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 vi-

[illegible]

profitability through experience) and compare

[illegible]

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

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

where $E(\pi^{hyv}_{ijt} - \pi^{o}_{ijt})$ 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

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

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

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

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

[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

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

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

$$y_{ij} = q_o + q_1x_i + q_2w_{ij} + e_{ij}. \tag{6}$$

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 av-

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-

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 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},

---

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.

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

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

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 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

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]

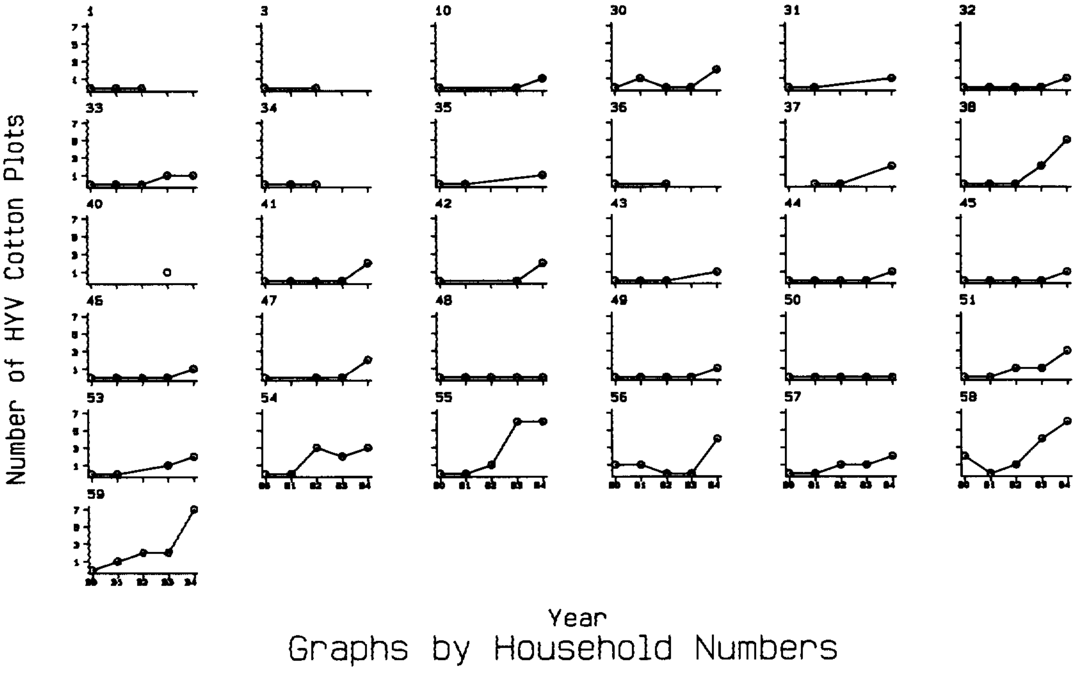

Figure 1. Graphs by household numbers

---

[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.

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.

[illegible]

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

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

These results can be compared with results ob-

[redacted]

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

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 dy-

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 Cap-

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) if HYV (t − 5) = 1	−1.167	−1.531	−1.609	−0.0497
	(−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 workers (t − 5)		−0.130	0.220	−0.00613
		(−5.950)	(1.565)	(−1.708)
H'hold and hired female workers (t − 5)		−0.0398	−0.341	0.00574
		(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 black soil	−5.075	−51.237	−463.87	−2.718
	(−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 the household	−20.546	−1.646	−657.56	−2.925
	(−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 (×1,000)	−0.0186			
	(−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.