PN-AAN-564
ISN: 31766

THE CONSEQUENCES OF SMALL RICE FARM MECHANIZATION PROJECT

Working Paper No. 12

Modelling the Consequences of Future Mechanization: An Outline of Possible Procedures

by

John A. Wicks

AGRICULTURAL ENGINEERING DEPARTMENT
International Rice Research Institute
Los Baños, Laguna

October 1979

The Consequences of Small Rice Farm Mechanization Project is supported by the United States Agency for International Development under Contract tac-1466 and Grant No. 931-1026.01 and is being implemented by the International Rice Research Institute and the Agricultural Development Council, Inc.

The Consequences of Small Rice Farm Mechanization Project is a cross country study of the impact of engineering technologies on agricultural output, employment and income distribution. The project, a centrally funded activity of the United States Agency for International Development, has two major components. The first is a series of case studies addressing specific topics relating to the application of mechanization. The majority of the case studies are funded through awards from the Agricultural Development Council to graduate and post-graduate research personnel in the Asian region. A second component is a three country comparative study being administered by the International Rice Research Institute in cooperation with agencies in Thailand, Indonesia and the Philippines. This portion of the project consists of a series of stratified cross sectional surveys and daily recordkeeping activities at sites in the above countries. Details of the field design and research methodology are contained in the operations manuals noted on the last page of this report.

Reports contained in the Working Paper Series are released without formal review or editing to facilitate communication of preliminary results and to elicit constructive comment.

Additional copies of the reports listed may be obtained by writing to:

Mechanization Consequences Project
Agricultural Engineering Department
International Rice Research Institute
P. O. Box 933
Manila, Philippines

or

The Agricultural Development Council, Inc.
1290 Avenue of the Americas
New York, New York 10104
U.S.A.

MODELLING THE CONSEQUENCES OF FUTURE MECHANIZATION: AN OUTLINE OF POSSIBLE PROCEDURES

J.A. Wicks

One of the principle objectives of the Consequences of Mechanization Project is to integrate the data collected into a model(s) which can subsequently be used to make projections of the impact of mechanization on employment, incomes and income distribution. It is the purpose in this paper to present some of the techniques which appear to have potential for achieving this objective. Rather than attempting to provide a complete review of all possible techniques, a brief summary is first made and attention then devoted to developing the most likely candidate. However, this approach should not be assumed to exclude the alternatives from future consideration.

Before becoming engrossed in model structure, it is expedient to examine the exact role of the proposed models. Within this context there are several points requiring clarification.

First, statements such as "this model will be used to predict changes that will occur in the agricultural sector" are frequently made when discussing model development. Such statements attribute a quality to the model which it never possesses. The most that can ever be expected of a model is the capacity to generate predictions of the future given a large number of underlying assumptions. Such predictions may be more adequately called "conditional predictions". Moreover, it is imperative that the underlying assumptions both of the model and of each conditional predictions are clearly stated before attempting to draw any conclusions.

Only then can comparability between models and the validity of conclusions be assured.

Second, why attempt to model the future and all of its uncertainties? After all a large volume of historic data will be collected, and there will be a lot of work involved in analyzing the data in a historic context. The answer to this question must surely be that there is only one scenario, set of events, which can be examined directly from historic data. The scenario comprises the events which occurred during the period. Further the possibility of these events recurring ranges from highly unlikely to impossible. Thus a role becomes clear for a model which can be used to project the future subject to a range of alternative assumptions. However, it is critical that a credible model be constructed from the available data.

Third, there is the problem of whether the model should be normative or positive. That is should the model builders attempt to formulate an appropriate set of norms and then develop a model corresponding to these, or should they attempt to build a model which is capable of simulating what farmers actually do. Setting aside the tendency in normative model formulation (e.g. Lin et al. 1974) for norms to be set in such a way as to tend to force the model to mimic farmers, and the opposite tendency in positive models, that is to incorporate an assumption of profit maximization, it is useful to examine the costs and benefits of these different types of model. Quite naturally, the normative model requires that a set of norms be established prior to formulation and solution of the model. In particular, such norms relate to the objectives of the farmers, be they profit maximization, utility maximization or cost minimization. The only

manner in which such norms can be identified, and quantified, is through consultation with the farmers involved, a process which is very complex and exacting. In terms of data requirements for the normative model, problems are not likely to be too severe. Indeed it is possible that the most preferred position is one where there are sufficient data to enable a reasonable estimate of appropriate coefficients, but not so much as to encourage great reliance on apparently meaningful average coefficients. When such coefficients are used to construct an "average" model, the results may well have little real meaning (Upton and Casey 1974). A more meaningful way to use the data is by intelligent use of parametric facilities. Finally, although relatively simple to formulate, the solution procedures for normative models are frequently complex and require substantial computer facilities. Hence simulations using the model can be time consuming and costly. By contrast, the positive model does not require prior definition of norms. However, data requirements are likely to be more complex, since the estimation procedures require sufficient observations for determination of significant and consistent coefficients. The estimational procedure of these coefficients may be prolonged and complex. However once estimated, use of the model for simulation exercises is usually a relatively simple matter, providing care is taken that exogenous variables are set within an acceptable range.

The fourth area for clarification concerns the appropriate temporal aspects of the models. That is should the model be structured as single period, say one year, or multiperiod, perhaps five years. Moreover if the model is for a single time period, should it be the next period or a period some time in the future. Prior knowledge of implications of the lumpiness of machinery investments, the time taken to recover the cost

of the investment, and the expectations framework typically accredited to farmers, all suggest the superiority of a multiperiod model. To its detriment is the fact that such a model is always far more complex to formulate and solve than a single period model. In addition, the cross-sectional structure of the data set limits the specification of multi-period model. Links must be established between time periods, and these might be unsatisfactory with the given data set. The alternative options do not appear very satisfactory either. Modelling the adoption procedure for mechanization within a single year, be it next year or some year in the future, appears to be an unrewarding task, since the critical inter-temporal effects must be ignored. One possible way to overcome some of these problems would be to construct a simple recursive model (see e.g. Day 1963). The recursive model would provide links between periods but would have to be expanded further to incorporate multi-period decision making. This matter is discussed in subsequent sections.

To move to the fifth point requiring discussion let us consider the scope of the proposed models. Theoretically, we might attempt to build models at one or more of the single farm, village, region or multi-region level. Without doubt such models would likely be of great interest to model builders and possibly to some users. However, some cautionary notes are in order. Unless extreme care is taken the models will develop to a level of complexity governed only by the available computer facilities. For multiregion models this situation is easily reached (see e.g. Wicks et al. 1978) without the model ever simulating the level of detail required for realistic policy analysis. A further point is that the models constructed should be consistent with the type and orientation of the data being collected. Since formulation of the models prior to

specification of which data was to be collected was not possible, it is now necessary to formulate the models in such a way that they make the greatest possible use of the available data, and permit the most relevant analyses to be conducted. Inevitably there will be significant gaps which will only be closed by conducting additional surveys. These data and manpower requirements rule out the possibility of constructing meaningful regional and multiregional models. Hence the emphasis for the remainder of the paper will be on single farm and village level models.

With the above points duly noted the following sections provide some discussion of possible analytic models. In order to limit the ultimate length of the paper, discussion has been restricted to those formulations (one for each section) which appear most applicable to our situation. However, this is not intended to limit the scope of subsequent discussions as it may be rewarding to discuss alternative approaches.

Before taking the first altering steps towards construction of an edifice to the proposed model builders' ingenuity, let us recall that each of the models represents a form of simulation. Thus it is appropriate to recall the three laws of simulation promulgated by Dillon (1971).

(i) Simulation, like statistics, cannot prove anything.

(ii) Simulation, like statistics, can nearly prove anything.

(iii) Once started, simulation will continue until available funds are exhausted.

Also let us not forget Anderson's three additional hypotheses (Anderson 1974):

(i) Every simulation study has its trenchant critics

(ii) The more aggregative the simulation, the more liable it is to criticism.

(iii) Study through simulation always absorbs more resources than anticipated *a priori*.

MODELLING THE INDIVIDUAL FARM: A NORMATIVE APPROACH

To facilitate presentation this section is limited to mathematical programming and related techniques. The techniques are briefly outlines in the first part of section, and the most appropriate selected. The data demands of that model are then examined further in the subsequent subsections. At this point finality of the choice of technique is not critical since most of the alternative options require very similar data sets and assumptions.

Alternative Specifications for the Model

Linear Programming is perhaps the simplist form of normative model which we might wish to consider. The linear programming model may be simply expressed as

$$\text{Max } U = c'x \qquad (1)$$

$$\text{Subject to } Ax \gtreqless b \qquad (2)$$

$$\text{and } x \geqslant 0 \qquad (3)$$

where x is a vector of activities

A is a matrix of input-output coefficients

b is a vector of resource availabilities

and c is a vector of gross margins per unit of activity

If appropriate, a more general formulation may be obtained by defining c as a vector of utility per unit of activity. However, the linear programming model remains constrained by the assumptions of linearity, divisibility, independence (additivity) and certainty. Briefly linearity implies that there must be no economies of size for activity levels and that the marginal utility of increasing the value of the objective function must be constant. These assumptions may be relaxed in certain cases through use of linear segmentation and seperable programming. The assumption of divisibility is more troublesome since it requires that activity levels can be divided into infinitely small amounts. For packages of land and items of machinery such as assumption may cause many problems. Furthermore, attempts to get around the assumption by forcing in equalities are fraught with danger. Independence implies that the total resource requirements and the total product of a number of activities taken together are simply the sums of their individual inputs and outputs. If the assumption needs to be relaxed constraints may be formulated to tie the activities together, but care must be taken to avoid altering the structure of the model. Finally the assumption of certainty implies that all elements of A, b and c are known with certainty. Such is rarely the case in reality.

Throughout the remainder of this section some alternative models, which represent generalisations of the linear programming model, are presented.

Risk Programming. Within the single category of risk programming all formulations which allow relaxation of the certainty assumption for one or more of A, b and c are grouped. A review of relevant techniques

is not sought here since they are provided in Anderson et al. (1977) and also by Wicks and Guise (1978). Although increasing the realism of solutions obtained, the use of even the simplist of these techniques calls for the availability of substantially more data and knowledge of the planning environment. However there appear to be few computational problems except in the case of extremely complex nonlinear problems. All others can be solved by linear or quadratic programming algorithms, both of which are readily available for most computers.

Integer Programming provides a technique through which the previously stated assumption of divisibility may be relaxed. Intuitively the integer programming model has great appeal for the current problem, where machinery has such a critical role. To its detriment there is a need for care since the search for an optimum integer solution is inevitably a complex problem, and many of the available algorithms are not particularly efficient. However, careful formulation of the model, and selection of the algorithm for solution, will do much to assist successful implementation.

Integer risk programming, at least theoretically, attempts to overcome many of the criticisms of linear programming by an amalgamation of the techniques discussed above. Possible approaches to defining a suitable algorithm are either by using a linear approximation of quadratic programming, within an integer programming algorithm, or by combining the principles of quadratic and integer programming into a single algorithm. Both of these techniques are theoretically feasible, but neither has been tried to any extent.

Monte Carlo Programming was developed in the 1960's in an attempt to overcome many of the then limitations of mathematical programming (for details of the procedure see Carlsson *et al*. 1969 and Donaldson and Webster 1968). In their formulation Donaldson and Webster indicated that the Monte Carlo model could readily accommodate integer values. More recently Anderson (1976) has extended the model to include risk. In all cases the Monte Carlo approach provides an almost optimum, rather than an optimum solution, whether this is acceptable, must be decided by the individual analyst. However, the number of solutions typically required to ensure that the best is almost optimum is so great that the availability of computer resources tends to restrict applications.

Multiperiod versions of the above models. As was indicated earlier, one of the limitations of the linear programming model is that it yields short-run equilibrium solutions. No account can be taken of longer run plan or the feasibility of the adjustment process. Multiperiod models of the above formulations can overcome these problems, although at a cost in additional computing time. The model may be summarized as

$$\text{Max } U = c_t' x_t + (1-\lambda) c_{t+1}' x_{t+1} + \cdots + (1-\lambda)^n c'_{t+n} x_{t+n} \qquad (4)$$

$$\text{Subject to } A_{1t} x_t \gtreqless b_t \qquad (5)$$

$$A_{2t+1} x_{t+1} \gtreqless b_{t+1} \qquad (6)$$

$$\vdots$$

$$A_{nt+n} x_{t+n} \gtreqless b_{t+n} \qquad (7)$$

$$\text{and} \quad x_t, x_{t+1}, \ldots\ldots x_{t+n} \geqslant 0 \qquad (8)$$

where A, b, c and x are defined the same as in the linear programming formulation,

subscript t refers to the current time period, t+1 to the next time period and so on up to n time periods ahead,

and λ is rate of time preference for the components of the objective function.

In addition further activities and constraints are introduced to permit the flow of cash, land, machinery and other resources through time.

The value of n to be used in any model has been subject to debate (Modigliani 1975), but determination of an appropriate value remains highly subjective. Within the current analysis five years would appear to be a reasonable compromise between what can be achieved and the desirable period. Since computational requirements increase roughly in proportion to the cube of the size of the matrix, this extension is likely to greatly increase the problems of optimizing the model.

Recursive programming is often presented as a technique through which the problems of matrix dimensions associated with multiperiod programming may be eliminated. The recursive model comprises a set of single year optimizations connected by an independent set of linkage equations. The flexibility thus obtained allows the user to change model coefficients or prices during execution of the job, thus imparting considerable reality to the solutions achieved. However, this flexibility is not achieved without some cost. Most importantly the option solutions derived are short run optimum for each of a series of single

years rather than a single long run optimum. Consequently, as prices change the short run optimum is subject to considerable fluctuation and this may cause drastic changes in plans. Unless some preventative action, such as the inclusion of flexibility constraints, is taken the sequence of model results frequently implied may be logically unacceptable.

Recursive Multiperiod Models. One important criticism of the multiperiod model is that, in spite of its complex structure, the results obtained actually indicate the appropriate decision for the first planning period consistent with a long run expected optimum (Modigliani 1951) and not the long run plan. To explain further, once the first time period is complete, actual results will be available to the farmer, and his actual level of resource availabilities for the next time period will differ from those previously expected. Hence the optimum long run plan will likely have changed, and the appropriate strategy for the next planning period will have to be recomputed. Intuitively, a combination of the principles of recursive programming and a multiperiod model into a single formulation should overcome many of the conceptual problems of the simpler models. Additionally, selection of the most appropriate optimizing algorithm for inclusion into the model will improve specification. This greatly expanded formulation appears to have many of the attributes of real world decision making, including a considerable degree of complexity in specification. Use of the approach has been extremely restricted to date. For one example see Chien and Bradford (1976), who incorporated a linear programming model into such a framework.

Doubtless there are other formulations which might be considered

as potential candidates for a normative model, but the above range appears to cover an adequate range of scope and complexity.

Identification of a Suitable Formulation

In the previous subsection, a number of different potential formulations were outlined. This number would be greatly increased if all of the twists were to be included. The purpose in this section is to go at least some way towards defining a suitable formulation for normative models envisioned. As an aid to the decision making process some criteria for selection of the preferred formulation are now presented.

1) It must be possible to construct the major part of the model from the data sets currently being collected. Although some further survey work may be required this should not be in the form of an extensive survey.

2) The model developed must appear potentially capable of providing meaningful solutions to the types of questions which engineers and economists will wish to ask.

3) It should be possible for reasonably well educated economists to at least understand the underlying principles of the model. Hence they should be able to carry out simulations with it, make minor adjustments to reflect particular situations, and sensibly interpret model results. The model should not be a black box.

4) The model to be developed should not exhaust available computer and personnel resources. This implies both the possible simplication of the framework if the selected methodology becomes unwieldy and attention to computational aspects of the model which may well yield large benefits.

5) It should be possible to develop, test and implement an operational model within the life of the project.

In Table 1 we attempt to evaluate the extent to which each of the above criteria are likely to be met, for each of the previously discussed models using a scale from 1 to 5. A value of 1 represents very likely to be met whereas 5 represents very unlikely to be met. Naturally, all of the values contain elements of the model builder's bias. The subjective valuations are then summed, using two different procedures, in order to derive subjective measures of the value of the approaches. In the raw total all criteria are weighted equally, whereas in the weighted total criterion number 2 is weighted to be as important as all other criteria combined.

The ordering of the options may be readily derived from Table 1. Salient points are that the top eight options are separated by four points in the raw total and five points in the weighted total. Linear programming (LP) and Recursive Integer Risk Programming-Linear Approximation (RIRP-LA) are the only two options to appear in one list but not in the other. The other seven options merely change rank between the listings. Hence the choice of an appropriate methodology for modelling the adoption of machinery remains unclear and it appears appropriate to make a personally biased assessment. The assessment is that we should proceed with developing a model using multiperiod recursive integer programming. This is selected on the basis that the model will be constructed in stages, and if it becomes too complex we can always stop at an earlier stage.

The Multiyear Methodology

The multiyear methodology has been outlined elsewhere by Chien and Bradford (1976), and essentially the same approach will be followed in the current study. The most important changes which we intend making are the specification of integer values for some activities and the possible use of a non-linear utility function as the maximand. These are both discussed later.

In Figure 1 the structure of the "multiperiod" component of the model is shown. It comprises a series of single year activities and constraints linked between successive years by transfer activities and constraints. Apart from in the first year resource availabilities are predominantly zero, the actual level being determined by transfer from the previous year. The objective function of the model is a summation of objective functions for individual years, weighted in the manner described earlier. Each of these topics is treated in more detail in subsequent sections. Solution of the component is undertaken each year and provides a plan for the current year which is consistent with long run utility optimization subject to current expectations.

In Figure 2 a generalized flowchart of the recursive programming model, incorporating the multiperiod model as the optimizing component is presented. To briefly follow through the processes of the model, when a run is initiated the value MAXN is set to define the number of years in the run, and N is set equal to one, the first year. A multiperiod integer programming model is then constructed according to a set of predefined rules. These cover the initial stock of land, labour and machinery, the set of potential activities, and expected

prices and yields. In order to solve the problem a starting basis is defined. This may be the conventional first basis or, as experience is gained, may be an alternative basis. The multiperiod model can then be optimized and the relevant planned values for the first year printed.

We now move into the simulation portion of the model. The farm plan for the first year is subjected to a set of randomly selected influences. These are drawn from known underlying distributions for factors such as labour required, machinery breakdowns, yield, price and climatic effects. The result of the simulation is set of "actual" outcomes, which differ from the planned outcomes. These are printed for purposes of later comparison. N is then checked to see if it is equal to MAXN. If not, N is incremented by one and the multiperiod model updated to reflect a change in year. A new final year for the model is also constructed. Initial resource levels are updated to reflect the changes that have occurred during the year, as may be other coefficients. The multiperiod model is then re-optimized starting from an (adjusted) current basis.

This process is continued for the required number of years. Given a well documented model and well thought out coefficients, it should be possible to answer many of the questions being posed by economists and engineers without a large number of complex modifications to the basic model.

The next two subsections outline the specification of a single multiperiod, optimization model and the between year linkages in greater detail.

Specification of the optimization model.

A single year. At this early stage in defining the model structure it hardly seems appropriate to become involved in defining specific activities. Rather, since we are still to confirm the exact form of the model, it would be better to examine particular groups of coefficients. As a first step a diagrammatic structure of one year of the multiperiod matrix is shown in Figure 3. At this stage, no facility has been incorporated for loans but they could readily be incorporated through the cash section. The matrix is the same as a conventional LP matrix except that fixed costs are included, some adjustments are required for utility optimization, and some of the variables (.e.g 2-wheel and 4-wheel tractor ownership) will be included as integers. By contrast tractor hire is non-integer.

The general scheme of the model is that one integer unit of tractor will supply a fixed maximum level of resources. Up to this level of resources can then be combined with certain types of labour to produce animal equivalents of power. The animal equivalents can be used in crop production. Similarly, labour for non-mechanized aspects of production will be utilized directly by the crop. Unfortunately the methodology prevents capture of some of the believed quality advantages of mechanization. However, it does constrain the size of the matrix from increasing too rapidly. As a next stage it is planned to expand some of these portions of the matrix in order to establish the exact requirements of the matrix. Final determination of coefficients will not be possible until towards the end of the year, unless the model is constructed on a seasonal basis. Note that no livestock activities are included at this stage.

Consideration of the problems associated with the objective function are included in this description of the multiperiod model, since they relate largely to optimization over time.

Interyear linkages. Each of the resources available on the farm at the end of the year must be transferred to the following year for use in future production. This is achieved by a set of transfer constraints which transfer the closing stock of resources in the current year over to the next year where they become an openning stock. The openning stock may be increased or decreased by purchase and sale to obtain the stock to be used for crop production. Intuitively it would be appealing to incorporate the aging process into the flow, so that a tractor which is n years old this year becomes n+1 years old next year. Changes in performance of the machine could then be modelled. Such attention to detail will need to be investigated but may cause the matrix to expand beyond control.

The multiperiod model. Once the many single periods are linked together by the transfer constraints we have the essentials of a multiperiod model. Whether the single periods should be years or cropping seasons is open to debate. Much will depend on the flexibility desired in the actual model and anticipated time horizon of decision making by farmers. In order to get a model working it may be advantageous to start with an annual model and then shift to a seasonal model if any benefits appear likely. Eventually the optimization period could be defined in a flexible manner and implemented by particular requirements.

We are now left only with the problem of the objective function.

In Figure 3 this was indicated for one year and no attempt was made to define actual values. However, the implication was of monetary values and profit maximization. Some extension of, and problems with, the model are now considered.

Models are frequently constructed on the assumption that farmers have perfect knowledge of future prices and yields. For yields an average of past values, if sufficient data are available may be adequate, but for many types of rice crop this would not be so. Perhaps it would be preferable to gather such data as are available and to adjust these according to prior knowledge. Expected prices are much more complex to determine. Under the current situation of rapid inflation, a simple average of the prices for recent years would not be accurate. Some alternative approaches need to be investigated. One would be the use of a Nerlovean Adaptive expectations model such as that adopted by Kingma and Kerridge (1977) in their model of the farm firm. In their model expected prices were defined as

$$E(P_t) = 0.5\ P_{t-1} + 0.33\ P_{t-2} + 0.17\ P_{t-3} \qquad (9)$$

and

$$E(P_t) = 0.9\ P_{t-1} + 0.1\ P_{t-2} \qquad (10)$$

for livestock products and grains, respectively. A somewhat more complex hypothesis is that expected price may be expressed as a function of farmers' prior experience and a government forecast. The model then becomes

$$E(P_t) = \alpha_1 F(P_{t-1}, P_{t-2}, \ldots P_{t-n}) + \alpha_2 F_p \qquad (11)$$

where F_p is the forecasters expected price

and α_1 and α_2 are scaling coefficients such that $\alpha_1 + \alpha_2 = 1$

The portion of the equation relating to the impact of past prices on expectations of farmers may assume many different functional forms. Clearly the values of α_1 and α_2 will vary over time, depending on the performance of the forecasters.

Without doubt quantification of any of these models would be extremely complex, and require data beyond that currently available. The only feasible alternatives may be to subjectively assess expected prices or to parametrically vary prices.

We now move on to the topic of defining farmer's objectives. The methodology involved will initially be developed for a single year model and the extended to a multiperiod model. Most similar studies have assumed that the objective is either profit maximization or utility maximization, where utility is defined within an Expected Income-Variance of Income (E,V) framework. For the current study it was suggested earlier that the benefits from incorporation of risk were not likely to be sufficient to justify the costs. Hence profit maximixation remains as the initial objective, and one which can be readily implemented.

However, from prior knowledge of the farmers with whom we are working, we anticipate that profit maximization is not the sole, and frequently not even the most important, objective. For example leisure time, the employment of local landless labourers and the ownership of machines may all add to a farmer's total utility. Identification and quantification of these attributes of the utility function is a complex problem. Indeed, it might be best tackled as an individual research problem. Therefore we now concentrate on the type of formulation which will be amenable to incorporation in the multiperiod model. This follows, and extends, the outline of Rae (1971).

As a first step those attributes which describe utility functions of farmers, in each of the countries, must be formally identified. Such research could benefit greatly from the expertise of sociologists and anthropologists.

The second step involves quantification of the major attributes identified in the first step, and incorporation of the function as the objective function of the multiperiod model. Development of the methodology here draws heavily on Anderson, et al.(1977, Chapter 4), although much of the detail has been omitted.

First let us assume that the utility function for one period of the multiperiod model may be written as

$$\theta = U(a_1, a_2, \ldots\ldots, a_n) \qquad (12)$$

where a_i is the i'th utility attribute and these are a total of n such attributes.

If it is possible to show, or reasonable to assume, that these attributes are jointly preferentially independent and utility independent for all n attributes then estimation of the function may be achieved by the "quasi-separable" approach, i.e.

$$U(a_1, a_2, \ldots, a_n) = f[U(a_1), U(a_2), \ldots U(a_n)] \qquad (13)$$

where $U(a_i)$, scaled from zero to one, is a utility function of the i'th attribute depending only on that attribute. Where utility independence does not hold it may be possible to combine two or more attributes into a single attribute, and to proceed in the same manner.

All that remains is to estimate a set of n scaling factors k_i,

$i = 1, \ldots, n$ which will scale the individual attributes in accordance with their relative importance, together with a second scaling factors k. Each of the k_is ranges between zero and one. In addition total utility is scaled to range between zero and one. These will yield a function either of the additive form

$$U(a_1, a_2, \ldots, a_n) = \sum_{i=1}^{n} k_i U_i(a_i) \tag{14}$$

$$\text{when } \sum_{i=1}^{n} = 1 \text{ and hence } k = 0$$

or of the multiplicative form

$$U(a_1, a_2, \ldots, a_n) = \{\prod_{i=1}^{n} [Kk_i U_i(a_i)+1]-1\}/K \tag{15}$$

$$\text{when } \sum_{i=1}^{n} k_i \neq 1 \text{ and hence } k \neq 0$$

In the multiplicative form K and may be estimated by considering equation (15) with all attributes at their most preferred level we get

$$K + 1 = (Kk_i+1)(Kk_2+1) \ldots (Kk_n+1) \tag{16}$$

Since k_i $i = 1, \ldots, n$ are know, and for $\Sigma k_i > 1$ we must have $-1 < K < 0$, and for $\Sigma k_i < 1$ we must have $0<K<\infty$, we can solve the equation. Unfortunately the multiplicative form, unlike the additive form, is not amenable to the linear programming formulation. An alternative approach must therefore be sought.

A brief examination of the logic of the structure of these utility functions indicates that the additive form of the function is rarely obtained. Hence as it appears computationally prudent to work with this

form, the only solution is to scale the k_i's to give

$$U(a_1, a_2, \ldots, a_n) = \sum_{=1}^{n} k_i' U_i(a_i) \tag{17}$$

$$\text{and } k_i' = k_i/\Sigma k \text{ for all } i \tag{18}$$

$$\text{such that } \Sigma k_i' = 1$$

As has been shown elsewhere (Huber 1974), the additive components of the model are usually so great as to render the approximation acceptable.

So long as all components of the function are concave, the additive function can be readily handled within a mathematical programming framework by using the grid linearization approach applied by Duloy and Norton (1975). To expand the function over several years requires an assumption of utility independence between years. Total utility is estimated for each year and a sum for all years derived[1]. Although this might also be scaled from zero to one there appears to be little benefit in so doing.

All that remains is to estimate the functions $U_i(a_i)$, such that they range from zero to one, and the k_i's. Estimational techniques are well documented (Anderson et al. 1977) and need not be reviewed here. Suffice to note that the negative exponential function

$$U_i(a_i) = 1-e^{-ca}, \; c > 0 \tag{19}$$

is probably the most appropriate form. Alternatives are the quadratic and logarithmic, respectively

---

[1]As an intermediate stage between profit maximization and full utility maximization, it may be prudent to estimate a form for the utility of profit only.

$$U_i(a_i) = a_i + ba_i^2 \tag{20}$$

and $$U_i(a_i) = a_i^c, \; 0<c<1 \tag{21}$$

It was mentioned earlier that the utilities for individual time periods could be aggregated to produce total, future, expected utility. However, reverting to the money argument it is clear that the aggregation is not a simple addition. This is due to the subjective rate of time preference for decision makers, a rate which must be reasonably assessed. To illustrate, suppose we were to offer a farmer the choice of $100 today or $100 one year from today, he would almost always, as we would expect, accept the $100 today. However, if the amount to be given today were to be gradually decreased a point would eventually be reached where the farmer would be indifferent between the money either today or in the future. From knowledge of this point the subjective rate of time preference may be computed. It should be noted that this will rarely be equal to the prevailing interest rate, and will typically be considerably higher.

With all of these components adequately specified the multiperiod model can be constructed. Attention is now directed to the recursive components.

Development of the recursive linkages

The purpose of the recursive linkages is to connect the plans for the current year to the outcome for the current year to the plans for the next year.

Once the plan for a given year has been implemented, the impact of a number of stochastic variables begins to take effect. These will include:

a) Machinery breakdowns,

b) Illnesses in draft animals,

c) Labour shortages,

d) Availability of machines and animals for hire,

e) Vagaries of the climate,

f) Random yield fluctuations, and

g) Price changes.

The effect of each will be modelled and used to obtain "actual" results from the plan. Linkage equations will be used to modify the levels of resource availabilities and farmers' price and yield expectations for the next period. It may also be possible to insert a random variable to control the availability of machinery in the next period. The presence of these stochastic variables will necessitate several runs of the model to obtain reasonable answers to any problem. Moreover, these answers will appear as ranges rather than single values.

Checking the sensitivity of the normative model

Well defined optimizing models of the type outlined above are frequently criticized for generating solutions which have little relevance to the real world. This follows from the fact that, although there are no other solutions better than the optimum, there may be many alternative solutions as good, or almost as good, as it. Furthermore, if these solutions are represented by plans having a vastly

different structure we may conclude that our "plans for development" are highly unstable. Hence care will be required when formulating recommendations.

Although the multiple runs of the multiperiod recursive model will assist in answering these questions, a more detailed examination would be of great value.

A potential approach would be to incorporate a Monte Carlo procedure (such as that developed by Donaldson and Webster, 1968) into the recursive model. An optimum solution to the multiperiod problem would first be obtained, using the methodology already outlined. Monte Carlo simulations would then be performed on the same matrix until a solution within a prespecified range of the optimum is located. This would be treated as the actual plan to run through the recursive sector in order to generate "actual" results with which to update the multiperiod matrix. The entire cycle could then be repeated. Carefully controlled experiments, with sufficient replications, would enable us to evaluate the stability of the solutions and alternative routes to development.

As a final note of warning, the methodology sounds relatively simple, but getting the model operational and actually generating results will almost certainly be far more complex.

MODELLING THE INDIVIDUAL FARM: A POSITIVE APPROACH

Up to now we have been concerned with how we might model the optimum, or almost optimum, response of farmers towards mechanization. In this section the problem will be examined from a different perspective.

Assuming that farmers make rational decisions, even if their rationale is not profit maximization or some definable form of utility maximization, the purpose now is to investigate whether an alternative model can be developed from the available data set. Most importantly the model will be defined as positive, that is describing what farmers actually do and subject to an unspecified underlying utility function. Although the estimated form of the utility function may derive from subsequent econometric analysis of the data, no attempt will made to pre-judge the appropriate values.

Having used appropriate econometric techniques to estimate the model, it is to be hoped that it can be used fairly readily for projecting changes at the farm level.

The model is proposed to contain the following components, each of which will be described in turn.

1) A Gross Household Income identity
2) A (Farm) Production Function for rice
3) Identities and functions to explain the observed levels of independent variables in the Production Function
4) Production Functions for other Crops and Livestock
5) A mechanism for distribution of Gross Household Income

Within the formulation it must be possible to link subsequent periods together, and the linkage must be compatible with the observed situation in our study villages. As is demonstrated later the link is currently proposed to be between the production of one season and the use or sale of that produce to generate Gross Household Income for use in the next season. Note here that the period for modelling is probably better considered as a season than a year, and hence at least

two sets of functions will be estimated.

The Gross Household Income identity

An identity may be specified for estimation of Gross Household Income available for use in season t. A reasonable form would appear to be

$$GHI_t = P(R_{t-1}).Y(R_{t-1})+P(OC_{t-1}).Y(OC_{t-1})+P(L_{t-1}).Y(L_{t-1})+OFI_t+NFI_t+INT_t+RL_{t-1}+LO_t$$

where GHI_t is the Gross Household Income available for use in the current season

P indicates price

Y indicates yield

R_{t-1} is rice in the previous season

OC_{t-1} is other crops in the previous season

L_{t-1} is livestock in the previous season

OFI_t is off farm income in the current season

NFI_t is non farm income in the current season

INT_t is interest on any savings received in the current season

RL_{t-1} is the rent received from land in the previous season. It is defined as zero if land is rented in, actual value if land is rented out and an imputed value for owned land used for crop production

LO_t is the value of any loans received in the current season

Since the relationship is an identity, there is no need for statistical estimation. It should be computed directly from the model. However, this implies that each of the righthand components must either be endogenously determined (e.g. the yield of rice) or exogenously set (e.g. the price of rice).

It will be noted that the right hand components are a mixture of values for the current season and values for the previous season. To determine exactly the correctness of these assumed temporal arrangements would be a complex problem, and so we should rely on local knowledge of the situation. Overall it does not appear unreasonable to assume that non-farm and off-farm income are spent at the time they are earned, whereas income from sale from crops is used in the next cropping season. This reflects the fact that income from sale of crops is rarely derived until the end of the cropping season at the earliest. However, these definitions may be confounded by the use of production loans.

Finally it is worth noting that the variable INT may be removed if it is found to be very low and that the variable RL_{t-1} has a very specific purpose. The latter is intended to provide a consistent definition of the model. Hence all land is assumed to be rented and those observations on owned land will contain balancing components in both parts of the model.

A (Farm) Production Function for rice

At this stage the estimation of a farm production function for rice, rather than a set of crop production functions is proposed. This is due to the type of data which will likely be generated as well as

the many logistical problems of estimating functions for individual parcels. Additionally, the use of more than one production function estimate would complicate the rest of the model and may lead to unnecessary problems in the subsequent simulations.

The usual form used for the production function estimates may be generally written as

$$Y(R) = f(L, P, La, F, S, I, O) \tag{22}$$

where Y(R) is the yield of rice

L is a measure of the land variable, area of value or maybe two variables

P is a measure of power inputs or stock, either machine or animal together with associated labour

La is all labour input less that used in conjunction with power inputs

F is fertilizer input

S is seed input, probably a dummy for type

I is irrigation input (dummy)

and O is a measure of all other inputs

All of these variables may be readily estimated from the survey data, in particular that on resource utilization. Estimation of the power component for land preparation will require aggregation of the four different types power, namely manual, animal, power tiller and four-wheel tractor. Alternative approaches would involve either the use of the purely economic criterion of cost (e.g. see Ramachandran 1979) or the use of physical equivalents, that is the

average number of man days required to cultivate one hectare. Because of necessary underlying assumptions neither method is completely satisfactory.

The representation of labour by a single independent variable within the model, presents a further set of problems. First there is the assumption that these coefficients can in fact be meaningfully aggregated. A strict interpretation of the aggregation is that if the amount of labour used for one process is increased, then the amount used for another process may be decreased by the same amount without influencing yield. This assumption does not appear particularly plausible. The problem may be overcome by specifying various disaggregated labour variables in the model. The second problem is that labour for harvesting is normally included in the aggregate labour variable and is hence assumed to influence yield. A more reasonable relationship would be to assume that labour required for harvest be a function of yield. The revised model thus obtained becomes a simultaneous equation system which may be represented as

$$Y(R) = f(L.\ P,\ La_{nh},\ F,\ S,\ I,\ O) \tag{23}$$

and $$La_h = f(Y(R),\ L) \tag{24}$$

where all variables are as before except that

La_{nh} represents one or more variables to define all labour except for harvest labour, and

La_h represents harvest labour

Because of the revised specification, harvest labour should be excluded from gross farm income.

The exact functional form to be used for the estimation is not yet clear. Much will depend on the ultimate quality of the data collected, and the estimational procedures available. Clearly many functional forms utter possibilities. Most widely used is the simple Cobb-Douglas which has both several beneficial and several detrimental features. Alternatively, there are the Transcendential Logarithmic Function (Forgenson, et al. 1973), the Constant Elasticity of Substitution (CES) function (Arrow, et al. 1971) and the Constant Ratio of Elasticity of Substitution Homothetic (CRESH) function (Hanoch, 1971). Each of these adds complexity to the estimational process and there is a need to first determine if there are likely to be sufficient additional benefits. On a different tack we should note recent work investigating the underlying assumptions of the error term in these equations, and their consistency with prior expectations (see e.g. Just and Pope, 1979). For example, in the Cobb-Douglas production function, it is implicit that the variance of output will increase with increasing input. However, for machinery and insecticide inputs we would expect the reverse. Perhaps pursuit of these problems at a later stage would be beneficial.

Suffice to suggest that for initial estimation, a simple function whether on the basis of per unit area or total area should be adequate.

Identities and functions to explain the observed levels of independent variables in the Production Function

Although it is relatively easy to define a blanket variable such as power input to production, an explanation of the breakdown is somewhat complicated.

The logical approach is to split the aggregated power variable (P) in the following manner.

$$P = P_n + P_o \tag{25}$$

where P_n = indicates hired power used

P_o = indicates the amount of owned power used.

Then we may define

$$P_n = La_n + A_n + M_n \tag{26}$$

and

$$P_o = A_o + M_o \tag{27}$$

where La_n indicates hire of labour for land preparation (n.b. physical preparation not operation of land preparation equipment),

A indicates animal input, and

M indicates machinery input.

Again the five variables on the right hand sizes of those identifies must be further examined to determine any effects. To suggest one set of possibilities we might investigate:

$$La_n = a_{11} + a_{12}\,(Pr_{La}/Pr_A) + a_{13}\,(Pr_{La}/Pr_M) + a_{14}L \tag{28}$$

$$A_n = a_{21} + a_{22}\,(Pr_A/Pr_{La}) + a_{23}\,(Pr_A/Pr_M) + a_{24}L \tag{29}$$

$$M_n = a_{31} + a_{32}\,(Pr_M/Pr_{La}) + a_{33}\,(Pr_M/Pr_A) + a_{34}L \tag{30}$$

$$A_o = b_1 TA_o \tag{31}$$

$$M_o = b_2 TM_o \tag{32}$$

where Pr_{La}, Pr_A and Pr_M are the unit area costs of land preparation by manual labour, animal and machine respectively, and

TA_o and TM_o are measures of the quantity of animals and machines owned, respectively.

In equations (28), (29) and (30) the relative prices of the relative prices of the various methods of land preparation are suggested as important determinants of the relative amount of the type of land preparation used. Farm size is also proposed as possibly having an effect. Before these equations can be estimated careful consideration will have to be given to their form since there may not be sufficient price variability to permit estimation. In the latter two equations we simply assume that use of machines or animals is a function of the amount owned. Further estimates are necessary to determine the amount owned. This might be accomplished by breaking the ownership variable down into the total amount owned at the end of the period t-1 plus changes in ownership in period t. This is represented by equation (33).

$$TM_{o,t} = TM_{o,t-1} + \Delta TM_{ot} \tag{33}$$

The total amount owned in t-1 is given exogenously but we must again determine a function for purchase of machinery. Except to suggest that there might be a relationship with relative prices, farm size and credit availability, we will proceed no further.

Returning to labour, the coefficient again consists of a number of variables. It may be expressed as

$$La = La_{hh} + La_{h} \tag{34}$$

where hh indicates household, and

h indicates hired

Because of the formulation used for the power input, labour used for power in land preparation will be excluded from the variable. Household labour may be determined exogenously, whereas hired labour will require additional estimation.

What is becoming clear from this outline is that it should be possible to disaggregate each of the independent variables in the production function to such a stage that logical causal relationships may be defined. However, the process is complex and much work is required to identify the appropriate statistical techniques. No doubt many problems will be encountered in the search for an adequate solution.

Production Functions for Other Crops and Livestock

In the two previous subsection a considerable amount of attention was devoted to estimation of plausible production function system for rice. Should we now pursue the same search for other crops and livestock? The answer to this question must surely depend upon the proportion of farm income generated by these activities. Intuitively it might be considered to be so small as to warrant nothing more than to be set exogenously, and any necessary adjustments made to the data set.

Alternatively, a simple form at

$$Y = \alpha X^{\beta} \tag{35}$$

where Y is output, and

X is a single aggregate input.

Distribution of Gross Household Income

In successive sections portions of the model, to determine the level of gross household income (GHI) and appropriate production functions, have been outlined. There now remains to tie these components together, and this is where problems are likely to appear.

Clearly gross household income must be allocated between various competing requirements. These may be summarized as:

1) Household expenditure,
2) Rice crop production expenditure, and
3) Other crop and livestock production expenditure.

Since the current study is predominantly concerned with rice production, it is only relevant to subdivide the second component into more detailed sections, namely

1) Expenditure on power (including the labour component),
2) Expenditure on labour (excluding labour for cultivation and harvesting),
3) Expenditure on fertilizer,
4) Expenditure on seeds, and
5) Expenditure on other inputs.

Seven categories of expenditures are therefore identified.

The most commonly encountered approach to derive an appropriate allocation follows the lines, "assuming profit maximizing conditions it may be shown that . . .". However, there is little to indicate that the farmers do follow profit maximization strategies, and hence the assumption would impose a nominative rather than positive framework on the ultimate model. All that it seems reasonable to assume is that, since the farmers are located in fairly small, closed communities, they might act in accordance with fairly common underlying utility function. Let us not try to define the objectives implied by the utility function but briefly to review some potential solutions to the impasse.

From consideration of the structure of the model it is clear that all of the GHI must be disbursed in some way. (Analysis of the data may reveal the need for a savings component to be included). Hence the system to be estimated becomes very tightly defined, but several solutions have been proposed.

Single Commodity Consumption Functions

The first possible approach follows a methodology similar to Lee and Philips (1971) who used the following system of simultaneous equations.

$$lrX_i = \alpha_{0i} + \alpha_{1i} (rY_p + U_i, \qquad i=1, \ldots, n \tag{36}$$

$$lrC = \alpha_0 + \alpha_1 LrY_p + V \tag{37}$$

where X_i is the value of consumption of item i

Y_p is the value of permanent income

C is the value of total consumption expenditure

Since Y_p is unobservable, the equations can be soleved for the observable variables X_i and C to give

$$La_{ni} = \alpha_i + \beta_i \ln C + W_i \tag{38}$$

where $i = \alpha_{0i} - \alpha_{1i} \cdot \alpha_0 / \alpha_1$

$\beta = \alpha_{1i} / \alpha_1$

and $W_i = U_i - \beta_i V$

Since lnC and W_i are functions of V they are correlated and the OLS estimate of β_i is biased and inconsistent. Lee and Philips present two possible approaches for overcoming the problem, one uses OLS and the other uses two stage least squares (TSLS). Furthermore, they show that a variable for household composition should be included.

In order to implement the methodology in the mechanization study we would need to obtain information on total income and total consumption for each household. However, it would not be necessary to obtain consumption expenditures for all categories of consumption since only those of direct interest would need to be estimated. An additional simplification of the model, which might be sufficiently realistic, would be to assume that all income is consumed. The system could then be reduced to the single equation:

$$\ln X_i = \alpha_{01} + \alpha_{1i} \ln C + u_i \tag{39}$$

which could be estimated by OLS. Other dummies could also be included to measure the differences between mechanized and non-mechanized farms.

The remainder of the systems outlined represent attempts to

improve upon the first approach by simultaneously estimating a complete expenditure system. However, the problems in estimation become successively more complex and the data requirements more restrictive.

The Linear Expenditure System (LES)

The LES, as set out by Howe (1977), may be summarized as follows. Suppose our farmers respond to a utility function of the form

$$U(X) = \sum_{i=1}^{n} a_k \ln(x_k - b_k) \tag{40}$$

with $a_i > 0$, $\Sigma a_i = 1$ and $x_i - b_i > 0$.

This is known as the Stone-Geary utility function and may be interpreted such that the a_i's are marginal budget shares and the b_i's are "necessary" or "committed" shares. Clearly the function is not defined for $x_i - b_i < 0$ and we may interpret this as implying that a certain basic level of consumption must be attained before any utility is derived. However, Pollock and Wales (1978) have cautioned against the definition of necessary shares since b_i's may be negative, and this would imply that the necessary level of expenditure is negative.

From the utility function a set of demand functions for each item of expenditure may be obtained. These take the form

$$x_{ih} = b_i + \frac{a_i}{P_i}(\mu_h - \sum_{k=1}^{n} P_k b_k) \text{ for } i = 1, \ldots, n \tag{41}$$

where x_i and P_i are the quantity and price, respectively, of good i and μ_h is total expenditure (income) of household h. The function may be rewritten as

$$P_i x_{ih} = P_i b_i + a_i(\mu_h - \sum_{k=1}^{n} P_k h_k). \tag{42}$$

The alternative form, in equation (42), is the one used for estimation. The categories defined as x_i's above are aggregate and thus it is unlikely, and indeed even unnecessary given the data are from a single cross-section, that price and quantity will be separable. However, this limitation also implies that price elasticities cannot be estimated and, hence, it will not be possible to simulate the impact of changing relative prices. If data from two years were available price indices could be used as proxies for actual prices.

The LES may be computed intensively using OLS regressions. Without going into the mathematics of expanding the equations, let us assume that a_i's are given. Substituting these into the LES produces a set of equations in n unknowns, the b_i's. OLS may be used to estimate these. The values obtained are then substituted into the equations to produce sets of equations each with one unknown, the a_i's. Again OLS may be used for estimation. Once revised a_i's have been computed, they may be substituted into the equations and a revised set of b_i's estimated. The cycle is repeated until a stable solution is obtained, a state which is usually achieved fairly rapidly. At the same time the conditions imposed on the a_i's will have been satisfied.

As noted earlier the LES may be computed from a single set of cross-section data. However, some of the data items required for the estimation are not currently being gathered. The most important of these are the income from the previous season's crop and household expenditure in the current year (season). Both will be required in order to estimate a complete system.

A final point of refinement relates to the fact that certain socio-economic variables will differ between households in the

sample, and these may be expected to influence the values of the b_i's. For example, it is likely that the size and age structure of the household will influence consumption. We may, therefore, wish to make the b_i's a function of household composition that is:

$$b_{ih} = K_i + \sum_{g=1}^{n} c_{ig} z_{gh} \tag{43}$$

where K_i is the basic level of consumption for any household,

c_{ig} is the effect of the g'th characteristic on the subsistence quantity of good i, and

z_{gh} is the independent variable.

Estimating the values of the k_i's and c_{ig}'s rather than b_i's produces a system with (m+1)n rather than n variables, and care must be taken to ensure sufficient degrees of freedom, in the estimated equations, for meaningful coefficients. One approach would be through judicious selection of the b_i's likely to be influenced by socio-economic variables.

Conceptually, the marginal budget shares (a_i's) may also be expressed as functions of household characteristics, but the estimation procedure is far more complex and will not be attempted.

Quadratic Expenditure Systems (QES)

Perhaps one of the most severe criticism of the LES is that the marginal budget shares (a_i's) are constant for all levels of expenditure. Intuitively it would be expected that marginal budget shares will differ at different levels of expenditures. This refinement is accommodated by the QES. Ignoring the form of the utility function, the demand equations

of the QES are given by

$$P_i x_{ih} = P_i b_i + a_i(\mu_h - \sum_{k=1}^{n} P_k b_k) + (c_i - a_i)\lambda \prod_{k=1}^{n} P_k^{-c_k}(\mu_h - \sum_{k=1}^{n} P_k b_k)^2$$

$$i = 1, \ldots, n \qquad (44)$$

$$\Sigma a_k = 1, \qquad \Sigma c_k = 1$$

where all symbols are the same as in the LES except that two new sets of variables, c_i's and λ, have been introduced. If $c_i = a_i$ for all i's, the system reduces to the LES.

The marginal budget shares are given by

$$\frac{\delta}{\delta\mu} P_i h^i(P,\mu) = a_i + 2(c_i - a_i)\lambda \Pi P_k^{-c_k}(\mu - \Sigma P_k b_k) \qquad (45)$$

and hence they depend on the value of μ.

The QES cannot be estimated from a single cross-section of data by the same neat approach used for the LES. However, if data are available for two years an alternative approach is possible. Rewriting the demand equation as

$$P_i x_{ih} = \theta_{1i} + \theta_{2i}\mu_h + \theta_{3i}\mu_h \qquad (46)$$

the θ's can be estimated for n-1 items in both years. From these results the values of the a_i's and $\Sigma P_k b_k$ may be obtained and hence the individual b_i's, c_i's and λ.

Similarly to the LES the b_i's may be expressed as functions of socio-economic variables and a more refined estimate obtained.

As commented previously the QES provides a more adequate representation of expenditure patterns but requires a greater volume of data, that is data for a second year. For the current study it appears unlikely that the quantity of data required will be collected.

Linear Logarithmic Expenditure System (LLES)

As a final point we note the existence of the LLES (Lan et al. 1978). The system provides a possible improvement in specification over the other systems but, as well as requiring an extensive data base, involves complex series of data manipulations and estimation procedures. Although socio-economic variables can also be included, it appears doubtful that the approach will prove of value in the near future.

Estimation of the model and use in simulation

Having developed a potential outline for the positive model we must endevour to estimate the relevant variables. Clearly some thought will need to be given to which equations can be estimated independently and which require simultaneous estimation. As a first step the system should be estimated in the simplist possible way consistent with obtaining reasonable estimates of coefficients. This will also provide the opportunity to check the credibility of the estimated components. Once a "crude" working model has been evolved, there should be time to examine more closely the estimational problems.

From the outline given it is clear that gross household income is to provide the link between seasons. Whether such a situation exists in practice must be determined on a subjective basis and at an early stage. Any modifications necessary can then be made without too many problems.

Finally the use of the crude model in some preliminary simulations at the earliest possible time would appear to be wise. Such an approach would go a long way towards determining the potential benefits from additional research work.

MODELLING VILLAGE DEVELOPMENT

Emphasis to this stage has been on developing acceptable normative and positive models of the farm household. As indicated earlier this represents the only first part of the analysis of mechanization. In order to view the problem in a broader perspective. and thereby answer some of the very complex questions on employment and income distribution, we need to construct a more encompassing model. It is intended that such models should be constructed at the village level. The reasons are first, that the data which we are collecting are reasonably consistent with village level analysis. Second, to extrapolate beyond the village level would tend to over-extend the credibility of the data set. Third, unless gross simplifications are made in the specification of the farm models, aggregation beyond a single village would result in too large a model for available resources to handle, particularly computer and human inputs.

At IRRI, a considerable effort has been made to monitor recent changes in village structure (see for example Hayami et al. 1978, Kikuchi et al., 1978 and Kikuchi et al., 1979). All of these studies emphasize what has already occurred. In doing so, they provide important information on the age structure of the village population and the dynamics of land ownership and use. For the study of the impact of mechanization, it is proposed to attempt to expand the framework so as to

provide projections of the future.

As a starting point to the analysis, consider the village level framework (Figure 4) proposed by Gemmill and Fischer (1973). In order to capture the different types of labour and the intercacies of the impact of population growth over time we have modified the population component. The revised model is shown in Figure 5. Clearly there are still many relationships, such as the effect of income from marketed surplus on choice of technology, crop mix and intensity over time, which have been ignored. However, the diagram hopefully provides a starting point for defining some of the more important points to be included.

The problem is to rationalize each of the village interactions over time, starting from 1) the previously defined farm models, 2) exogenously set rates of population growth, and 3) the current village structure.

The Village Structure

At an early stage we will need to describe the structure of our sample villages in detail. The census data should be extremely valuable in this respect, and also for generating "representative" villages in each country, should they be needed. As a starting point, it may be useful to characterize typical irrigated and rainfed village. Within the general structure, villages in each country will have specific characteristics. To illustrate briefly from my visits to the countries involved, it appears that the Philippines will be the only site with significant amounts of non-farm labour. By contrast, Thailand and South Sulawesi will not even have permanent hired labour. Also, the use of land preparation equipment will vary greatly between sites. In spite of these physical differences we should strive, if possible, to develop a general framework.

Age structure. The age structure, and family composition are important variables in our model. If we are to investigate the impact of mechanization on future employment, we need to have some indication of the likely future supply of labour and also to run the model with and without mechanization. Only then can we avoid attributing the impact of underlying changes in labour availability to mechanization. Further-

zation, a very important factor.

Given that the forward projection is likely to be for no more than ten years, so long as information is available on the current age structure of the population, the main problem in projecting the future availability of labour will be in terms of the mortality rate. To obtain information on the future population structure will require additional data on birth rates. But since these individuals will all be less than 10 years old it should be possible to make an acceptable estimate.

Non farm labour. Non-farm labour appears to be one of the areas for potential expansion. In order to gauge the possibilities we need to obtain sufficient information on the current incomes, etc. of non-farm labour. From this information we might be able to judge the extent to which the sector can expand and what resources might be required. Given current farm size and a possible shift towards mechanization, the non-farm sector appears increasingly important.

Migration. An alternative means to relieve the pressures caused by population growth, and possibly mechanization, would be by increased migration to urban centres. Initially an attempt should be made to determine current rates of migration, and whether these are permanent or temporary. It should be possible to derive sufficient data from the demographic questionnaire in the survey. Although an explanation of the migration rate is not possible from the data set, an exogenously set variable would do much to assist the analysis.

The Village Level Farm Plan

In the earlier discussion both normative and positive models for individual farms were proposed. It would be conceptually tempting to link the normative plans together, using data from the household census to provide weights, and hence to generate a general equilibrium solution. I intend to resist the temptation, for a variety of reasons. First, the resultant model would be far too large. Second, it is unlikely that villages ever are, or will be in equilibrium. Hence the additional computations are of little practical value to policy makers. Third, it would be an almost impossible task to obtain the meaningful supply functions for inputs, and demand functions for output, required for specification of the model. Finally, the formulation implies that inputs will always go to the farmer able to pay the highest price. Although that may be the case in the long-run, it is often not in the short-run.

The alternative approach is to use a "disequilibrium" model, with solutions obtained for a set of farms having different levels of fixed resources available. These solutions would indicate the quantity of variable resources required by each farm. A simulation approach could then be used to allocate the variable resources available and generate the actual levels of production. Depending on whether supply is greater than, equal to or less than demand, prices can be adjusted in the following year. Much of the information to establish such a system would have to be gleaned from the questionnaire results.

The Final Model

At this very early conceptual stage much of the final structure of the village level model is hazy. There are, perhaps, so many

questions to be answered at the individual farm level, and only when sufficient progress has been made on these will we be able to piece together the parts of a village level model. Even so it is already apparent that such a model will prove an important vehicle for examination of the likely future impact of mechanization of incomes, income distribution, employment and production.

ADDITIONAL DATA REQUIREMENTS

As has become apparent from the outlines of the various model, there are a number of sets of data which have not yet been collected but will be required for a thorough analysis. In this section we bring these together and attempt to assess the feasibility of filling the gaps. Where possible we will also draw some conclusions regarding the implications of not being able to collect these data.

Farmers' Expectations and Utility Over Time

It is quite clear that there is a major gap in the subjective data which we are collecting. As has already been indicated in the text we would follow the pattern used by other researchers of simplifying the objective function, or assuming a form for the utility function. As an initial step, this strategy would permit some analysis, but it seems highly likely that the resultant solutions would be unacceptable biased. Although the manner in which farmers form expectations is not clearly understood, the quantification of multi-attribute utility functions is well documented (see for example Herath, 1978). Extension of the utility function to a multiyear framework should also be feasible.

The amount of work required to adequately specify the functions at each site, for only a small sample of farmers, suggests that it could most readily be achieved by a series of studies. This would also ensure that the quality of enumerating required was actually achieved.

Income and Expenditure Data

During the current survey no information is being collected on consumption expenditure. If the pattern of household expenditure is to be fully analyzed, these additional data will be required. In addition, income from sale of crops produced the previous season will be required in order to construct the vital link between seasons. These data could possibly be collected during the next, and succeeding rounds but there might be doubt as to their accuracy.

Number of Years for Which Data are Collected

A simple cross-section of data should provide most of the data for a normative model, and a reasonable amount for a positive model. However, if the model is to be used to examine questions such as the impact of fuel price increases on the level of mechanization, additional data will be required. At the minimum this should be a cross-sectional survey for a second year. Alternatively it might be decided to remain with the single cross-section and answer questions relating to price changes as well as possible within the restricted framework. Further assumptions would be inevitable.

The Village Structure

Currently, our information relating to the villages and their structure is relatively sparse. Some information was collected during the household census, and more is being collected in the surveys. In addition we have qualitative information on the villages derived from our experiences. If a meaningful village model is to be constructed, there will be need to carry out a supplementary survey to gather details of the age structure and other relevant parameters, as well as to survey those members of the village previously excluded. Hence, a critical area of modelling requiring concerted thought is the structure of the village model.

CONCLUSIONS

The purpose of this paper has been to provide an initial outline of some of the approaches available for analysis of the future impact of mechanization. Analytical methods are presented for both the individual farm and village level. From this initial work it is clear that much remains to be done to adequately specify the farm level models, and even more is required for the village level model. Since all of these are inextricably link work must proceed almost simultaneously in each area.

As a minimum (personally) acceptable level of modelling the following items should be achieved:

1) development of a working normative model, preferably the recursive multiperiod integer model outlined;
2) development of a positive model; and

3) development of a village level model.

Hence, the proposed Monte Carlo model will receive the lowest level of priority. Perhaps it could be tackled as a separate research project.

One of the principle objectives in developing these models should be to produce a framework which will prove adequate for making comparisons of the impact of mechanization in the various participating countries. As appears to be almost inevitable in this type of study, an adequate specification of the models will probably take a long time. Use of the models for test simulations may be more restricted than would be desirable. However, we hope that the models will be sufficiently flexible and tested that they can form the basis for many subsequent applications.

Finally, it should be useful to plot out a plan of action for the next few months. The over-riding intention at this stage is to restrict work to the Philippines data set, since expansion to all four sets would likely cause problems. Once the techniques have been reasonably tested they can be extended to the sites in the other participating countries. Furthermore, it seems unlikely that the data from the other countries will be ready prior to that from the Philippines. Hence the plan becomes:

1) Fully identify the activities and constraints to be used in the multiperiod farm model.
2) Specify a recursive sector for the multiperiod model.
3) Start refining the specification of the positive model.
4) Using data from the first season develop some initial estimates of the econometric system. Check for consistency.
5) Further refine the positive model.

6) With a full set of data available for the Philippines specify and test the recursive, multiperiod model.

7) Recompute the positive models.

8) Use the models for some initial farm level simulations.

9) As time permits develop the village level model.

10) Extend the modelling to the other participating countries.

REFERENCES

ANDERSON, J.R. 1974. Simulation: Methodology and Application in Agricultural Economics. Review of Marketing and Agricultural Economics, 42(1):3-55.

ANDERSON, J.R. 1976. Risk Efficient Monte Carlo Programming - A Manual. Agric. Econ. & Bus. Mgmt. Misc. Pub. 6. Armidale, Australia. UNE

ANDERSON, J.A., J.L. Dillon and J.B. Hardaker. 1977. Agricultural Decision Analysis. Iowa State University Press, Ames.

ARROW, K.J., H.B. Chenery, B.S. Minhas and R.M. Solow. 1961. Capital-Labour Substitution and Economic Efficiency. Review of Economics and Statistics. 43(3):225-50.

CARLSSON, M., B. Hormark and I. Lindgren. 1969. A Monte Carlo Method for the Study of Farm Planning Problems. Review of Marketing and Agricultural Economics. 37(2):80-103.

CHIEN, Y.I. and G.L. Bradford. 1976. American Journal of Agricultural Economics. A Sequential Model of the Farm Firm Growth Process. 58(3):456-65.

DAY, R.H. 1963. Recursive Programming and Production Response. Amsterdam. North Holland Publishing Co.

DILLON, J.L. 1971. Interpreting Systems Simulation Output for Managerial Decision-Making in J.B. Dent and J.R. Anderson eds. Systems Analysis in Agricultural Management. Sydney: Wiley. p.85-120.

DONALDSON, G.F. and J.P.G. Webster. 1968. An Operating Procedure for Simulation Farm Planning - Monte Carlo Method. Department of Agricultural Economics, Wye College, Kent, U.K.

DULOY, J.H. and R. Norton. 1975. Prices and Incomes in Linear Programming Models. American Journal of Agr. Eco. 57(4):591-600.

GEMMILL, G. and C.K. Eicher. 1973. A Framework for Research on the Economics of Farm Mechanization in Developing Countries. Dept. of Ag. Econ. Michigan State University. AID/Washington Contract (AID/csd 3625).

HANOCH, G. 1971. CRESH Production Functions. Econometrica. 39(5): 695-712.

HAYAMI, Y. in association with M. Kikuchi, P.F. Moya, L.M. Bambo and E.B. Marciano. 1978. Anatomy of a peasant economy; a rice village in the Philippines. IRRI. Los Baños, Philippines.

HERATH, H.M.G. 1978. Resource Allocation by Rice Farmers in Sri Lanka: A Decision Theoretic Approach. Unpublished Ph.D. Thesis. University of New England. Armidale, Australia.

HOWE, Howard. 1977. Cross-Section Application of Linear Expenditure Systems: Responses to Socio-Demographic Effects. American Journal of Agricultural Economics. 59(1):141-48.

HUBER, G.P. 1974. Multiattribute Utility Models: A Review of Field and Field-like Studies. Management Science. 20(10):1393-402.

JORGENSEN, D.W., L.R. Christensen and L.J. Lau. 1973. Transcendental Logarithmic Production Functions. Review of Economics and Statistics. 55(1):28-45.

JUST, R.E. and R.D. Pope. 1979. Production Function Estimation and Related Risk Considerations. American Journal of Agricultural Economics, 61(2):276-84.

KIKUCHI, M., N. Fortuna, N.M. Bamno and Y. Hayami. 1978. Polarization of a Laguna Village. IRRI. Ag. Economics Paper 78-01. Unpublished.

KIKUCHI, M., A. Hafid, C. Salch, S. Hartoyo and Y. Hayami. 1979. Class Differentiation, Labor Employment and Income Distribution in a West Java Village. Supported jointly by the Agro-Economic Survey Dynamic Project in Indonesia, the IRRI in the Phils., and the IDC of Japan. Unpublished.

KINGMA, O.T. and K.W. Kerridge. 1977. Towards an Analytical Base for Studying Farm Adjustment Problems - A Recursive Stochastic Model of the Farm Firm. Quarterly Review of Agricultural Economics. 30(2):91-116.

LAU, L.J., W.L. Lin and P.A. Yotopoulos. 1978. The Linear Logarithmic Expenditure System: An Application to Consumption Leisure Choice. Econometrica. 46(4):843-67.

LEE, F.Y. and K.E. Phillips. 1971. Differences in Consumption Patterns of Farm and Non-Farm Households in the United States. American Journal of Agricultural Economics. 53(1):573-83.

LIN, W., G.W. Dean and C.V. Moore. 1974. An Empirical of Utility vs. Profit Maximization in Agricultural Production. American Journal of Agricultural Economics. 56(3):497-508.

MODIGLIANI, F. 1952. "The Measurement of Expectations" paper presented at the American meeting of the Econometric Society in Boston, December 1951. (Abstract in Econometrica 20:481-2).

POLLAK, R.A. and T.J. Wales. 1978. Estimation of Complete Demand Systems from Household Budget Data: The Linear and Quadratic Expenditure Systems. American Economic Review. 68(3):348-59.

RAE, Allan N. 1971. Stochastic Programming, Utility and Sequential Decision Problems in Farm Management. American Journal of Agricultural Economics. 53:448-60.

RAMACHANDRAN, V. 1979. The Economics of Farm Tractorization in India. Unpublished Ph.D. Thesis. North Carolina State University at Raleigh.

UPTON, M. and H. Casey. 1974. Risk and Some Pitfalls in the Use of Averages in Farm Planning. Journal of Agricultural Economics. 25(2):147-52.

WICKS, J.A. and J.W.B. Guise. 1978. An Alternative Solution to Linear Programming, Problems with Stochastic Input-Output Coefficients. Australian Journal of Agricultural Economics. 22(1):22-40.

WICKS, J.A., R.A.E. Mueller and I.R. Crellin. 1978. APMAA7 - Project Development, Model Structure, and Applications of the Aggregative Programming Model of Australian Agriculture. APMAA Report No. 11. Armidale, N.S.W., Australia.

CONSEQUENCES OF SMALL RICE FARM MECHANIZATION PROJECT

Operations Manual

1. Moran, P. and D. Unson. Farm Survey and Recordkeeping Procedures for Consequences of Small Rice Farm Mechanization Project. Revised May 1980.

2. Wicks, J.A. Procedures for Constructing Two-Way Tables for Consequences of Small Rice Farm Mechanization Project. January 1980.

3. Wicks, J.A. Specification of Variables to be Included in Data Files for Consequences of Small Rice Farm Mechanization Project. October 1980.

Working Papers

1. Juarez, F. and B. Duff. The Economic and Institutional Impact of Mechanical Threshing in Iloilo and Laguna. October 1979.

2. Pathnopas, R. The Economics of Rice Threshing Machines in Thailand: A Case Study of Chachoengsao and Supanburi Provinces. October 1979.

3. Gardezi, J.,A. Rauf, M. Munir, K. Altaf, Q. Mohd-ud-Din, and B. Lockwood. A Study of Mechanical and Traditional Wheat Threshing in Multan District, Punjab, Pakistan: Some Preliminary Results. October 1979.

4. Habito, C. and B. Duff. A Simulation Model to Evaluate Mechanization of Rice Postharvest Operations in the Philippines. October 1979.

5. Chapman, J. The Potential of Mechanization for Crop Intensification in a Rainfed Area - Iloilo, Philippines. October 1979.

6. Thapa, G. The Economics of Tractor Ownership and Use in the Nepal, Terai. October 1979.

7. Jongsuwat, N. Productivity Growth and Farm Machinery Adoption in Thai Agriculture. April 1980.

8. Bernsten, R.H. and R. Sinaga. A Methodology for Identifying Lowland Rice Farms that would Benefit from the Mechanization of Land Preparation. October 1979.

9. Bernsten, R.H. and A. Rochim. Labor Shortage as a Constraint to Increasing Cropping Intensity. Revised March 1980.

10. Ayob, A.M. The Economics and Adoption of the Combine Harvester in the Muda Region of Malaysia. October 1979.

11. Lubis, R. Impact of Cropping Pattern Technology on Income, Employment and Production: A Case Study on Expanded Crop Production in Lampung. October 1979.

12. Wicks, J. A. Modelling the Consequences of Future Mechanization: An Outline of Possible Procedures. October 1979.

13. Khoju, M. R. and J. A. Wicks. Economics of Pump-Irrigation in Eastern Nepal. August 1980.