Institute of Economic Development and Research
SCHOOL OF ECONOMICS
University of the Philippines

Discussion Paper No. 69-21 December 5, 1969

A CROSS-SECTIONAL STUDY OF THE DIFFUSION OF NEW RICE VARIETIES IN CENTRAL LUZON

By

Kelly
Mahar Mangahas, 1944-

Acknowledgements

This paper is intended as a first draft towards a Ph.D. thesis in the Economics Department of the University of Chicago. The work in the Philippines has been supported by the U.P. School of Economics, in the form of exemption from all other duties during academic year 1968/69, and by the U.P. Computer Center, in the form of free machine time; the International Rice Research Institute and the U.P.-Wisconsin Program in Development Economics contributed jointly towards research expenses. The work in the United States has been supported by the Rockefeller Foundation. Among the individuals who have helped obtain this support, V.W. Ruttan, T.W. Schultz, A.A. Castro, A. Kintanar, Jr., R. Barker and J. de Castro (more or less chronologically) must be cited with appreciation.

The author is particularly indebted to Dr. Leonardo Paulino, of the Bureau of Agricultural Economics, who provided the survey data used in this study. Messrs. Perfecto Villamor and Mamerto Damasco, also of the Bureau, gave much assistance on technical aspects of the survey.

Supplementary information was obtained from several other institutions and government agencies, among them the Rice and Corn Production Coordinating Council, the Rice and Corn Administration, the Agricultural Credit Administration, the Department of Rural Banks of the Central Bank of the Philippines, the Fertilizer Institute of the Philippines, Acknowledgements...contd.

ESSO Fertilizer, the Seed Corporation of the Philippines, and the mapping division of the Infrastructure Operations Center. Many thanks are due to the individuals from these and from other institutions who gave their cooperation.

A number of persons at the School of Economics provided much assistance, but of course are in no way responsible for defects in the work. Deo Reyes helped in several phases, most especially in data-gathering. Teresita Co helped in the preliminary computer processing. Arsenio Iloreta helped with some computations. Typing was done by Jerricha Caballes and Samuel Senturias. Lastly but certainly not least of all, thanks are due to Emmitt Summers, Jr. of the U.P. Computer Center for help in the difficult stages of editing the survey data.

Mahar Mangahas
19 December 1969

1. DIFFUSION OF INNOVATIONS IN PHILIPPINE AGRICULTURE

1.1 Introduction

The appearance in the latter 1960s of the new high-yielding varieties of rice, led by IR8 or "miracle rice", certainly has been a major event in the development of Philippine agriculture. This study is concerned with the factors which led to farmer acceptance of the new varieties, with focus on six major provinces in Luzon -- Bulacan, Nueva Ecija, Pangasinan, Tarlac and Laguna -- during crop year 1967/68.

The data consist of a sample of 866 farmers chosen roughly at random from a group which had been surveyed by the R. P. Bureau of Agricultural Economics in 1968.[1] The factors irrigation, season and extension assistance were used to divide the sample into a number of strata for separate analysis. Then, within each stratum, regression analysis was used to study links between the use of new varieties and the following additional factors: farmer-expertise, farm size, owner-operatorship, the interest rate, the marketed surplus ratio, and irrigation by pump. Using a one-zero adoption dummy as the dependent variable, the regression results may be used to obtain either probability functions or discriminant functions.

The next section is a review of the literature on diffusion of agricultural technology in the Philippines. It

---

[1]The survey and the subsample taken are described in Appendix 2.

is followed by an examination of ex ante considerations and hypotheses respecting acceptance of the new varieties, then by a description of the empirical method, and finally by a presentation of the results of the computations.

1.2 The Literature

There are a number of studies, most of them very recent, regarding the diffusion of innovations in Philippine agriculture. They are summarized in this section. Those which deal with agricultural innovations in general are by Madigan (1962), Feliciano (1968) and Liao (1968a), and those which deal with rice in particular are by Covar (1960), de Guzman and Dimaano (1967) and Liao (1968b).

Madigan took a sample of 521 from three barrios in Misamis Oriental, Northern Mindanao. He sought to explain variations in the predisposition to be resistant to community development (CD) projects requiring individual cooperation, at the time of entrance of the CD worker. Projects were limited to innovations in agriculture, livestock care and health practices. An individual's resistance score, call it R, was measured as the sum of the ratings of the individual's verbal response to 14 statements, with ratings on each response having the integer range (1,7), the predisposition to resistance (in the interviewer's opinion) increasing with the rating. An explicit purpose of the research was to provide CD workers with a procedure for estimating the R's of the individuals in a new CD project area.[2]

A number of variables were tested, mainly one at a time, for explanatory power with respect to R. Table 1.1 lists the variables and the significance levels obtained under several statistical tests.[3] The resistance score, R, was finally regressed against highest grade completed, a leadership ideals score and a house and grounds rating (determination coefficient = .171). These variables were chosen for their high simple correlation coefficients.[4]

---

[2]See pp. 247-93 (Madigan, 1962), entitled "A Practical Methodology for Use of the Predictive Instrument by Community Workers."

[3]The cost of computation at the time of the research weighed heavily on the choice of statistical method. Grouped data techniques were often used, and regression equations run only "in cases where correlation coefficients were sufficiently large to justify the labor" (p.35).

[4]The included variable leadership ideals was measured by the individual's responses to only 6 questionnaire items, and declared relatively inexpensive to collect. Income, size of largest field, leadership appeal, and popularity were rejected as explanatory variables due to anticipated problems of obtaining accurate information on the latter two variables (pp. 230-234). Qualitative variables were excluded from regression outright, apparently due to unawareness of the dummy variable technique (p. 229).

Table 1.1

Tests of possible determinants of Likert resistance scores[a]
Madigan (1962), Northern Mindanao

	r	r's Signif. level(%)	Signif. levels in % ANOVA	Chi-square	Diff.of Means	Diff. of Proporti
xpected negative relationship						
1. Literacy				.1	.1	.1
2. Education	-.37	.1				
3. Income	-.19	5				
4. Occupational status			1			
4a. Tenure status			NS			
5. Farm Area	-.11[b]	5				
6. Area of largest field	-.13[b]	1				
7. Social class (subjective)			5			
8. Migrant (yes/no)			NS		NS	
9. Held barrio office(yes/no)			.1		.1	
10. Popularity	-.11[c]	5				
11. Leadership appeal	-.16	.1				
12. Rating of house and ground	-.16	.1				
13. Rating of farm	.02	NS				
14. Landlord's interest in farm			NS		NS	
15. Health			5		5	
16. Energy			1		1	
xpected positive relationship						
17. Age	-.16[d]	1				
18. Number of children	.06	NS	-	NS[f]		
19. Barrio isolation			NS		NS	NS
20. Leadership ideals (pref. for authoritarian)	.32	.1				
21. Authoritarian personality	.04	NS			6[g]	
22. Distance of landlord's residence			NS		NS	
o relationship expected						
23. Sex			5[e]			5[e]

[a]See pp. 10-35, Madigan (1962) for expected relationship, and pp. 295- or the r's and the significance levels. For the tenure status test, see 67-168. NS means not significant at = .05.

[b]Persons without farms omitted.

[c]Only three classes used.

[d]Ages 25 and above.

[e]Positive for females.

[f]Resistance classes by numbers of children.

[g]Two groups: upper and lower scorers.

Means of subsets of the sample were cross-classified within each of the following three 'clusters':

1. Occupational cluster: means cross-classified according to literacy, holding of barrio office, occupation, tenure status, and size of farm.

2. Health cluster: means cross-classified according to health status, age and social class.

3. Popularity cluster: means cross-classified according to popularity and leadership appeal scores. These scores may or may not be available.

It was then suggested that an individual's R be estimated by a certain weighted average of his regression-computed R and his cluster means.

Feliciano (1968) studied four sets of barrios (in Laguna, Batangas and Albay) in which the UPCA Farm and Home Development (FHD) Project[5] operated. The FHD Project had chosen barrios whose farmers could be expected to have favorable attitudes towards extension personnel. In these barrios the

---

[5]The project started in 1958. Its aims were to investigate the problems of applying research findings on farms, to train subject-matter extension specialists, and to strengthen in-service training of government extension personnel at Los Baños (pp. 5-6).

people were generally cooperative, and had farm organizations of some sort; the land was sufficiently irrigated and drained; there existed peace and order; and the type of farming was typical of surrounding communities. Within the chosen barrios, certain farmers were selected as "cooperators", to be recipients of more intensive assistance than non-cooperators, on the basis of ability and willingness to adopt new farm practices (p. 9).[6] Feliciano studied 36 barrios in all, and in each took random samples of 10% of the farm households. The adoption variable was defind as the number of farming practices adopted. The range (0,14) was classified as low adoption, and the range (15,25) as high adoption; χ^2 - tests were used to test for relationship between the listed variables, one at a time, and adoption (Table 1.2).

The test results are generally of rather low statiscal significance. Not all of the expected directions of relationship between adoption and the independent variables materialized. In particular, the tests indicated that, individually, farm size and length of farming experience were

---

[6]Cooperators thus form a sample expected to be biased towards a high adoption rate. The fact that no material aid was extended (p. 10) would not alter the bias, for this proviso also holds for non-cooperators.

Table 1.2

χ^2 tests of possible determinants of levels of adoption of new farm practices: Feliciano (1968), Laguna, Batangas and Albay.

	Sign of relationship		Range of Signifi[cance]
	Expected[a/]	Actual[b/]	levels (%)[b/]
mographic variables			
Age	_[c/]	?	30,80
Education	+	+	10,50
Religion dummy (1 = Catholic)	-	+	10,50
mily variables			
Familism	-	?	50,90
With children in 4-H Clubs or in vocational agricultural school	+	+	30,70
cial variables			
Membership in farm organizations (Yes/No)	+	+	1,30
Index of such membership (Many/Few)	+	?	30,90
Participation in farmer-classes (Yes/No)	+	+	20,70
Frequency of such participation (High/Low)	+	+	1,50
Frequency of attendance at farm meetings	+	+	.1,80
onomic variables			
Farm size	+	-	10,70
Farming experience	+	-	30,70
Type of labor:			
Family labor only	-	?	
Family plus hired labor	+		2,20
Tenure class:			
Tenants	-	-	
Owners and part-owners	+	+	.1,5
mmunications variables			
Availability of Media	+	+	10,90
Time spent listening to the radio	+	-	2,30
Time spent reading printed media	+	+	.1,80

[a/] Pp. 89-90. Only those variables for which test results are available [ar]e listed.

[b/] See pp. 172, 316-331. Crop, livestock, and farm management practices [ar]e treated separately, hence the ranges.

[c/] Originally +, probably a misprint.

negatively ... Coralan, Laguna

of the ... rmers planting new varieties[8]

pe ... Wet season

5
72
94

... an important factor

... 14% of farmers plant-

ce ... figure was 49%).[9]

vide ... alogue was made of

De ... e or non-accep-

under the U. ... such as "demon-

seven out of 66 ... practice",

both 1965/66 and ... ility of re-

15 years of farming ... argued in

and cultivated about 2 ... rovide

nants. Ninety percent did ... le corre-

nization. Nevertheless the s ... prac-

was quite great both absolutely ... ethod,

practices recommended: ... e in-

... unt

---

[7]Barrio Coralan, one of eight barrios ... Maria-Mabitac Development Project. The Cora... nician began work in August 1964. The crop te... first accomplishment in this barrio was an adju... the planting schedules to the seasonal supply of... water; then varieties were changed to suit the new... ing months.

Barrio Coralan, Laguna

Percentage of farmers planting new varieties[8]

	Dry season	Wet season
1964	12	5
1965/66	31	72
1966/67	74	94

The change in planting schedule was an important factor (before the change, in the dry season 14% of farmers planted new varieties; after the change the figure was 49%).[9]

As in Feliciano's study, a catalogue was made of the reasons given by farmers for acceptance or non-acceptance of the new practices, under headings such as "demonstrated superiority or effectiveness of the practice", "ease in carrying out the practice", "availability of resources", "personal influence".[10] (It will be argued in the next section that this procedure might not provide bases for generalization about the future.) Simple correlations were taken between adoption of recommended practices and several variables (Table 1.3). By this method, the most important variables influencing adoption were indicated to be farm size, number of dependents, and amount

---

[8] De Guzman and Dimaano (1967), pp. 17-18.

[9] Ibid., p. 25.

[10] Ibid., pp. 44,45,58.

Table 1.3

Correlations between adoption of recommended farm practices and selected variables: Barrio Coralan, Laguna, 1965-67.

	Correlation coefficient	Significance level
Size of farm	.446	.01
Number of dependents	.341	.05
Age of respondents	-.013	NS
Amount of loan acquired	.438	.01
Length of farming experience	.018	NS
Years of schooling	.224	NS

Source: De Guzman and Dimaano (1967), p. 67.

of loan acquired.

Liao (1968a) studied three lowland rice areas in Laguna: Biñan, mostly rainfed; Cabuyao, pump-irrigated; and Calamba, gravity-irrigated. The first survey was in 1965/66, prior to mass diffusion of the new rice varieties. Nearly all respondents were tenants. He considered the following practices: (1) use of the hand tractor, (2) straight-row planting, (3) use of fertilizer, (4) chemical weed control and (5) modern pest and disease control, and defined an adoption index as the percentage adopted of the above practices. This index was included as an independent variable in a production function. (Table 1.4).

The production function evidence indicates that returns decrease with respect to scale (negative land coefficients), decrease faster for rainfed areas than for irrigated areas, and faster in the dry than in the wet season. Why these may be so are still unanswered questions. The adoption index performed creditably as an included variable, relative to the other variables. Its coefficient is markedly higher for Cabuyao than for the other two areas. This may be an indication of the contribution of water control (from pump irrigation) to the effectiveness of the recommended practices.

Table 1.4

Lowland rice Cobb-Douglas production functions incorporating an index of adoption of recommended practices.

The dependent variable is yield in 44 kg units/ha, in palay. Land is measured in hectares, labor in man-days per hectare, cash inputs in pesos per hectare, and the adoption index in per cent[a] adopted out of 5 recommended practices.

	Variable					
	Land	Labor	Cash Inputs	Adoption Index	R^2	Sample Size
				WET SEASON		
Biñan (rainfed)	-.28 (.14)	-.41 (.28)	.20 (.06)	.13 (.07)	.35	51
Cabuyao (pump-irrigated)	-.16 (.10)	.38 (.24)	.12 (.05)	.36 (.13)	.37	60
Calamba (gravity-irrigated)	-.10 (.08)	.11 (.16)	.13 (.06)	.07 (.04)	.32	58
				DRY SEASON		
Biñan (rainfed)	-.41 (.14)	-.31 (.29)	.14 (.06)	.06 (.05)	.49	32
Cabuyao (pump-irrigated)	-.27 (.21)	.14 (.36)	.13 (.09)	.47 (.18)	.34	39
Calamba (gravity-irrigated)	-.15 (.08)	-.12 (.15)	.37 (.10)	.16 (.06)	.38	58

[a]In computations, 1% is substituted for 0%.

Source: Liao (1968a), pp. 31-32.

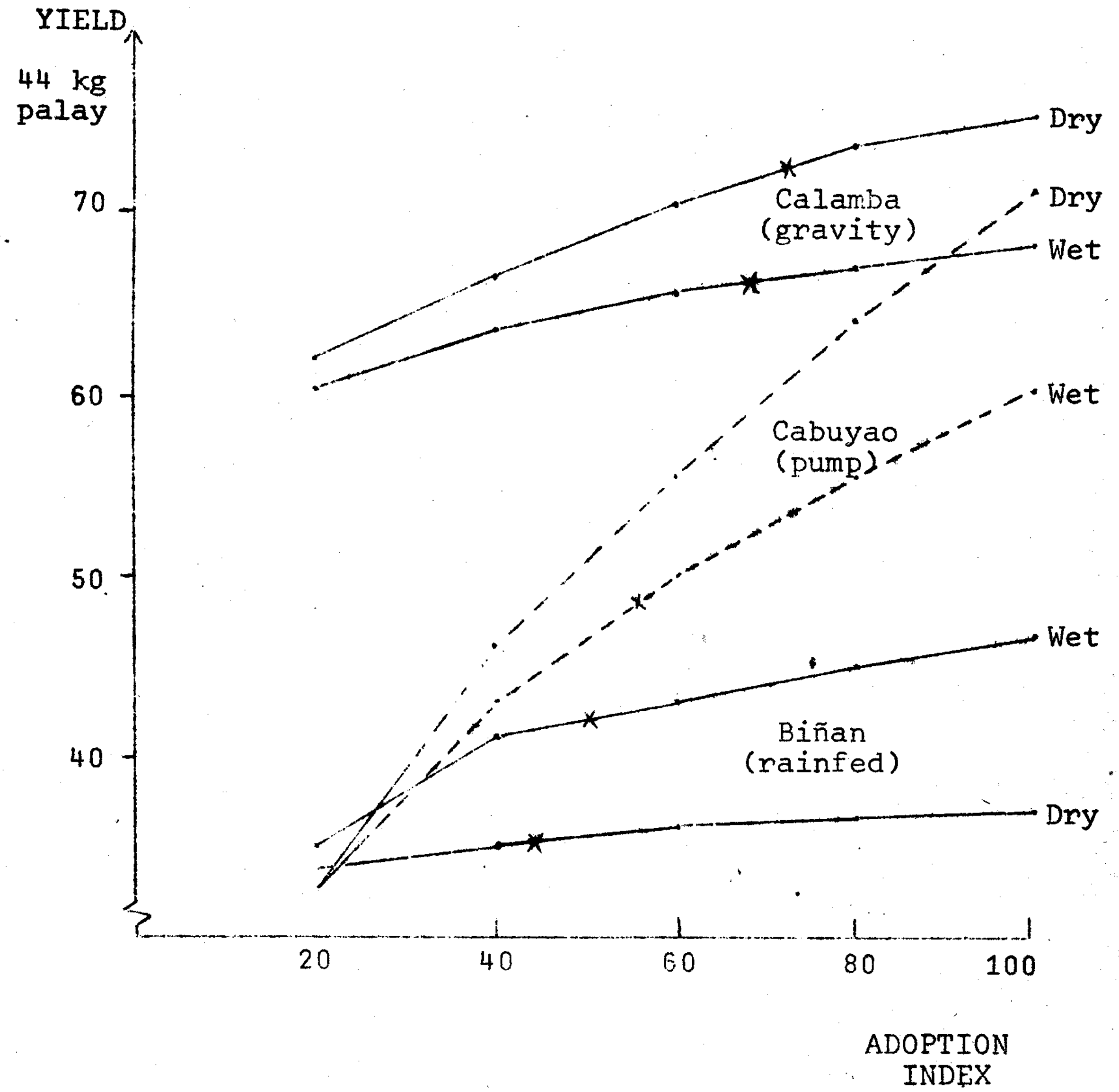
Figure 1.1 indicates graphically the potential contribution to yield of the adoption index. The partial production functions are traced with other inputs held constant at their arithmetic means, using the coefficients from the preceding table. X's mark mean levels of the adoption indices, by area and season; measured from these levels the absolute 'unused' potential of the adoption index, in terms of additional yield in cavans, runs as follows:

	Wet	Dry
Biñan (rainfed)	6	2
Cabuyao (pump-irrigated	12	17
Calamba (gravity-irrigated)	2	2

The absolute yield gains foreseen from expansion of average adoption indices to 100% are notably small, even for Cabuyao, relative to yield expected from the new rice varieties.

Liao then tried estimating a function determining the adoption index, the independent variables being schooling, age, farm size, landlord residence (1 = local resident, 0 = absentee), off-farm work (1 = part-time farmer, 0 = full time farmer) and government assistance. He considered government assistance with respect to (1) irrigation, (2) extension, (3) credit, (4) marketing, and (5) other forms of subsidy, and defined his assistance index as the percentage of the above services provided. The variables that

Figure 1.1



Source of data: Liao (1968a), pp. 34,103. The curves are partial Cobb-Douglas production functions (the adoption index is an included variable). Other inputs are held constant at their arithmetic means. X's mark the mean levels of the adoption indices, according to area and season.

turned out consistently significant across sample areas were schooling, off-farm work (negative coefficient) and government assistance. The coefficients of the government assistance variable were however rather small in size relative to those of the other variables. On the other hand the coefficients of landlord residence, though non-significant (positive in Cabuyao and Biñan) were large absolutely. There were no clear distinctions between the coefficients of the rainfed equation and those of the irrigated equation. Nevertheless the importance of irrigation is indicated by the following time-comparison (many of his respondents were also respondents in a much earlier survey):

	Average Adoption Index	
	Irrigated	Non-irrigated
1954/55	16.1	8.5
1965/66	63.5	49.4

Source: Liao (1968a). p. 60

These means were found significantly different, at the 10% level at least, both row- and column-wise.

Liao (1968b) continued his records of farmers in the three areas, concentrating on the process of diffusion of the new high-yielding varieties. Even though the data cover only 4 seasons, they clearly indicate that diffusion is swifter in the initial stage than later (Table 1.5). Liao used simple correlation coefficients (Table 1.6) to test for factors determining the area acceptance rate and the adoption period. Irrigation, schooling, and extension assistance were all individually significant; area planted was not. Age was significant with respect to the adoption period (older farmers adopted more swiftly) but not with respect to the area acceptance rate.

Covar (1960) dealt with the acceptance, in Laguna, of the Masagana and Margate (MM) systems of cultivating rice. These are comprehensive farming guides, from seed selection to harvest, that mainly involve greater labor cost than the traditional system: more careful seed selection, seedling treatment, transplanting, and weeding -- activities calling for added hired labor, a large ratio of which is paid for in cash. Heavy manuring is recommended; no particular chemical fertilization rates are mentioned.

Most of the diffusion problems reported involve farmers' fears that certain chemical treatments may be harmful to seed and that the low seed rates (wide planting dis-

Table 1.5

New rice varieties: area acceptance rates and adoption periods, Laguna (Biñan, Cabuyao and Calamba)

	Season	Area acceptance rate	Adoption Period
Pre-1967	Dry	.42	4 months
1967	Dry	.66	7
1967	Wet	.74	13
1968	Dry	.83	19

Adoption period = Date of adoption minus Date of awareness

Source: Liao (1968b), Figure 1 (unpaged).

Table 1,6

New rice varieties: simple correlation coefficients between area acceptance rates, adoption periods and other variables, Laguna (Biñan, Cabuyao and Calamba)

	Area acceptance rate		Adoption Period	
	Correlation	Significance	Correlation	Significance
Irrigation[a]	.34	1%	-.35	1%
Area planted	.05		.05	
Number of consultations with extension worker	.15	10%	-.19	5%
Schooling	.25	1%	-.14	10%
Age	-.03		-.10	1%

[a]Dummy variable: 1 = Irrigated, 0 = Rainfed.

Source: Liao (1968b), Tables 8 and 9 (unpaged).

tances) may be wasteful of land.[11] Covar felt that these fears were objectively unfounded.[12] He took data on attitudes, beliefs, practices and aspirations as well as on tenure and the usual demographic variables. Strangely, he did not present correlations between acceptance of the new MM systems and any of the variables collected. Nevertheless he offered the following generalizations (pp. 89-91): acceptance of the systems is positively related to (1) the number of dependents in the household, (2) farmer youthfulness, and (3) farmer education and training; and negatively related to (1) the farmer's age less the extension agent's age, and (2) the farmer's schooling less the extension

---

[11]Covar also has a section misleadingly entitled "Magico-religious Practices Associated with the Choice of Rice Variety" (pp. 16-18). The listed practices cover special treatments applied to seed, after selection, that hardly seem harmful to productivity (e.g.,seed selection by trusted men; seed storage in a special location in the home; seed blessing by a priest), and probably would not interfere with the acceptance of a new seed variety. Nevertheless this action is cited by Feliciano (1968, p. 66) as evidence that such practices are important barriers to diffusion of innovations.

[12]Two of the complaints reported appear however to be quite sensible: (1) MM system requires much labor for rigid seed selection; a smaller but more select quantity of seed than usual is then planted in the seedbed. Under the traditional system, a large quantity of seed is selected, more or less indiscriminately, and then seeded; when the seedlings come up it is relatively easy to select the best seedlings for transplanting (p.48). (2) The farmers simply do not have enough animals to provide the recommended dosage of manure (p. 54).

agent's schooling. The last two variables indicate a hypothesis that diffusion is easier when farmer and extension agent have similar ages and schooling.

1.3 Comments on the Literature

First of all, it seems unnecessary to state that the acceptance of an innovation is determined by numerous factors. Hence, for example, the variable-lists of Feliciano (1968, pp. 89-90) or Madigan (1962, pp. 34-35). It probably will also be accepted that the search is for a set of the most important determining variables (probably more than one). If A refers to acceptance and (x_1) to the set of explanatory variables, then the search is for relevant parameters of a function such as $A = f(x_1, x_2, \ldots)$ rather than functions such as $A = g(x_1)$, $A = g(x_2)$, etc.

Reliance on single variable analysis of variance or simple correlation coefficients or the chi-square test of independence of distributions is analogous to doing the separate regressions

$$(1) \quad A_t = a_1 + b_1 x_{1t}$$

$$A_t = a_2 + b_2 x_{2t}$$

etc.

when the appropriate equation is

$$(2) \quad A_t = a + b_1x_{1t} + b_2x_{2t} + \ldots$$

Doing regressions (1) instead of (2) will of course give biased coefficients. The signs of b_1, b_2 etc. as estimated from (1) may even be different from the corresponding signs as estimated from (2), and the significance levels may well be low for (1) but high for (2). R^2 will of course be higher for (2) than for any of the equations of (1), and may be statistically significant even though all the R^2's derived from (3) are non-significant. Almost all of the χ^2- tests of Feliciano in fact turned out highly non-significant; yet it is possible that significant multiple-variable analysis would result from the very same set of data. This is also possible for some of the variables declared non-significant in the De Guzman-Dimaano study or in Madigan's study.

One may note in the latter study that on occasion 'standardization' techniques analogous in purpose to multiple regression are applied when (and apparently only when) a preliminary test does not support the *a priori* hypothesis -- for example, the corrections made for age and education when it was found that women had significantly higher resistance scores than men.[13]

---

[13] Madigan (1962), pp. 156-9.

A second criticism concerns the common practice in Philippine agricultural surveys, of asking farmers to specify which item in a set (which is often provided beforehand) is the most important reason for making such-and-such a decision, or not making it. The objection is not with respect to the subjective truthfulness of the replies, but rather with respect to the usefulness of such data for generalizations regarding the future.

In general, a theory of farming decisions will state that some decision y depends in some manner on a certain set of variables x_1, x_2, etc. If many farmers state that x_1 is <u>the</u> "most important" variable affecting y, what does this imply? It certainly does not necessary imply, say, that dy/dx_1 is greater than dy/dx_2. For suppose that x_2 is the price differential between old and new varieties of rice, and also suppose that, at the time of the survey, $x_2 = 0$. Then farmers are not likely to state that x_2 is important; and yet dy/dx_2 may be 'large'. Respondents whose activities are not restricted by the current level of a variable are likely to take it for granted and not report it as "important", even though their future decisions may be quite sensitive to it.

It thus seems hardly appropriate to use single-variable statistical tests when one knows beforehand that a group of several variables is responsible for a farming decision, all of which are free to vary from sample unit. This is the usual problem when the data are non-experimental, and so it is more acceptable to take a multiple-variable method such as regression analysis.

2.0 VARIETY ACCEPTANCE FUNCTIONS FOR RICE IN CENTRAL LUZON AND LAGUNA: EX ANTE CONSIDERATIONS

In this section the economic advantage of new varieties over old is studied with reference to the experimental evidence, farm "status" variables, land tenure, the supply of inputs, and the degree of farm commercialization. Diffusion-determining variables, in capital letters, which are available in the survey data are introduced at relevant points in the discussion. (The available variables are listed in Appendix 4.)

2.1 Experimental evidence

Traditional Philippine varieties are generally tall, leafy, profuse tillering, photoperiod-sensitive, and susceptible to lodging. They are able to produce moderate but stable yields under adverse conditions as deep water and intense weed competition, and are adapted to conditions of low fertility and minimum care. Photoperiod-sensitive varieties have been naturally selected where water control is slight, since they may be planted whenever monsoon rains begin, yet will always mature at a fixed date after the monsoon season; harvesting is easier and the chance of typhoon damage at late stages of growth is minimized. Yield increases through a high rate of application of nitrogen or dense planting are not large due to lodging susceptibility. Lodging also raises the cost of harvesting and lowers the quality of the grain. The new varieties being developed are short and stiff-strawed, with short, narrow, erect and dark green leaves, are medium tillering and of early maturity, and are capable of substantial grain response to nitrogen without lodging.[1]

The probability of acceptance of a new variety by a given farmer may be assumed to be determined jointly by its agronomic superiority over old varieties, by external economic conditions affecting the feasibility of his taking advantage of this agronomic superiority, and by his so-called status variables. Demonstration of varietal superiority, prior to the seasons covered in this study, occurred chiefly on the experimental farm.[2] In order to maximize yields, experiments typically took advantage of well-irrigated lowland conditions and used high levels of modern complementary inputs such as fertilizer and protective chemicals.

---

[1]University of the Philippines (College of Agriculture) and International Rice Research Institute, Rice Production Manual, 1967.

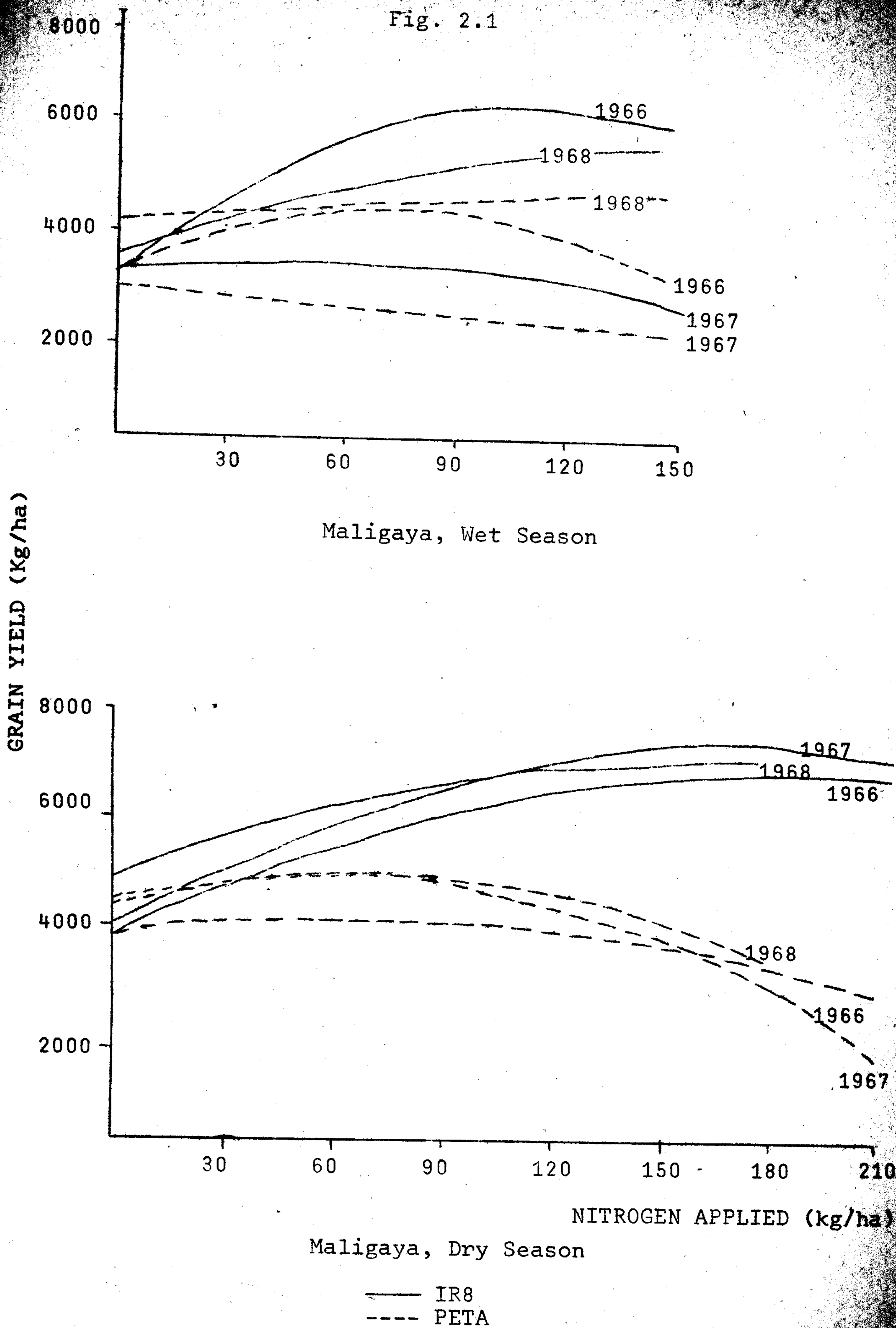
[2]The important experimental areas are the International Rice Research Institute (IRRI) and the University of the Philippines' College of Agriculture (UPCA), both at Los Baños, Laguna, and the Maligaya Rice Research and Training Center, at Muñoz, Nueva Ecija.

Response to nitrogen is the most important attribute for which the new varieties are bred, and experiments are often replicated for varying nitrogen levels. In the light of available experimental evidence, this section considers hypotheses on the possible effects on varietal adoption of the following variables: crop type, climate, soil, protective chemical inputs, and the degree of land preparation.

It is assumed that agronomic superiority is sufficiently described by the absolute difference between the yield of a representative new variety and that of a representative old variety at least with respect to irrigated land. The evidence following indicates that on irrigated land there is no significant difference -- as may have been previously feared -- between yield variances of new and old varieties on account of weather and disease factors.

Experiments at the Maligaya Rice Research and Training Center during 1966, 1967 and 1968 have provided a range of nitrogen response functions for IR8 (a new short-stemmed indica) and Peta (a traditional tall indica). These functions are reproduced in Figure 2.1. Response function variation is much greater in the wet seasons, as there is a greater incidence of typhoons, floods, droughts, and concomitant diseases. The 1966 wet season functions are the "best"; there was no unusual weather or crop damage. And the 1967 wet

Fig. 2.1



8000
6000
4000
2000
1966
1968
1968
1966
1967
1967
30
60
90
120
150
Maligaya, Wet Season
GRAIN YIELD (Kg/ha)
8000
6000
4000
2000
1967
1968
1966
1968
1966
1967
30
60
90
120
150
180
210
NITROGEN APPLIED (kg/ha)
Maligaya, Dry Season
IR8
PETA

season functions are the worst; there was a bad typhoon, incidence of bacterial leaf blight, and lodging in the case of Peta. Although there was a serious drought in Central Luzon in the wet season of 1968, irrigation was adequate at Maligaya. For both seasons and at any given nitrogen level it is clear that the range of achieved yields is no greater for IR8 than for Peta. As expected, the mean yield corresponding to any nitrogen level at least slightly greater than zero is always greater for IR8.[3]

There are three rice crop types of importance in Central Luzon and Laguna: (1) rainfed lowland, wet season, (2) irrigated lowland, wet season, and (3) irrigated lowland, dry season (second crop).[4] The importance of irrigation is obvious. Cet. par., the irrigated dry season yield is expected to be significantly greater than the irrigated wet season yield, because of the greater supply of solar energy and the absence of typhoons.[5] Hence the analysis is

---

[3]Data is from mimeo copy of section on Agricultural Economics of IRRI Annual Report for 1968, forthcoming.

[4]There are almost no sample observations for upland or for rainfed lowland, dry season.

[5]But often dry season irrigation is less adequate, due to the shortage of water storage facilities. It was hoped that the variable IRCOND (1 for adequate irrigation, 2 for inadequate) would allow for this. However, almost all farmers replied that their irrigation was inadequate.

stratified for these three crop types.

Fig. 2.2 presents experimental nitrogen response functions for two locations and for each season. At Maligaya, the difference between the yield of IR8 over that of Peta is greater for the dry season, for most nitrogen levels. This is also true for IRRI, for nitrogen levels up at 80 kg/ha; at 60 kg N/ha or less, the wet season yield difference is greater. The recommended levels (1967) of nitrogen for IR8 are 100 kg N/ha in the dry season and 70 kg N/ha in the wet season,[6] and at these levels the yield differences are greater for the dry season for both locations. The inference is that for irrigated area the probability of adoption is greater for the dry season.[7]

---

[6] Rice Production Manual, p. 180.

[7] The climate throughout the area of study is quite similar except with respect to the dry season, for which two climate types may be distinguished (Huke (1963) pp. 44-51):

A. Long low sun dry season (5 or 6 months with less than 2.4 inches of rainfall per month)

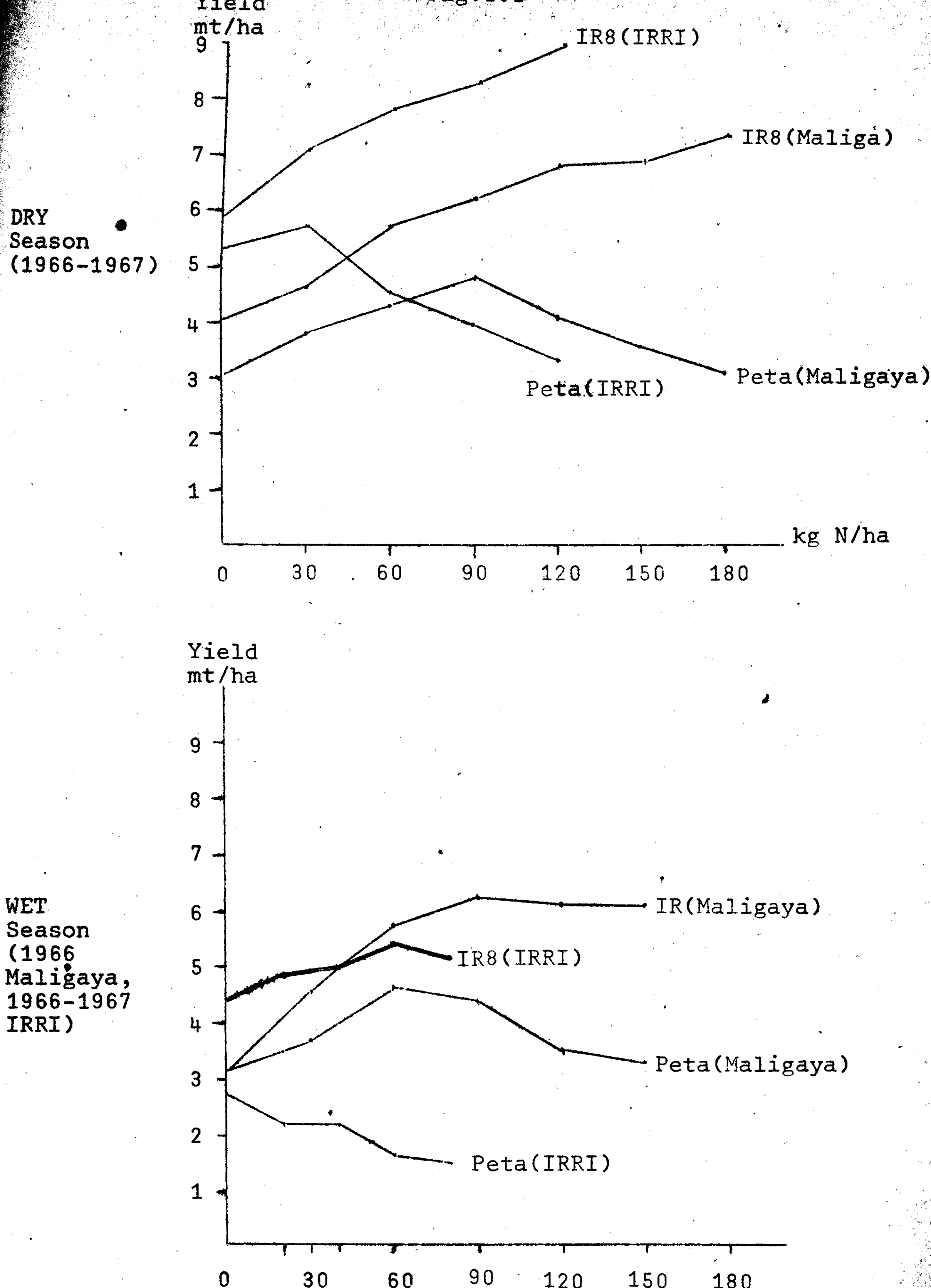
B. Intermediate low sun dry season (4 months with less than 2.4 inches of rainfall per month)

The provinces may be classified roughly:

1.	Bulacan and Laguna	Climate B
2.	Nueva Ecija and Tarlac	Half A and half B
3.	Pampanga and Pangasinan	Climate A

Thus water storage or pumping facilities should be less crucial for Laguna than for Pangasinan, since the dry season is shorter. Under the assumption that old varieties are hardier under dry conditions, varietal change should be more likely, cet. par., for areas with a milder dry season. This was not confirmed by the data, however. Diffusion rates for the dry season were high in Tarlac, Pangasinan and Laguna, moderately high in Nueva Ecija, and relatively low in Bulacan and Pampanga.

Fig.2.2



Source: IRRI, Yield response to nitrogen, mimeo, VC 68(3): 3, Tables 3.1-3.3, 3.5, 3.7, 3.8.

Cet. par., unfertilized *soil* of a given quality is likely to give a higher yield than unfertilized soil of a lower quality. In Fig. 2.2 unfertilized experimental yields from IRRI may be compared to unfertilized experimental yields from Maligaya, and it is seen that in all cases except for the variety Peta in the dry season, the higher-quality IRRI soil gives higher yields. (These are irrigated lowland fields.) It is also seen that the absolute difference between the yields of IR8 and that of Peta[8] is greater for IRRI than for Maligaya. This is evidence (scanty to be sure) in favor of the hypothesis that farms with high quality soils, other things equal, are likely to be earlier adoptors of new rice varieties than farms with low quality soils. The dummy variables SOIL1 (1 for first-rank soil series, ..., 4 for fourth-rank soil series) and SOIL2 (1 for first-rank soil texture, ..., 3 for third rank soil texture)[9] are used to test this hypothesis.

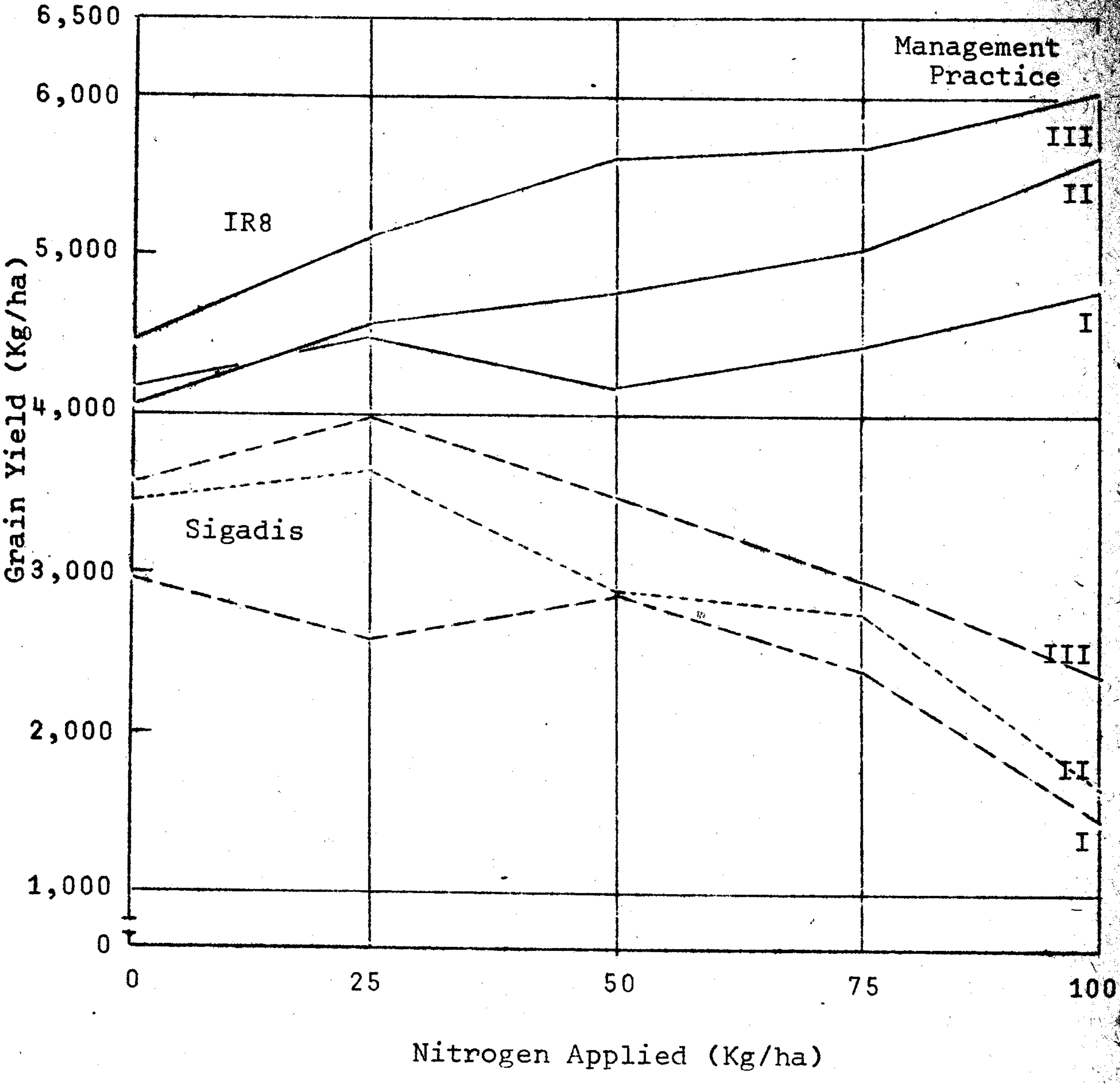
The application of *chemical* protective inputs may have a slight effect, if any, on the *difference* between yields of new versus old varieties. Fig. 2.3 presents the results of a

---

[8] A Philippine Seedboard variety of several years' standing, commonly used in experiments concerning varietal comparison.

[9] See Appendix 3. There is a difference of opinion among soil scientists as to whether soil *series* or *texture* is the more important discriminatory variable, with IRRI experts favoring *texture*.

Fig. 2.3



Effect of varietal types, levels of Nitrogen, and management practices on the grain yield of rice. IRRI, 1966 wet season

Source: De Datta and Barker (1967)

wet season experiment using three levels of chemical inputs (I,II,III)[10] on IR8 and on Sigadis, a tall weak-strawed variety.

However, there are serious problems in the fertilizer and chemicals data concerning the interpretation of the quantities applied. In the case of fertilizer, the identification variable NPKID includes brand names, e.g., Esso Engro, as well as generic names, e.g., Urea, thus making it difficult to compare across farmers the quantities applied (NPK1). In the case of chemicals the identification variable CH includes one herbicide in addition to several insecticides. Scientists at IRRI[11] felt that the quantity applied, CH, was hardly meaningful since effectiveness of the chemicals depends on some factors which are not in the data: the number of applications (so that chemical strength per application may be computed), the type of pest, and the extent of the infestation. In the final computations, fertilizer and chemical application levels were excluded from the diffusion functions.[12]

---

[10]Although De Datta and Barker (1967, pp. 16 ff.) refer to three levels of "management", it is clear from their description that the differences between "management" levels are basically differences in chemical application levels. The only non-chemical component in "management", as they use the term, is handweeding.

[11]Dr. S. K. de Datta, Mr. J. Calderon.

[12]The farmer-expertise variable EXPERT, described in the following section, is however partially determined by usage of fertilizer and chemicals.

A minimum amount of land preparation appears necessary for a new variety to keep its yield advantage. De Datta and Barker (1967, pp. 25 ff.) report that, in an experiment, IR8 gave consistently higher yields than BPI-76, regardless of the number of times land was harrowed, provided it was harrowed at least once. IR8 lost its advantage only in the case of land plowed but not harrowed. Data on number of harrowings is not available in my sample; and in any case the minimum amount of land preparation required for IR8 is quite small. Hence land preparation will not be considered as a factor determining varietal acceptance.

In summary, crop type, and soil are possible significant factors affecting varietal acceptance; they shall be considered in the empirical analysis as characteristics for stratifying the sample. The yield advantage of the new varieties is definitely linked to the fertilization level. Chemical protective inputs may have only a slight effect on varietal advantage. Unfortunately the survey data do not indicate fertilization and chemical levels in a fashion useful for inter-farmer comparisons. The degree of land preparation cannot be tested as a factor affecting acceptance, but in any event there is little a priori necessity for such a test.

2.2 Farm status variables

AGE is traditionally included although its relevance is not too clear-cut. Madigan found it significantly positively related to adoption (negatively related to resistance), but Feliciano, De Guzman-Dimaano and Liao (1968b) all found it negatively related and non-significant.

These authors all found education to be a positive factor, in most cases significant. This variable may be represented by SCHOOL, measured in number of years of formal schooling. Morrow (1966), however, found no relationship between years of schooling and yield in Bay (Laguna) and Mayantoc (Tarlac), considering wet and dry seasons separately.

Thus it may be fruitful to consider other variables which may be more indicative of the level of farm expertise: MLAND (2 or 3 indicates mechanized land preparation), MSEED (2 or 3 indicates seed selection using a salt or other chemical solution), MSDBED (2 or 3 indicates use of the recommended dapog system of seedbed preparation), MPLANT (2 or 3 indicates straight-line planting), MWEED (2, 3, or 4 indicates use of herbicides or a rotary weeder), NPK1 (quantity of inorganic fertilizer applied), and CH (quantity of protective chemicals applied).

/

A farmer expertise variable, EXPERT, measured by the proportion accepted of these recommended practices, was constructed using the above seven variables. EXPERT is a substitute variable for SCHOOL.

Liao (1968a) found that off-farm work is a significant negative factor with respect to an index of adoption of modern farm practices. However, almost none of the farmers in the sample seemed to earn substantial amount from off-farm work,[13] hence this factor could not feasibly be tested.

Area planted (AIN) is a variable of interest since one would expect that a negative relationship exists between it and yield,[14] indicative of decreasing returns to scale.[15] The expected absolute gain per hectare due to varietal change must be multiplied by the number of

---

[13]The variable JOB1, for primary occupation, takes values 6, 7 or 8 if the respondent is primarily a farm owner, a tenant, or a hired farm worker. If these values define a dummy variable, where a zero value indicates when off-farm work is a more important source of income, almost no farmers are assigned a zero.

[14]See the empirical production functions in Table 1.4 supra from Liao (1968a). Such a relationship has also been shown (1) for data pertaining mainly to share tenants (Ruttan, 1966) and (2) for time series data aggregated to the regional level (Venegas and Ruttan, 1964.)

[15]With respect to all inputs besides the farm operator, that is.

hectares planted to obtain the total gain for the farm. Assuming that the gain is directly related to the varietal yield-differential, larger farms will enjoy greater absolute gains from varietal change than smaller farms, unless the relative advantage in hectarage is overcome by the relative disadvantage in yields. But this advantage seems unlikely to be overcome,[16] and hence it is hypothesized that larger farms are more likely to be earlier adoptors than smaller farms.

The assistance of extension workers is reflected by variables ID3 (indicating whether the barrio is included under the rice production program, or is a land reform barrio, or is under a special project), and COOP (indicating whether the farmer, if belonging to a program barrio, is a program "cooperator"). The hypothesis of course is

---

[16]Let h refer to hectarage, y to the new variety's yield, x to old variety's yield, and subscripts 1 and 2 to two farms, with the second farm being the larger. Then the absolute gain in output due to varietal change is greater for the large farm as long as

$$h_2(y_2 - x_2) - h_1(y_1 - x_1)$$

is positive. Assuming an elasticity of yield with respect to hectarage, e , equal for both varieties, and denoting by d the relative difference in hectarage with farm 1 as base, then this condition is satisfied provided that the absolute value of e is less than 1/(1 + d). It appears from Table 1.4 supra that the absolute value of e is at most .4; in which case d would have to be greater than 150% (and this value of e would have to be applicable to a 150% difference in hectarage) for the condition not to be satisfied.

that those assisted by government are more likely to be earlier adoptors. It is recalled that Liao (1968a, 1968b) found government assistance to be a significant variable. Also, Mina and Tiongson (1967), considering Nueva Ecija and Iloilo, have found yield positively related to the length of time spent as a program cooperator. Specifically, three farmer-groups may be considered: (1) those in barrios not under any program, (2) those in program barrios who are not cooperators (but presumably neighbors of cooperators), and (3) program cooperators and those in other special projects, all of whom receive "intensive extension". If extension assistance is a critical variable, then clear differences may be expected among acceptance coefficients for these groups.

2.3 Tenancy

The effect of varietal change on the expected return to the inputs owned by the decision-making unit may vary with the farm's tenure status. This status is indicated by the variable TENCY. Other things equal, it is expected that owner-operators would be the earliest adoptors of the new rice varieties, since in these cases one individual receives the absolute gains accruing to two inputs, land and labor, and the decision for varietal change is concentrated in him. This is supported by Feliciano (1968), who, studying general farm practices in Laguna, Batangas and Albay, found using a chi-square test that owner-operators were "high adoptors" while lessees, part-owners and share tenants were "low adoptors".

In determining the expected ranking of share farms and lease farms with respect to varietal change, the role of landlords should not however be underrated. Polson and Pal (1956, p. 85) report, regarding the Dumaguete Area (Negros Oriental):

> "The use of new practices or innovations can be initiated by the tenant only with the permission of the landlord. The results of innovations are considered so uncertain that it is a big responsibility for a tenant to adopt new practices without the landlord's permission and support."

Covar's evidence for Laguna (1960, pp. 10-11, 13-14) is that landlords are dominant over tenants with respect both to varietal choice and to provision of seed. Feliciano (1968) reports that tenant- "low adoptors" need landlord-permission in order to adopt an innovation, and that failure to obtain permission is not entirely due to landlord absenteeism.[17]

The dominance of landlords in production decisions also has legal bases. Republic Act 1199 of 1954, Section 25 (1), states that the landlord has the right to choose the type of crop and the seed to be worked with by the tenant. The present Agricultural Land Reform Code RA 3844 of 1963, in Section (3) provides that the agricultural lessor has the right

> "To require the agricultural lessee, taking into considerations his financial capacity and the credit facilities available to him, to adopt in his farm proven farm practices necessary to the conservation of the land, improvement of its fertility and increase of its productivity: Provided, that in case of disagreement as to what proven farm practice the lessee shall adopt, the same shall be settled by the Court according to the best interest of the parties concerned."

Varietal choice, in the case of either share or lease tenancy, is therefore made jointly by the owners of land and labor inputs, and not by the owners of labor

---

[17] For a comparison of "managing" versus "non-managing" landlords, see Bernal (1967), pp. 62-64.

alone. If the latter event were true, then any difference between the return to labor[18] under one tenure system and that under another would be a more important factor affecting varietal change of one tenure group relative to the other. The evidence is thus not too strong in favor of a hypothesis that members of one tenure group are more likely to be earlier adoptors. The hypothesis to be tested is that no significant difference exists between the two tenancy groups with reference to adoption of the new rice varieties. The variable used is OWNER (1 for owner-operators and zero otherwise).

2.4 The supply of inputs

The data do not include costs of purchased inputs, and hence input prices are not available for analysis. There do not seem to be reasons to believe, however, that the prices paid for seed, fertilizer and farm chemicals differ significantly among farmers in Central Luzon.[19] Table 2.1 contains seed selling prices of a sample of 43 out of an intial group of 93 private farms which received new variety

---

[18]Cheung (1968) expresses some doubt as to whether in general such a difference exists.

[19]With respect to fertilizer and farm chemicals, the information is from N. Bonoan of Esso Fertilizer.

Table 2.1

Selling price of palay-seeds, by outlet, 43 private seed multiplication farms, RCPCC program, 6 provinces, Philippines, 1966-67.

Outlet	Laguna	Bulacan	Nueva Ecija	Pampanga	Tarlac	Pangasinan	All provinces
IR8							
RCPCC (ACA, BPI & RCA)	30.00	-	30.00	30.00	28.67	30.00	29.86
Philippine National Bank	-	30.00	30.00	30.00	30.00	-	30.00
Farmers Cooperative Marketing Associations	-	-	-	-	30.00	-	30.00
Philippine Sugar Institute	30.00	-	-	-	-	-	30.00
Seed producer and other farmers	25.75	-	28.16	30.00	29.39	28.72	27.85
All outlets	29.19	30.00	29.55	30.00	29.19	29.33	29.34
Other varieties							
Bureau of Plant Industry	-	-	-	-	29.73	-	29.73
Seed producer and other farmers	25.00	23.20	25.42	-	31.20	30.00	28.59
All outlets	25.00	23.20	25.42	-	30.35	30.00	29.01
All outlets All varieties	29.19	28.45	28.47	30.00	29.44	29.57	29.31

Source: Mina and Tiongson (1967), p. 35.

seed from the government for multiplication. The price differences across provinces for IR8 are quite small.[20]

Additional evidence is from the replies to survey queries as to reasons for not using new seed, fertilizer or chemicals (variables named SDWHY1, SDWHY2 and (HWHY). Although "unavailability" was listed among the set of anticipated replies, hardly any of the farmers surveyed gave this reply.

The variable INTRST, or interest rate, is used to measure the cost of farm credit. The range of this variable extends to zero since interest-free loans are fairly common, e.g., landlords providing credit as part of the tenure bargain with tenants. Farmers who self-finance their operations are also assigned an INTRST level of zero. For lack of definite information, it is assumed that their opportunity earnings from financial investments are negligible.

---

[20] It is interesting to note that in Pangasinan and Tarlac nearly 70% of seed sales were made directly to other farmers. The proportion was 44% for Nueva Ecija and much smaller for the other provinces. The provinces are thus differentiated with respect to the ratio of seed obtained through the market vis-avis the ratio of seed obtained through the government. A relatively strong private demand for the new varieties is indicated in the cases of Pangasinan and Tarlac. It may be noted that aside from Laguna these two provinces are the diffusion leaders in Luzon.

In the case of irrigated farms, the variable PUMP (one if pump-irrigated, zero otherwise) is also introduced. We recall that among Liao's (1968a) empirical production functions, the coefficient of his adoption index was greatest for the pump-irrigated town Cabuyao.

2.5 The marketed surplus

It is commonly stated that agricultural modernization is a function in part of the degree of market orientation of the farmer. This degree may be measured in various ways from output disposition data:

DISPOS = Total palay disposed

OUTLL = Landlord's share

OUTPY1 + OUTPY2 + OUTPY3 = Payments in kind to harvesters, threshers, etc.

DEBT = Debtors' share

DISPOS - (OUTLL + OUTPY1 + OUTPY2 + + OUTPY3 + DEBT) = Operator's share

OUTSEL = Quantity sold by operator

It is not clear whether a significant portion of the payments in kind received by harvesters, etc., enters the cash market, while it seems certain that most of the land rents and the debt repayments do. Hence the <u>gross marketed surplus</u>

ratio probably falls within the limits given by (1) MQ1 = = (OUTLL + DEBT + OUTSEL)/DISPOS, (2) MQ2 = (OUTLL + DEBT + + OUTSEL + OUTPY1 + OUTPY2 + OUTPY3)/DISPOS.

The operator's marketed surplus ratio may be defined

$$MQ3 = \frac{OUTSEL}{DISPOS - (OUTLL + OUTPY1 + OUTPY2 + OUTPY3 + DEBT)}.$$

And the marketed surplus ratio for both operator and landlord is

$$MQ4 = \frac{OUTLL + OUTSEL}{DISPOS - (DEBT + OUTPY1 + OUTPY2 + OUTPY3)}$$

Under these definitions, owner-operators shall of course have MQ3 = MQ4. These two variables are alternately introduced into the diffusion functions to determine which provides a specification which is "better" in the statistical sense. Under the hypothesis that landlords have an important share in decisions regarding varietal acceptance, and if farm commercialization is a significant determinant of adoption, then MQ4 would be expected to register greater statistical significance than MQ3.

3.0 EMPIRICAL ANALYSIS

3.1 Methods

The varietal diffusion variable may be represented either by a dummy (1 for adoption of new varieties, 0 otherwise) or by a proportion (of area, of farmers, etc.) referring to new varieties. Both are cases of limited dependent variables[1] and are subject to certain methodological problems. Considering the large number of independent variables to be accounted for, and the necessity of selecting certain strata among the data for separate analysis, the size of the sample did not seem large enough to warrant the formation of groups such that proportions could be computed from each group.[2] (Table 3.1 lists the number of barrios and Table 3.2 the number of farmers contained in the subsample.) The analysis following therefore employs the dummy dependent variable.

Ladd (1966) has shown that such a dependent variable may be used for estimating either probability functions or discriminant functions, and that the computations entailed in both cases are very similar and can be produced by an

---

[1]Cf. Goldberger (1964), pp. 248-255.

[2]Cf. Theil (1967), pp. 71-76.

Table 3.1

Number of barrios, subsample from 1968 Evaluation Survey

	Program[a]	Non-program
Bulacan and Laguna	30	29
Nueva Ecija and Tarlac	42	21
Pampanga and Pangasinan	50	37
	122	87

[a]Includes special project areas.

Table 3.2

Sample Sizes by Crop Types,
Extension Group, and Adoption of New
Rice Varieties in Crop Year 1968,
Six Provinces

	Rainfed Lowland	Irrigated Wet Season	Irrigated Dry Season	Row Totals
I. Program Cooperators and farmers in special projects				
Adoptors	17	73	61	151
Non-adoptors	52	145	56	253
	69	218	117	404
II. Program area non-cooperators				
Adoptors	7	22	21	50
Non-adoptors	143	114	29	286
	150	136	50	336
III. Others				
Adoptors	17	20	15	52
Non-adoptors	176	80	24	280
	193	100	39	332
Column totals				
Adoptors	41	115	97	253
Non-adoptors	371	339	109	819
	412	454	206	1,072

ordinary regression routine. A brief description of these functions, closely following Ladd (1966), may be warranted at this point.

The linear probability function is estimated by assuming

$$Y = i\beta_o + X\beta + u$$

where

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \quad Y_1 = i_1 , \ Y_2 = 0_2 , \ X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

and we have i = an $(N_1 + N_2) \times 1$ unit vector, i_1 = an $N_1 \times 1$ unit vector, 0_2 = an $N_2 \times 1$ vector of zeros, X_1 = an $N_1 \times K$ matrix, X_2 = an $N_2 \times K$ matrix, and u = an $(N_1 + N_2) \times 1$ error vector. There are thus K variables plus an intercept as explanatory factors. The equation is estimated by regression and termed a probability function because computed values of y are likely to (but not necessarily) fall between zero and one.

In a sample such as this, where $N = N_1 + N_2$,

(1) $\overline{Y} = N_1/N \quad , \quad S_{yy} = N_1N_2/N = n.$

Define $\overline{X}_1$ and $\overline{X}_2$ as the column vectors of means deriving from X_1 and X_2 , and define $d = (\overline{X}_1 - \overline{X}_2)$. The slope coefficients from this regression are

(2) $b = S^{-1}nd$

where $S = (X'X - (1/N)\ X'ii'X)$ and the intercept is

(3) $b_o = \overline{y} - b'\overline{X}$

where $\overline{X} = (N_1/N)\overline{X}_1 + (N_2/N)\overline{X}_2$.

As Goldberger (1964, pp. 249-50) has shown, the weakness of this procedure is that the usual assumption that $E(u) = 0$ presupposes a certain density function[3] for u such that its covariance matrix has diagonal elements $E(y_t)\ (1 - E(y_t))$. Goldberger has suggested a two-stage procedure: (1) estimate $E(y_t)$ as $\hat{y}_t$, the computed least squares value; (2) use $\hat{y}_t(1 - \hat{y}_t)$ as diagonal elements of an estimated covariance matrix W, and obtain the Aitken estimator $b_* = (X'W^{-1}X)^{-1}X'W^{-1}y$.

In practice, however, this suggestion is very often not feasible. The Aitken estimator requires diagonal terms of H where $H'H = W^{-1}$, and these terms would be $1/\sqrt{\hat{y}_t(1 - \hat{y}_t)}$. If, as often happens, y is outside the (0,1) range, then the corresponding diagonal term contains a complex number.

In discriminant analysis it is assumed that the population to which X_1 belongs is distributed as $N(\mu_1, \Sigma_1)$

---

[3]This function also suffers from the assumption that the 'true' vector $(i\beta_o + X\beta)$ has elements all in the (0,1) range.

and that the population to which X_2 belongs is distributed as $N(\mu_2, \Sigma_2)$, and that $\Sigma_1 = \Sigma_2 = \Sigma$. Denote the first population by I and the second by II. The problem is to classify a given sample unit represented by the vector X_* as belonging either to I or to II. It may be shown that, if the loss from misclassification into I equals the loss from misclassification into II, then the expected loss due to misclassification is minimized by use of the rule:

(4) If $X_*'\beta \geq \frac{1}{2}(\mu_1 + \mu_2)'\beta$, classify the sample unit into I, and into II otherwise,

where $\beta = \Sigma^{-1}(\mu_1 - \mu_2)$, and $X_*'\beta$ is termed the discriminant function. In order to distinguish as far as possible between the two populations from a given sample, β is estimated by maximizing so-called <u>between-group</u> variance relative to <u>within-group</u> variance. Setting $Z_1 = X_1\beta$ and $Z_2 = X_2\beta$, between-group variance is defined as $(\bar{Z}_1 - \bar{Z}_2)^2$ and within-group variance as $(Z_1'Z_1 - N_1\bar{Z}_1^2 + Z_2'Z_2 - N_2\bar{Z}_2^2)$. It is clear from the rule in (4) that it is sufficient to estimate any scalar multiple of β, and it may be shown that the estimator $b = S^{-1}nd$ (cf. eqn. (2)), which satisfies the maximization criterion, is such a scalar multiple. In practice, the classification rule is then:

(4') If $X'b \geq \frac{1}{2}(\bar{X}_1 + \bar{X}_2)'b$, classify the sample unit into I, and into II otherwise.

Thus the parameters of the discriminant function can be estimated by the least-square <u>slope</u> coefficients of the linear probability function.

The following tests of significance are available:

a. The significance of the discriminant function is tested by the same statistic which tests for the significance of regression, viz.,

$$(5) \quad F_1 = \frac{R^2}{1 - R^2} \frac{N - K - 1}{K}$$

where R^2 is the coefficient of determinant from the regression.

b. To test whether a set of q variables makes a significant contribution to either the discriminant function or the regression, after a set of p variables has already been included, one may use:

$$(6) \quad F_2 = \frac{R^2_{p+q} - R^2_p}{1 - R^2_{p+q}} \frac{N - p - q - 1}{q}$$

where the subscript to R^2 indicates the set of variables included.

c. To test the hypothesis, for the discriminant function, that $\beta = A$, A being a vector of preassigned numbers, one may use:

$$(7)\quad F_3 = \frac{R^2 - R_c^2}{1 - R^2}\,\frac{N - K - 1}{K}$$

where $R_c^2 = A'd$.

The probability of misclassifying a sample unit may be estimated by computing

$$(8)\quad t = \frac{1}{2}\left(\frac{F_1}{n}\,\frac{K}{N - K - 1}\right)$$

and finding the probability of obtaining a t-statistic, with $N - K - 1$ degrees of freedom, at least as large as this.

A probability function may be used for discrimination, say by the rule that if $b_o + b'X_* \geq .5$, then classify the sample unit into I, and into II otherwise. A discriminant function may also be used to compute what may be interpreted as a conditional probability:

$$(9)\quad P_* = (b_o + b'X_*)/2y_z$$

where $y_z = b_o + b'(\frac{1}{2}(\overline{X}_1 + \overline{X}_2))$; thus if $X_* = \frac{1}{2}(\overline{X}_1 + \overline{X}_2)$, then the probability of event I taking place is computed as .5.

Ladd states (p. 884):

> "The basis for choosing between [the probability function and the discriminant function] should be largely distributional. If one knows, and is willing to assume, nothing about the distribution of the X's, but is willing to assume the u_t to be independently distributed with mean $E\ u_t = 0$ and variance $E\ u_t^2 = \sigma_t^2$, it would be appropriate to apply the two-stage procedure suggested by Goldberger. This procedure suffers from the disadvantages inherent in using the same set of data twice.
>
> "If one knows, or is willing to assume, that the X's come from two normal populations with common covariance matrix, yet different means, but has no idea on the distribution of the u_t, discriminant analysis is appropriate."

Discriminant analysis would thus seem more appealing for the study at hand, except that some of the variables in X are unavoidably dummy variables themselves, such that the assumption that X is drawn from two normal population gets rather tenuous. Note should be taken of these several problems when reviewing the regression results.

Table 3.3 lists the explanatory factors to be considered. The sample was first classified into groups using the several combinations of crop type, extension group and soil ranking. Then probability/discriminant functions were computed with farm status variables, land tenure etc. as independent variables. Since many of the independent variables are zero-one dummies, tests of normality for the inde-

Table 3.3

Summary of Variables

Stratification variables	
3 Crop types	
3 Extension groups	
Soil ranking (a. 4 soil series ranks; b. 3 soil texture ranks)	
Dependents variables	
NEW1(68)	Dummy for use of new varieties in 1968
NEW2	Dummy for planned use of new variety in 1969
Farm status	
AGE	Ten-year units, with 1 = 20 years approx.
AIN	Area planted, hectares
SCHOOL	Number of years of formal schooling
EXPERT(67,68)	Ratio adopted of 7 recommended farm practices, new variety use excluded
Land tenure	
OWNER	Dummy for owner-operatorship
Credit and farm irrigation	
INTRST	Interest rate
PUMP	Dummy for pump irrigation
Marketed surplus	
MQ3	Operator's marketed surplus ratio
MQ4	Operator-landlord marketed surplus ratio

All variables are dated crop year 1968 except where indicated (see EXPERT).

pendent variables do not seem too meaningful; the discriminant functions were estimated on the assumption that departures from normality are not excessively serious.

The strata-determining variables are listed in order of a priori importance. If coefficient-vectors were not different according to the Chow test, regression data were pooled, first across soil strata, then across extension groups. Crop types were not pooled on a priori grounds. The soil strata are a priori the weakest because the soil data are from a secondary source.

Determination of the maximum number of meaningful strata may be considered one level of analysis. A second level deals with the relative explanatory contribution of each of the sets of variables, in which case the relevant test statistic is F_2 as given in eq. (6). A third level deals with the relative explanatory contribution of each of the variables within a set X_j, in which case the simple t-ratio and, again, F_2 may be used as test statistics. It may be noted that some of the variables within a set are substitutes for each other, e.g., the farm status variables SCHOOL and EXPERT, such that one function of these tests is to determine the most useful substitute.