variables.

Another limitation of cross-sectional data is that it cannot control for unobserved household heterogeneity.[8] This introduces further omitted variable bias if the unobserved heterogeneity is correlated with the explanatory variables.

Estimating the Model on Panel Data

Unlike cross-sectional data, panel data can produce consistent estimates of the underlying parameters in equation (5).

[7] Although equation (4) is not explicitly derived from a Bayesian learning model, it is in the spirit of Bayesian learning.

[8] If the cross-sectional data are at the plot level then it is possible to include household fixed effects, but doing so involves dropping all household variables and all households that sow only one plot in that year. The resultant estimates are thus of limited use.

Empirically Defining the Dynamic Learning Term

The first question to be addressed when estimating the learning model is how to define the learning term. As mentioned above, this study uses the average profit differential between the new and the old seed that has been experienced by the farmer as the dynamic learning term. This variable has a number of shortfalls but is the best available variable given the data limitations. Farmers may actually form expectations of the profitability of the new seed relative to the old seed for each plot, in which case one may want to use a plot-level learning variable. However, plots cannot be tracked over time in the ICRISAT data so this is not an option. Hence, learning is modeled at the household level. Using household-level profits does confer the advantage of being able to compare household profits from the new seed with those of the old seed. There is no obvious counterfactual with which to compare profitability of the new seed at the plot level.

Another potential drawback of the proposed learning term is that one may want to allow for the possibility that learning from own experience may be more important in earlier years than in later years when the information is more widely dispersed. There is the prospect of including the profit differentials experienced in each of the previous years separately and allowing their coefficients to differ. This is not feasible when using a small data set like ICRISAT. Including an additional lag involves dropping a year of data. Using the average of the profit differentials implicitly weights the differentials so that the earliest observations have the largest impact. Subsequent observations result in smaller and smaller changes to the farmer's best estimate of the profit differential.

A possible alternative to the chosen variable is to only use a one-period lagged profit differential to reflect the farmer's knowledge. A major drawback of this measure though is that it makes the implausible assumption that all earlier experience with the new seed is irrelevant given knowledge of the previous period's profits. Nevertheless, a model using the one-period lagged profit differential was estimated. The results were similar to those reported below, but the average profit differential was a better fit for the data.[9]

Potentially the biggest drawback from the learning-from-own-experience model above is that it fails to take into account learning from other sources such as neighbors. The ICRISAT data provides no information on the geographic situation of farms and so immediate neighbors cannot be identified. With data originating from only one village, village-level learning cannot be explicitly modeled. Any village-level learning variable will act as a year dummy and pick up the effect of village-level learning and any other factors influencing the village as a whole, such as weather shocks. Hence, the year dummies in the regressions below capture village-level learning but cannot separate its effect from a variety of other factors.

One can think of the farmers supplementing learning from sources at the village level with learning from own experience. The farmer's own experience is more likely to provide specific information on the productivity of the new seed on his own plots. Village-level learning may explain why some households are late adopters and plant their first plots to the new seed only in the later years of the sample. A model based solely on learning from own experience has difficulty explaining this phenomenon.

Note also that, as is the case in any empirical work, it is possible that a correlation between adoption and the average profit differential reflects an underlying process other than learning.

Controlling for Unobserved Household Heterogeneity

Unobserved household heterogeneity biases the coefficient on any variable with which it is correlated. An additional advantage of panel data is that it becomes possible to control for unobserved household heterogeneity. Consider the case where some farmers have an inherent ability to be more profitable with the new seed than other farmers and that this ability is not specifically reported in the data nor proxied by any of the reported variables. This may arise if the data do not reflect land quality, unobserved farmer skills, or initial beliefs of crop profitability. The average profit differential learning term and the adoption decision are then both partially determined by the heterogeneity in household profitability. Consequently, the profit differential learning term can be statistically significant due to

[9] Results obtained when using the one-period lagged profit differential are available from the author upon request.

household heterogeneity even when learning is absent.

The consequence of unobserved household heterogeneity is that the error term in equation (5) has a household-specific component and so is not independently distributed. This component can normally be modeled as a random effect or as a fixed household-specific constant. The random effects method, however, is not valid in this context because it assumes that the household effects are uncorrelated with the explanatory variables. This assumption is violated by a lagged dependent variable (or a variable related to a lagged dependent variable, such as the average profit differential) because it is correlated with the unobserved heterogeneity and, hence, with the random effects.[10] Therefore, the error will be modeled as a household-specific component using household fixed-effect dummy variables.[11] The fixed-effects model corresponding to equation (5) is depicted in equation (7):

$$(7) \quad y_{ijt} = \alpha_i + \beta_1 X_{it} + \beta_2 w_{ijt} + \beta_3 \sum_{n=1}^{t-1} [y_{i,t-n}(\pi^{hyv}_{it-n} - \pi^{o}_{it-n})/N_{it}] + e_{ijt}$$

where X_{it} includes only those x_{it}'s which are time-varying to avoid collinearity between the household-level variables and the fixed-effect dummy variables.

Instrumental Variables Estimation

Unfortunately, controlling for household heterogeneity using fixed effects in a dynamic model introduces another source of bias because the lagged dependent term is then correlated with the error term (Hsiao).[12] The method of instrumental variables will be used to remove this bias. For convenience, the bias arising from the use of fixed effects in a dynamic model is explained below in the context of a pure lagged dependent model. The logic holds for the model in equation (7) since the average profit differential is correlated with the lagged dependent variable. The fixed-effects model effectively converts all variables to deviations from their household mean over the entire period:

$$(8) \quad y_{ijt} - \bar{y}_{ij} = a_1(X_{it} - \bar{X}_i) + a_2(w_{ijt} - \bar{w}_{ij}) + a_3(y_{ijt-1} - \bar{y}_{ij,-1}) + (e_{ijt} - \bar{e}_{ij})$$

where $\bar{y}_{ij}$, $\bar{X}_i$, $\bar{w}_{ij}$, $\bar{y}_{ij,-1}$ and $\bar{e}_{ij}$ are the household means of the respective variables over the entire period.

Note that y_{ijt-1} is a function of e_{ijt-1} and $\bar{e}_{ij}$ is a function of e_{ijt-1}. Therefore it follows that $E[y_{ijt-1}, \bar{e}_{ij}] \neq 0$ and, hence, $(y_{ijt-1} - \bar{y}_{ij,-1})$ breaks the condition for unbiased estimates by being correlated with the error in the fixed effects model. The bias that results from this correlation can be eliminated by using instrumental variables. Instruments are needed for y_{ijt-1} that are correlated with y_{ijt-1} but not correlated with the error term through $\bar{e}_{ij}$.[13]

Instruments have to be found for the average profit differential for the learning model. The ICRISAT data are especially well-suited for finding instruments. They provide information on the households in the five years before the introduction of the new seed as well as on the first five years of the adoption period. Since the means in equation (8) are constructed using data from the post-introduction period, any variables from the period before introduction are valid instruments.[14] The instruments must also be time varying so they are not collinear with the fixed effects. Two-stage least squares is then performed. Experience with other HYVs five periods ago, HYV_{t-5},

---

[10] Although random effects estimation can be adapted so as to allow for correlation between the household effect and specifically chosen regressors—"correlated random effects"—this involves including all leads and lags of the variable that is suspected to be correlated with the household effect. Hence, this would involve including all leads and lags of the profit differential terms, which leaves us with coefficients on the learning terms that are difficult to interpret.

[11] Other methods exist for eliminating the fixed effects but are not applicable in the current context, for example, removing the household mean for the period from the observations (as opposed to the household mean over all time periods as is the result of using household dummy variables) or first-differencing the variables. The first would, however, radically reduce the sample size because some households sow only one plot to cotton in some years. The second is impossible because of the inability to track plots over time in the ICRISAT data.

[12] This bias could be avoided if it were possible to remove the fixed effects by subtracting the household mean for the period from the observations (as mentioned in footnote 11). Note that first-differencing would entail the same bias as described above.

[13] Foster and Rosenzweig did not avoid this bias in their empirical work. They first difference their data but then use instruments that are not sufficiently lagged to avoid the correlation with the error term.

[14] Note that in other data sets, if the data set is long enough it would be possible to reserve some of the earlier periods for constructing instruments and not include them as part of the main sample. This may, however, involve losing some of the most interesting observations because it is likely that a lot of the learning takes place in the early periods.

and the income derived from those HYVs, π_{t-5}, are used as instruments. Equations (9) and (10) show the first- and second-stage regressions, respectively.

$$(9) \quad \sum_{n=1}^{t-1} [(y_{it-n}(\pi_{it-n}^{hyv} - \pi_{it-n}^{o})/N_{it}] = \eta_i + \delta_1 HYV_{t-5} + \delta_2 \pi_{t-5} + \delta_3 x_{it} + \delta_4 w_{ijt} + u_{ijt}$$

$$(10) \quad y_{ijt} = a_i + a_1 X_{it} + a_2 w_{ijt} + a_3 \sum_{n=1}^{t-1} [(y_{it-n}(\pi_{it-n}^{hyv} - \pi_{it-n}^{o})/N_{it}] + e_{ijt}.$$

Arellano and Bond propose an alternative method of instrumentation to deal with the bias caused by the inclusion of fixed effects in dynamic models. However, their method cannot be used effectively with ICRISAT data because it involves eliminating the household effects by first-differencing the data, and plots cannot be traced over time in the ICRISAT data. It is worth pointing out a significant drawback of their method in the context of learning models. By requiring two period lags (two lags are required for their test of serial correlation) any potentially valuable information contained in the first two periods is lost. These periods are likely to be vital in a relatively fast process such as learning. Arellano and Bond's method was used to estimate the learning model using household averages of the variables but it did not produce any significant results.[15]

Controlling for the Possible Endogeneity of the Explanatory Variables

The final problem to be addressed is the possible endogeneity of some of the explanatory variables. Any explanatory variables that are choice variables are possibly endogenous and so may be correlated with the error term. Rather than determining the adoption of the new seed, their values are affected by the adoption decision. Instrumenting for the endogenous variables using pre-introduction variables removes such correlation for the same reason as it does when instrumenting for the learning term.

The variables most likely to be endogenous are wages received and total household assets.[16] If incomplete labor markets exist and the new seed is more labor intensive than the traditional seed, adoption may result in more family labor being allocated to the plot, reducing income from other sources. Similarly, if there is a significant difference in profits between the new and old seed then use of the new seed could affect the household's assets in the current period. The instruments used below are the total (household and hired) number of male and female workers five periods ago and the number of household members five periods ago.

Empirical Results

The data used are the ICRISAT village level surveys of Kanzara in Maharashtra, India. Thirty-one households are covered over a ten-year period, 1975–84. The data are at the plot level and provide information on various plot and household characteristics. This analysis investigates the introduction of a new HYV cotton seed, AHH468, in 1980.

Focusing on the decision between planting the traditional cotton seed and planting the new HYV seed, only those fields that were sown to cotton were included in the sample.[17] Three observations were dropped from the sample because some variables had missing values. The model is tested on the period

[15] The results are available from the author upon request. Arellano and Bond's method is driven by the fact that if one first-differences the data, then the dependent variable (or any other variable) lagged two periods or more is a suitable instrument provided there is no serial correlation. Arellano and Bond propose a test of serial correlation of the errors to show whether or not y_{ijt-2} is a valid instrument. Their optimal estimator is obtained using the generalized method of moments (GMM).

[16] Irrigated area per plot could also be suspected of being endogenous, but in this context it is judged to be exogenous. Irrigation rights are determined at birth, and, although marketable, water is not often sold to neighboring farms (Walker and Ryan, p. 41). It may still be possible, however, for irrigation resources to be switched from one plot to another within a particular farm. To assess whether the inclusion of the irrigated area variable affected the results, the regressions were estimated with and without this variable. The omission of the irrigation variable did not significantly alter the results.

[17] Focusing on the decision between planting the traditional cotton seed and the new high yielding variety cotton seed assumes that the farmers are only considering planting cotton and they just focus on the decision as to which kind of cotton to plant. This may be somewhat artificial in that it ignores the multistage nature of the planting decision. The decision to plant cotton is likely to be influenced by the type of cotton seed that is available. See McGuirk and Mundlak for a discussion of this point. However, there is some evidence in the data of experimentation taking place between the two seeds. Of the thirty-one households that plant cotton over the five-year period, only seven plant the same type of cotton each year.

Table 1. Summary Statistics of the Data by Year

	1980	1981	1982	1983	1984	Total
Total number of cotton plots	73	64	71	59	78	345
Number of plots sown to the HYV	3	3	9	21	58	94
Number of households	29	25	25	23	26	

1981–84.[18] The resulting sample has 272 observations.

The period 1975–79 is used to construct the "pre-introduction" instruments. Table 1 reports the number of cotton plots, households, and number of plots sown to the HYV in each year. Figure 1 shows the households' use of the new seed over the period.

The estimation procedure used is the linear probability model. This procedure is used because it is the only estimation procedure that produces consistent coefficient estimates when fixed effects are required to control for unobserved heterogeneity. Estimates from probit and logit models are biased and inconsistent when fixed effects are used.[19] The linear probability model does, however, have some potential, albeit less serious, problems. They are (*a*) heteroskedasticity of the error terms which is overcome by using the White correction for heteroskedasticity and (*b*) the inability to constrain the predicted probabilities to lie between 0 and 1. This problem is more serious when the mean of the dependent variable is close to zero or one. The mean of the dependent variable in this study is 0.335.[20]

[18] The year 1980 is excluded from the sample because at that stage no learning had taken place.

[19] See Hsiao (pp. 159–61) for an explanation of the bias arising in maximum likelihood estimation with fixed effects. Conditional logit models with fixed effects produce unbiased estimates; however, the method is practicably unfeasible when the number of observations per fixed effect varies over time as it does in the ICRISAT data.

[20] Results from the linear probability model were compared with those from a logit model (both estimated without fixed effects) as an informal validity check. The results were similar.

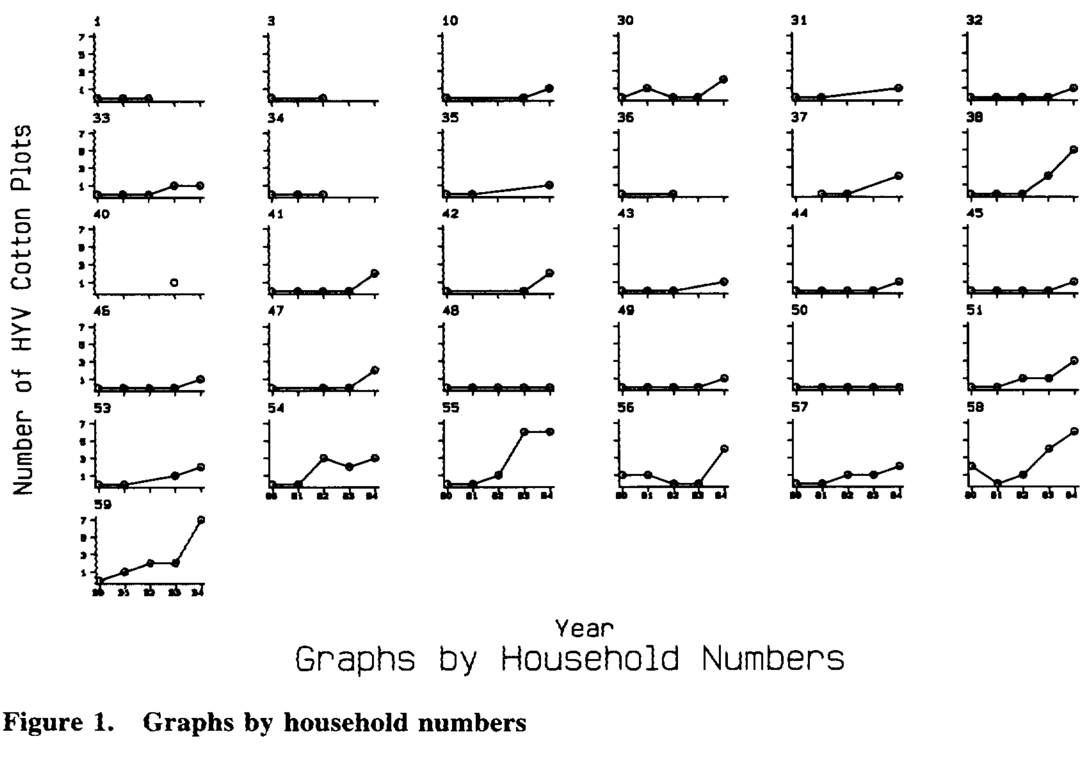

Figure 1. Graphs by household numbers

Table 2. Household Level Differences in Average Profits Between Seed AHH468 and the Traditional Cotton Seed

Year	Number of Households That Used Seed AHH486	Mean of Non-Zero Values (Rs/Acre)	Standard Deviation of Non-Zero Observations	Minimum	Maximum
1983	9	355.93	682.7	−408.1	1,882.3
1982	6	110.02	309.8	−240.8	503.9
1981	3	−123.47	196.6	−314.5	78.2
1980	2	−41.88	505.4	−399.2	315.5

The evidence of learning obtained from estimating the model on the panel is first examined below and these results are then compared with the cross-sectional results to assess the extent of the bias inherent in the cross-sectional estimates.

Evidence of Learning from Panel Data

The dependent variable in the following regressions, y_{ijt} is an indicator variable that equals 1 if plot j, owned by household i, was sown to the new cotton seed AHH468 in period t, and 0 otherwise. The explanatory variables can be broken into the following three categories:

1. the average profit differential "learning" term (in Rupees);[21]
2. variables that would appear in the farm's production function: indicator variables of soil type (deep black, medium depth black, medium-shallow black, or shallow red soil), the number of acres of the plot which are irrigated, the number of bullocks owned by the household, the years of schooling of the household head, the number of females in the household;
3. those variables that reflect households' access to credit, labor, output and input markets if these markets are imperfect[22] (the total household assets [Rs/1,000], the number of hectares of land owned by the household, and wages received in cash and in kind [Rs]). These variables may also affect the household's attitude toward risk and the household's decision making under uncertainty.

Summary statistics of the profit differentials by year are shown in table 2.[23] Table 3 shows summary statistics of the other explanatory variables and of the instruments as detailed above. Table 4 presents the estimation results.[24] Column 2 shows the results when one controls for household heterogeneity and instruments for the learning term and for the potentially endogenous variables. The coefficient on the learning term is statistically significant at the 5% level (p-value = 0.046). The fixed effects are significant at the 1% level. The instruments are jointly statistically significant at the 1% level in each first stage regression (as shown in table A.1 in the appendix).[25] A household that experienced an average profit differential of Rs500 per acre is 26.9% more likely to use the new seed in the current period than a household with no experience of the new seed (the

[21] There were only three households that in any year planted all cotton plots to the new seed. In these cases, all in 1983, the net income per acre from the traditional seed on average across the village (instead of the household average) was subtracted from the net income per acre from the new seed.

[22] Some of the explanatory variables, number of women in the household for example, may appear in the production function and reflect access to incomplete markets. Rosenzweig and Binswanger, in a study using Indian data, test the hypothesis of complete labor markets. For a study on the effect of credit limitations on product decisions that uses ICRISAT data, see Chaudhuri.

[23] It is unclear why the new seed was less profitable than the old seed in the early periods. It is possible that in the initial stages, learning about the profitability of the new seed involved learning about how the profitability varied with the use of different inputs as modeled in Foster and Rosenzweig. However, learning about inputs does not explain why the general reaction to the low profits was to not use the seed again in the following year. As suggested by an anonymous referee, it is possible that the farmers were improving the seed themselves over time. Another possibility is that farmers may be learning about inputs but that some inputs, irrigation for example, can't be immediately changed. The farmer may therefore not plant the new seed in the very next period while he increases his irrigation capabilities. However, the farmer plants the new seed the period after that when the new irrigation is in place.

[24] Note that the White correction for heteroskedasticity was used when estimating all of the reported results.

[25] The coefficient on the pre-introduction HYV indicator is positive, indicating that those who used a HYV cotton seed before 1980 were more likely to use AHH468. The income of those households who used HYV seeds in the pre-introduction period was negatively correlated with the average profit differential, possibly indicating that families who were more successful with the earlier HYV seed were less likely to change to the new HYV seed.

Table 3. Summary Statistics of the Regression Variables, Panel 1981–85

N = 272	Mean	Std. Dev.	Min	Max
Cotton seed AHH468 (indicator variable)	0.3346	0.4727	0	1
Irrigated plot area	0.3734	1.186	0	7
Deep black soil (indicator variable)	0.0772	0.2674	0	1
Medium depth black soil	0.8566	0.3511	0	1
Medium depth-shallow black soil	0.06250	0.2425	0	1
Shallow red soil	0.00369	0.0606	0	1
Owned bullocks	3.794	2.893	0	10
School years of household head	5.518	4.273	0	12
Number of females	2.710	1.334	0	6
Total household assets (Rs/1,000)	115.84	101.42	5.005	330.23
Area owned by the household (hectares)	10.669	9.407	0	26.5
Household wages (Rs)	2,013.7	3,380.33	0	13,256
Profit differential ($t-1$) (Rs)	40.509	292.944	−408.11	1,882.3
Average profit differential	11.864	192.429	−399.21	1,170.6
HYV ($t-5$)	0.217	0.413	0	1
Income per acre ($t-5$) if household sowed some HYV	93.995	206.23	0	734.14
Own and hired male hours ($t-5$)	2,009.08	1,741.66	0	6,032
Own and hired female hours ($t-5$)	2,729.86	2,625.33	0	9,695
Household members ($t-5$)	7.43	3.547	0	20

Definitions of variables: irrigated plot area = area of plot that is irrigated (acres); soil type: indicator of deep black soil (omitted variable), indicator of medium depth black soil, indicator of medium depth-shallow black soil, indicator of shallow red soil; owned bullock hours per plot; years of schooling = the years of schooling of the household head, females = the number of females in the household; total assets = total household assets (Rs/1,000); owned area = hectares owned by the household; wages = household wages received in cash and kind (Rs); HYV($t-5$) = 1 if the household sowed one or more plots to a HYV cotton seed five periods ago; income per acre five periods ago if household sowed some HYV cotton five periods ago (Rs/acre); own and hired female/male hours = number of hours spent on crops by household and hired workers.

value of the average profit differential ranged from −399.21 to 1170.6).

The finding that learning from own experience is a significant determinant of the probability of adoption and that unobserved household heterogeneity was also highly significant means that if the explanatory variables are correlated with the learning term or the unobserved household heterogeneity the estimates obtained from purely cross-sectional data will suffer from omitted variable bias. The magnitude of the bias is examined below by comparing the cross-sectional estimates with those from the panel.

Comparisons of Cross-Section and Panel Results

The coefficients on the other explanatory variables in the panel estimation indicate that an increase in the irrigated plot area of one acre will result in an increase in the probability of adoption of 5.62%. Medium-depth black soil decreases the probability of adoption by 30.6%, and medium-shallow black soil by 31.4%, relative to deep black soil. An extra female decreases the probability of adoption by 15.2%.[26] These results can be compared with results obtained from cross-sectional regressions.

Column 3 shows the result of estimating the model on the pooled cross-section with no dynamic term. The regression was estimated with year effects but no fixed effects. This method constrains the coefficients to be the same across all four years as they were in the panel estimations and allows for a fair comparison of the results. Column 4 shows the t-statistics for the tests of equality of the panel and pooled cross-section estimates. None of the estimates in the pooled cross-sectional regression differ significantly from the panel estimates. This is a surprising result given the magnitude and the statistical significance of the dynamic learning term and of the household fixed effects. An examination of the correlations between the learning term and the other explanatory variables reveals that none of the correlations were greater in absolute magnitude than 0.079. Similarly, the correlations between the household dummy variables and the explanatory variables were examined and the vast majority were less than 0.1 in absolute value.

The low correlation of the explanatory variables with the learning term is due to the large number of observations with a value of zero

[26] These are the coefficients that are statistically significant at the 10% level.

Table 4. Regression Results

	Dependent Variable = 1 if the new seed was planted, 0 otherwise			
Regression #:	[1] Instrumental Variables	[2] Instrumental Variables	[3] Pooled Cross-Section	t-stats [2] vs [3]
Instrumenting for:	Average Profit Differential	Average Profit Differential Wages, Assets		
Constant	0.462***	0.340***	0.331***	
	(3.098)	(1.862)	(3.127)	
Average profit differ'l (×1,000)	0.600***	0.537***		
	(1.996)	(1.998)		
Irrigated plot area	0.0469***	0.0562***	0.0442***	0.583
	(2.784)	(2.848)	(2.838)	
Medium black soil	−0.300***	−0.306***	−0.214***	0.710
	(−2.978)	(−2.529)	(−2.460)	
Medium-shallow black soil	−0.276	−0.314***	−0.108***	0.149
	(−1.614)	(−1.780)	(−2.584)	
Shallow red soil	−0.333	−0.0446	−0.1082	0.159
	(−1.218)	(−0.112)	(−1.007)	
Bullocks owned by household	−0.000479	−0.0195	−0.0160	0.0568
	(−0.136)	(−0.344)	(−0.797)	
Years of schooling			0.00347	
			(0.398)	
Females	−0.157***	−0.152***	−0.0438***	1.191
	(−1.798)	(−1.735)	(−2.114)	
Total assets	0.000813	−0.000440	0.000388	1.164
	(0.759)	(−1.083)	(0.457)	
Owned area	−0.0306	0.0386	0.00310	0.622
	(−1.356)	(0.692)	(0.295)	
Wages (×1,000)	0.0472***	0.0287	0.0155***	0.135
	(2.332)	(0.384)	(1.868)	
Year = 1984	0.539***	0.733***	0.631***	
	(7.186)	(4.434)	(11.207)	
Year = 1983	0.171***	0.434***	0.237***	
	(2.146)	(1.846)	(3.366)	
Year = 1982	0.0618	0.166	0.039	
	(1.041)	(1.428)	(0.813)	
Household fixed effects	yes, significant at 1% level	yes, significant at 1% level	no	
Adjusted R^2			0.3921	
Observations	272	272	272	

Note: *** indicates significance at the 10% level. T-statistics are shown in parentheses. The White correction for heteroscedasticity was used in all of the reported results.

for the learning variable. Hence, cross-sectional estimates obtained early in an adoption process when the new technology is not widely in use, and so learning is not widespread, may not be seriously biased by the omission of a learning term.[27] However, the extent to which the inability to control for household heterogeneity in cross-sectional studies leads to bias is much harder to predict. In this study, although strongly significant, the household fixed effects were not strongly correlated with the explanatory variables. The extent to which this will be true in other studies depends on the explanatory variables chosen and will need to be assessed on a case-by-case basis.

As discussed above, cross-sectional estimates are valid once all learning is complete. Hence, a consequence of the above finding is that cross-sectional estimates are likely to be

[27] The lower bias is also consistent with Bayesian learning regardless of the number of plots planted to the new seed in the early periods. In the first couple of periods adoption is primarily driven by prior beliefs, which are most likely correlated with the farmer's observable characteristics such as soil and human capital. In later periods, as more new information accumulates, the bias from ignoring learning increases.

the most seriously biased in the intermediate periods after significant adoption has taken place and before learning is complete.

Conclusions

This article uses panel data to study the dynamic nature of the adoption of a new high-yielding variety seed. The results suggest that learning from own experience plays an important role in the adoption decision. Unobserved household heterogeneity also plays a significant role.

Panel data are often difficult to come by and, as a result, researchers are often limited to using cross-sectional data. This article establishes that cross-sectional estimates are biased due to their inability to incorporate dynamic elements (such as learning) and because they have no way of controlling for unobserved household heterogeneity. However, the bias was found to be small due to (*a*) weak correlation between the explanatory variables and the learning term, and (*b*) weak correlation between the explanatory variables and the household effects.

It is hypothesized that the bias in cross-sectional estimates is likely to be the most serious in periods in which learning is still taking place but in which the new technology is already in relatively wide use. Replication of this study on other data sets will establish whether the small magnitude of the bias in the cross-sectional estimates is a general rule or a characteristic of this data set and learning model. Future work should also explore alternative learning models and other dynamic mechanisms.

[Received May 1997; accepted April 1998.]

References

Arellano, M., and S. Bond. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Rev. Econ. Stud.* 58(April 1991): 277–97.

Besley, T., and A. Case. "Modeling Technology Adoption in Developing Countries." *Amer. Econ. Rev.* 83(May 1993a):396–402.

———. "Taking Learning Seriously: A Diffusion Model for HYV Cotton." Unpublished, Princeton University, 1993b.

Chaudhuri, S. "Essays on the Consumption and Production Behavior of Households in a Poor, Agrarian Economy." PhD dissertation, Princeton University, 1996.

Dixon, R. "Hybrid Corn Revisited." *Econometrica* 48(September 1980):1451–61.

Feder, G., R.E. Just, and D. Zilberman. "Adoption of Agricultural Innovations in Developing Countries: A Survey." *Econ. Develop. and Cultur. Change* 33(January 1985):255–98.

Fischer, A.J., and R.K. Lindner. "The Effect of Distance to the Information Source or Information Quality and the Time to Adoption." Unpublished, University of Adelaide, 1980.

Foster, D., and M. Rosenzweig. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *J. Polit. Econ.* 103(December 1995):1176–1209.

Griliches, Z. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica* 25(October 1957):501–22.

Hsiao, C. *Analysis of Panel Data Econometric Society Monographs No. 11.* Cambridge: Cambridge University Press, 1986.

Lindner, R.K., and A.J. Fischer. "Risk Aversion, Information Quality and the Innovation Adoption Time Lag." Unpublished, University of Adelaide, 1981.

Lindner, R.K., A.J. Fischer, and P. Pardey. "The Time to Adoption." *Econ. Letters* 2(1979): 187–90.

McGuirk, A., and Y. Mundlak. "Incentives and Constraints in the Transformation of Punjab Agriculture." Washington DC: International Food Policy Research Institute, Research Report Series 87, 1991.

O'Mara, G. "A Decision-Theoretic View of Technique Diffusion in a Developing Country." PhD dissertation, Stanford University, 1971.

Rahm, M.R., and W.E. Huffman. "The Adoption of Reduced Tillage: The Role of Human Capital and Other Variables." *Amer. J. Agr. Econ.* 66(November 1984):405–13.

Rosenzweig, M., and H. Binswanger. "Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments." *Econ. J.* 103(January 1993):56–78.

Walker, T.S., and J.G. Ryan. *Village and Household Economies in India's Semi-Arid Tropics.* London: Johns Hopkins University Press, 1990.

Appendix

Table A.1. First-Stage Regressions

Dependent variable (N = 272)	Average Profit Differential	Avge Profit Differential	Wages	H'hold Owned Bullocks
Constant	−218.2	−272.41	−325.07	−22.486
	(−1.840)	(−2.363)	(−0.440)	(−1.189)
Avge income/acre (t − 5)	−1.167	−1.531	−1.609	−0.0497
if HYV (t − 5) = 1	(−9.084)	(−11.481)	(−1.883)	(−2.275)
HYV (t − 5)	497.47	699.69	215.44	0.655
	(6.934)	(9.133)	(0.439)	(0.052)
H'hold and hired male		−0.130	0.220	−0.00613
workers (t − 5)		(−5.950)	(1.565)	(−1.708)
H'hold and hired female		−0.0398	−0.341	0.00574
workers (t − 5)		(2.990)	(−3.993)	(2.629)
Household members (t−5)		28.843	112.44	1.509
		(3.399)	(2.066)	(1.084)
Irrigated plot area	4.598	−0.214	29.394	1.555
	(0.568)	(−0.028)	(0.592)	(1.224)
Medium black soil	10.776	1.425	−331.55	−0.280
	(0.267)	(0.037)	(−1.346)	(−0.044)
Medium-shallow	−5.075	−51.237	−463.87	−2.718
black soil	(−0.080)	(−0.840)	(−1.186)	(−0.272)
Shallow red soil	−300.65	58.651	3,466.83	82.856
	(−1.236)	(−0.120)	(2.161)	(2.019)
Bullocks owned by	−20.546	−1.646	−657.56	−2.925
the household	(−1.336)	(−0.120)	(−7.495)	(−1.304)
Females	166.19	124.33	−6.867	−4.240
	(4.177)	(3.232)	(−0.028)	(−0.672)
Total assets	1.133			
	(2.744)			
Owned area	24.300	45.167	108.91	13.525
	(1.812)	(3.894)	(1.465)	(7.111)
Wages	−0.0186			
(×1,000)	(−1.824)			
Year = 1984	104.73	166.96	952.68	31.548
	(3.264)	(5.994)	(5.334)	(6.906)
Year = 1983	−8.851	−17.043	1,177.91	40.201
	(−0.251)	(−0.596)	(6.421)	(8.568)
Year = 1982	−9.947	42.056	631.98	17.631
	(−0.337)	(1.470)	(3.445)	(3.757)
Test of Instruments (p-value)	0.000	0.000	0.000	0.000
R-squared (adj)	0.5718	0.6114	0.9344	0.9523

Note: t-statistics are shown in parentheses. Fixed effects are included and significant at the 1% level.