PN-AAN-730

isn= 31990

# RECURRENT COSTS

# AND

# BASIC NEEDS STRATEGIES

David Wheeler

John Harris

African Studies Center
Department of Economics
Boston University

June 10, 1980

This report has been prepared under contract AID/ORT-G-1733 for the United States Agency for International Development. The views expressed are those of the authors and do not necessarily reflect those of any department within USAID.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

INTRODUCTION 1

I. FINANCING BASIC NEEDS -- SOME STRATEGIC ISSUES 5

A. Public Commodity Lotteries and Evolutionary Economics 9

B. Fixed Yardsticks in Basic Needs Planning 19

C. Do the Poor Know What They are Doing? 25

II. FULL INCOME ACCOUNTING AND REVEALED PREFERENCE ANALYSIS 32

A. An Augmented View of Family Income 32

B. Full Income Accounts: A Detailed Look 46

C. Planning Strategy: Project Analysis and Net Subsidy Accounting 60

III. FULL INCOME ANALYSIS AND RESOURCE MOBILIZATION: STRATEGIC ISSUES 100

A. The Urban Problem: Net Subsidy Flows and the Capitalization Process 101

B. The Urban-Rural Split 105

C. The Rural Problem: Spatial Discontinuity and Opportunity Cost in BHN Provision 107

D. Integrated Rural Development: Some Illustrative Experiences 113

# INTRODUCTION

The "basic human needs" (BHN) approach to development has been concerned with the identification and delivery of goods and services which are particularly important to the health and welfare of the poorest people in poor countries. Besides the familiar list of necessities -- food, water, shelter, health care, education -- basic needs strategists frequently stress access to sustained income through employment and participation in decision-making as crucial to successful implementation. Unfortunately, attempts to mobilize public resources at a satisfactory supporting level frequently suffer from an inability to meet the heavy recurrent costs associated with many BHN projects. Equipment deteriorates for want of maintenance; teachers go unpaid; simple rural mechanization programs come up against unexpected foreign exchange constraints. In poor societies where income and the ownership of assets are very unequally distributed and national output fluctuates sharply in response to changes in the weather or international commodity markets, the poorest people are frequently the first to be cut off. Ambitious programs can easily decline to levels of funding severe enough to assure widespread discouragement, cynicism, and corruption.

After an intensive examination of several basic needs sectors and the experience of needs-related programs in a large number of countries, we conclude that such problems are by no means inevitable. However, they can be expected to arise very dependably when basic needs planners fail to make an appropriate analysis of benefit and cost flows. Frequently, projects are structured in ways which allow the capitalized

While the approach to analysis which has been developed in this report is inevitably far from complete, it is hoped that it will represent some contribution to the understanding of BHN project analysi from the perspective of recurrent cost support. The major theme is the need to take account of the revealed preferences of the poor themselves in project design and execution. In the absence of knowledge about these preferences, it is our feeling that wasteful allocations will be common and BHN projects will be unable to exploit the available willingness to pay which is undoubtedly the key to sustainable recurren support in the long run. Because of this emphasis in our own thinking, we have devoted a major portion of the research effort summarized here to the development of appropriate measurement methodologies. Hopefully the final product will be of assistance to interested planners and policy makers.

and sanitation services are provided in central places. The problems of the ujamaa village program in Tanzania are used to illustrate our analysis of this unavoidable conflict.

Once the fundamental problems associated with recurrent support in urba and rural areas have been identified and discussed, we devote Chapter IV to a consideration of organizational forms appropriate for the mobilization of recurrent support resources by the poor themselves. Since the spatial problem is fundamental, territory seems to provide a better basis than function for organization of resource allocation across BHN categories. Certain evident weaknesses of large decision hierarchies in information-poor societies join with difficulties in central tax collection to suggest relatively small scale, local resource responsibility, and non-monetization of resource flows whenever possible.

In this concluding chapter, the key principle is the notion of resource translation. In urban areas transactions are extensively monetized and (as previously noted) ground rents can be expected to absorb much of the benefits to the poor of BHN projects in the absence of explicit government intervention. The urban discussion focuses on appropriate mechanisms for this sort of intervention.

In rural areas, the main resource of the poor is time rather than money. Our examination of project experience has suggested one expedient which can aid in the translation of the available time of the poor into a resource which is significant for the support of recurrent costs: Wherever possible, it seems appropriate to complement rural BHN projects with an ancillary capital resource which can be combined with local labor to produce revenues sufficient to cover a large proportion of recurrent labor and materials costs. Several variations on this theme are discussed.

redefinition of income and the use of particular econometric techniques as aids to the identification of perceived benefits which may be of assistance in the allocation of subsidies across BHN activities. As in all planning problems, of course, it is only net benefits which count, and the concluding sections of Chapter II are concerned with an appropriate framework for cost-benefit accounting in the context of BHN planning. We propose a form of "net subsidy analysis" as one means of monitoring the flow of opportunity values into different BHN sectors through time.

3). Planning Strategy

In Chapter III, we move beyond methodology to a broader treatment of the strategy for resource mobilization. This chapter is devoted to a discussion of the fundamental issues in BHN planning in urban and rural areas. In our view, the most important of these issues may well be the spatial accessibility of BHN delivery services. In urban areas, we identify a source of financial inadequacy in the tendency for the value of access to services such as housing and water to be capitalized in payments from the poor to private land-owners. It is apparent that recurrent support for BHN projects and the willingness of the poor to pay for needed goods and services can be brought much closer together in the planning process if the problem of capitalization in land values is attacked more directly.

In rural areas, we also identify the problem of access as crucial to an understanding of cost problems in the provision of basic needs. A fundamental conflict can be discerned between dispersed landholding patterns which are characteristic of low-technology agriculture and the economies of scale which can be realized if health, education, water,

in policy areas of concern to BHN planners. The first technique, based on the notion of "utility equivalence analysis", owes much to the standard microeconometric approach to consumer behavior and is particularly useful in studies of housing, location, and infrastructure utilization. The second technique reflects a line of inquiry which has been at the frontier of econometric research until very recently. This is the econometric analysis of discrete choices, or situations in which consumers are forced to choose among a small number of alternatives in their attempts to maximize satisfaction. We are persuaded that this econometric technique represents an innovation whose importance to efficient BHN planning is likely to be considerable.

In domains as diverse as health care, education, transportation, and the choice of residential location, the poor are confronted with choices which are discrete rather than continuous. The incompatibility of this case with existing microeconometric techniques has been a major handicap to analysis in the past. Recently, however, important work on the analysis of discrete choices has been completed by a group of econometricians at MIT. Although their work has been inspired by a set of domestic policy problems, it is clear to the authors that it can be extremely useful for BHN planning. The available techniques are as yet relatively unknown in the professional planning community, and reports in the literature have been cursory at best. For this reason, discrete choice analysis is discussed in great detail in Chapter II. In the main body of the chapter, the methodology is sketched and its application illustrated. An 80-page Appendix to the chapter is devoted to a complete discussion of the state of the art in discrete choice modeling.

Implicitly, the subjects discussed in these first sections of Chapter II pertain to the measurement of benefits to alternative activities as these are perceived by poor families. We have proposed a

2). Methodology

Having concluded that the revealed preferences of the poor can provide a superior guide for BHN planning in most cases, we devote the second chapter to an extended examination of methodologies which are useful in linking preferences to planning strategies. This chapter is by far the longest in the report. It focuses on three topics which in our view are fundamental.

The first topic is conceptual, and may well be the most important of the three. After considering the existing empirical literature on the behavior of poor consumers, we have concluded that the conventional methods for measuring expenditure lead to biased estimates of allocation across expenditure categories. In the first part of Chapter II, we develop what we have termed a "full income analysis", which allows for a recalculation of budgetary allocations after own-production and the opportunity cost of time are taken into account. A simple example is used to show that "full income analysis" can reveal a substantially different pattern of behavior than that which is suggested by standard consumer surveys. It seems apparent that this re-definition contributes further to the view that the poor allocate their resources in ways which are sensible from the perspective of BHN.

The second part of Chapter II is concerned with the implementation of a broader view of consumer behavior in a way which is useful for quantitative planning. Two means of gauging the preferences of the poor are considered: Direct preference surveys, and econometric analyses of data drawn from observations of the actual behavior of poor families. We conclude that the econometric "revealed preference" approach to measurement has much to recommend it. The bulk of the chapter is then devoted to the elaboration of techniques which are particularly useful

information in fields such as education and health may create a rationale for this approach in some cases, we attempt to show that planning by fixed targets in the BHN context will all too frequently bog down at both the conceptual and practical levels.

After considering the limitations of the evolutionary approach, we devote a final section of the chapter to a consideration of the "neoclassical" view of economic behavior. Its central tenet is simply that the poor know their own self-interest. Although contrary anecdotal evidence certainly exists, we have found little or no careful empirical work which supports the conclusion that poor families generally behave in ways which are deleterious from the perspective of basic human needs. Evidence from the ECIEL study of consumer behavior in ten South American cities is cited in support of the claim that the budgetary allocations of poor families reflect values which are essentially in keeping with the spirit of BHN.

Our conclusion after considering alternative basic philosophies of BHN provision is that it is better to use the revealed preferences of the poor themselves whenever possible as a guide to planning. Although the theme of recurrent cost analysis does not seem to be directly affected by this theoretical distinction, we would maintain that the opposite is in fact true. It is our strong feeling that recurrent support problems frequently result from the failure of public planners to take the preferences of the poor into account in a creative way.

appropriate project design in the context of basic human needs. What we term the "neoclassical view of economic behavior by poor families assumes that they are well-informed, adaptive, and aggressively self-interested in their market behavior. Planners whose beliefs are essentially neoclassical will tend to push for maximal expansion in the availability of goods and services which the poor evidently prefer, while emphasizing the exploitation of willingness to pay as a way of minimizing the recurrent cost burden of BHN projects.

In contrast to the "neoclassical" behavioral model, the "evolutionary" model leans toward a view of consumer behavior dominated by limited information, uncertain tastes, and inertia. Naturally, those who hold such views are more likely to accept the preferences of government planners in the design of BHN programs. Evolutionary strategists lean toward the adoption of size targets for BHN consumption based on attempts to define "minimum decent physical standards." Because these "minimum decent standards" are generally beyond the financial reach of the poor themselves, and because the public fisc cannot support the extension of these standards to the entire population of poor families, evolutionary programs tend to be run like lotteries. Some criterion is employed to select particular families in the poor population who are awarded an entitlement to the technically-defined minimum standard. Any tendency of recipient families to use arbitrage in moving away from social norms to privately-preferred allocations is discouraged by some system of surveillance and penalty enforcement.

In a series of sections which comprise the first chapter of the report, this theoretical division is considered in some detail. We pay particular attention to the evolutionary approach, which is more novel for economists although it has always had many adherents in other quarters. While admitting the possibility that problems of consumer

# I. EXECUTIVE SUMMARY

This report is primarily addressed to policy makers who are concerned with the design and financing of basic human needs (BHN) projects. It stresses three central themes which have emerged from our research. First, we attempt to demonstrate that BHN program design will almost inevitably reflect one of two divergent philosophies of consumer behavior. The "neoclassical" approach regards the poor as active, knowledgeable consumers and advocates reliance on their revealed preferences as a guide to BHN planning. "Evolutionary" theory, on the other hand, bases a strong advocacy of interventionist policy on the view that poor consumers are handicapped by inertia and lack of information. Our research has suggested that reliance on revealed preferences is desirable in most cases.

Our second major purpose has therefore been to develop a consistent approach to planning which incorporates the preferences of the poor. We give major emphasis to methodological questions which we think are essential in this context. Our concern with "revealed preference" analysis has lead us to the third major preoccupation of the study: The analysis of the current flow of payments by the poor for BHN goods and services and an investigation of possible means for making more efficient use of the resources of the poor themselves in providing recurrent support for BHN projects.

## 1). Theoretical Issues

The first part of the report is devoted to an analysis of the opposing behavioral models which generally motivate debate on

IV. RECURRENT COSTS AND PROGRAM IMPLEMENTATION 120

A. Organizational Strategy: Planning Flexibility and Local Control 123

B. Delivery Strategy: Horizontal Equity and Arbitrage Problems 129

C. Recurrent Financing and Local Resource Mobilization 133

APPENDIX A: Maximum Likelihood Estimation of Modal Choice Models

APPENDIX B: Basic Needs Fulfillment and Economic Growth: A Simultaneous Model

value of benefits to be appropriated by private interest groups. At the same time, project costs are wrongly estimated because the full economic environment is not taken into account. Nothing can be done, of course, about tragic resource limitations in poor countries. The poor are going to remain poor for a long time to come. It is clear, however, that the poor can be better served with existing resources in many cases. This report is concerned with setting up an analytical framework within which it will be easier to generate the maximum on-going project effort with available resources.

Our examination of the circumstances under which BHN projects succeed or fail convinces us that more attention should be paid to two related factors -- The opportunity value of time for individual members of poor families, and the spatial accessibility of BHN delivery services. Since the relationship between access and time expenditure is clear, the key component in the analysis of BHN projects may well be the way in which the poor spend their available time. From this, we conclude that the standard definition of income has to be revised before a good analysis of BHN projects can be undertaken.

Currently, social groups are defined by observable money income or expenditure, sometimes coupled with the imputed value of own-production. For the purposes of basic needs analysis, however, our research has suggested that this way of categorizing people is inadequate. Rather than observable money income, it is the potential income of poor families which counts. Each member of a poor family has a certain number of daily active hours, and each has a time opportunity cost of some magnitude. The sum of these time opportunity costs for a poor family seems to provide for a much more relevant definition of income

from the BHN perspective.

With time opportunity cost factored into the system, the real price paid by the poor for various goods and services is thrown into sharp relief. It is then possible to identify preference patterns, and a full accounting of willingness to pay can aid in lowering the recurrent subsidy burden. In particular, more accurate information about the preferences of the poor can be used as a basis for comparison with the values and expectations implicit in current governmental approaches to basic needs delivery. Generally speaking, the greater the dissonance between the revealed preferences and expectations of the public authorities and their clients, the greater is the true recurrent cost (in an opportunity sense) of particular policies.

This revised approach to family income accounting also suggests a way of thinking about organizational forms which can sustain local BHN efforts. The underlying principle is in many ways analogous to the use of foreign exchange opportunity costs in analyzing investment projects. Organizations can be evaluated according to their ability to minimize the external monetary resources which are necessary to support a given level of basic needs delivery to the poor. Since the primary resource of the poor is time, the central problem is to organize the system of project resource flows so that internal labor is exchanged for external money-equivalents at the best possible rate.

In this report, a substantial effort has been devoted to developing and applying the revised accounting scheme in the general fashion outlined above. All of this is important for the design of prospective projects, but we are also persuaded that retrospective analysis needs further elaboration as a way of monitoring the true

subsidy flows in specific cases. We have therefore developed the outlines of a "net subsidy" analysis, which provides a way of tracking the true recurrent costs of BHN projects to society by monitoring the full opportunity flows in the system and the distribution of net benefits from these flows among socio-economic groups (again, re-defined as "potential income" groups).

## I. FINANCING BASIC NEEDS -- SOME STRATEGIC ISSUES

The basic needs approach to development incorporates a strong belief that the welfare of the poor can be enhanced more quickly through targeted intervention than through simple income-promotion. Four bases for this belief are generally cited: The relative ease of extracting funds for politically-appealing, concrete goals; tremendous stickiness on the supply side in poor countries; the existence of externalities; and "irrational" consumption behavior on the part of poor families. Certainly the accumulated experience of two decades supports the first two propositions, which bear on the supply side. Taken together, they provide a rationale for targeted governmental intervention.

The latter two propositions are focused on the demand side, and they have been much less firmly established by careful empirical work. Of the two assertions, the claim that important external benefits flow from the promotion of basic needs seems much less threatening to behavioral freedom than its counterpart. The inclusion of external effects in any social calculation of benefits and costs would, after all, be endorsed by the public under its existing structure of preferences if the important linkages could be demonstrated successfully. The claim that poor consumers sometimes act "irrationally", on the other hand, obviously points toward more ominous terrain. If BHN planners adopt this point of view, then the inevitable dissonance between publically-imposed preferences and those of the poor will be suppressed only at some cost, both in behavioral freedom for the poor and in the administrative resources needed to enforce public norms.

Since it seems clear to us that the choice of intervention

strategy will have crucial significance for project design and the necessary level of recurrent subsidy, we begin with a detailed consideration of the demand-side issues (1).

---

(1) While the existence of positive externalities in the provision of basic needs to the poor is of independent interest as a topic of inquiry, it is not directly related to the main flow of the discussion presented here. We did pursue this subject in some depth as part of the Basic Needs Project, because we were unable to find a satisfactory macroeconomic treatment in the empirical literature. The purported conflict between basic needs expenditures and national output growth wa of particular interest in this context, and a major subsidiary research effort was devoted to the specification and testing of an appropriate growth model. Our results are reported in Appendix B. They certainly suggest that basic needs expenditures and conventional investment flows should be regarded as complements rather than competitors in the proces of national development.

In the design and implementation of BHN strategies, the choices of policy planners will usually reveal a belief in one of two opposing behavioral models. Adoption of the familiar neoclassical view leads to the assumption of well-informed, adaptive, and relatively aggressive economic behavior on the part of poor families. However, much current thinking on basic needs leans more heavily on another viewpoint, which we have called the "evolutionary" approach. The evolutionary model substitutes a view of consumer behavior dominated by limited information, preference uncertainty, and inertia.

While a consideration of such abstract issues may at first glance seem inappropriate in a discussion of recurrent cost problems in policy design, it is our strong feeling that the opposite is true. The basic choice of paradigm by the planner will have automatic consequences for the kind of strategy adopted. Adoption of the neoclassical view will lead to delivery systems whose priorities are strongly influenced by the revealed preferences of the poor. Neoclassical advocates will tend to push strongly for maximal expansion in the general availability of desirable goods and services while emphasizing the exploitation of willingness to pay as a way of minimizing the recurrent cost burden of BHN projects.

In contrast to this approach, the evolutionary strategy leans toward the adoption of size targets for BHN consumption based on attempts to define "minimum decent physical standards." Because the emphasis is on the attainment of numerically-defined minimum consumption levels by the poor, evolutionary programs tend to be run like lotteries. Some criterion is employed to select particular families in the poor population, and these families are then awarded an entitlement to the

technically-defined minimum standard. Any tendency of recipient families to use arbitrage in moving away from social norms to privately-preferred allocations is discouraged by some system of surveillance and penalty enforcement.

Evidently, the adoption of a particular view of behavior influences the kinds of BHN programs which are adopted. In addition, the choice must have important consequences for the recurrent cost burden and the levels of benefit to the poor which are actually yielded by projects. A behavioral model which is wrong in particular circumstances can lead to situations in which few socially-desirable benefits accrue to the poor and the entire recurrent cost is borne by the government. This is obviously a formula for slow progress, since the ability of governments to use central fiscal mechanisms to support the recurrent components of particular projects will continue to be very limited.

In an attempt to delineate the issues in a way which is relevant for policy analysis, this section will discuss the arguments which are central to both viewpoints. Although we are willing to concede the possibility that the evolutionary view is more appropriate in some contexts (and some thoughts on needed field research are included), our own conclusion is that an approach which is more neoclassical in spirit will be more effective, provided that the economic circumstances of the poor are clearly understood by policy makers.

## A. Public Commodity Lotteries and Evolutionary Economics

Many public programs are designed to operate on quantities rather than prices, with goods made available either generally or by lot to those who can pass some kind of needs test. Even wealthy societies have not abandoned goods-based as opposed to income-based assistance programs, and in poor countries the tendency to handle public welfare activities in this way is quite pronounced. Housing projects, differential siting of water facilities, ration shops, free clinics -- the list is quite long and encompasses many of the goods which are commonly found on lists of "basic needs" commodities.

To the neoclassical economist, such programs seem misplaced because they must inevitably do little more than transfer income to selected poor individuals or families in an inefficient way (2). The poor know what they want, like other people, and they will exhibit a strong tendency to arbitrage goods for cash in order to expand their consumption of all commodities for which they have positive income elasticities of demand. Public programs which select particular poor individuals for the receipt of benefits in kind can be thought of as essentially goods lotteries, with the "winning" tickets redeemable for goods or cash (at a discount reflecting transactions costs) (3).

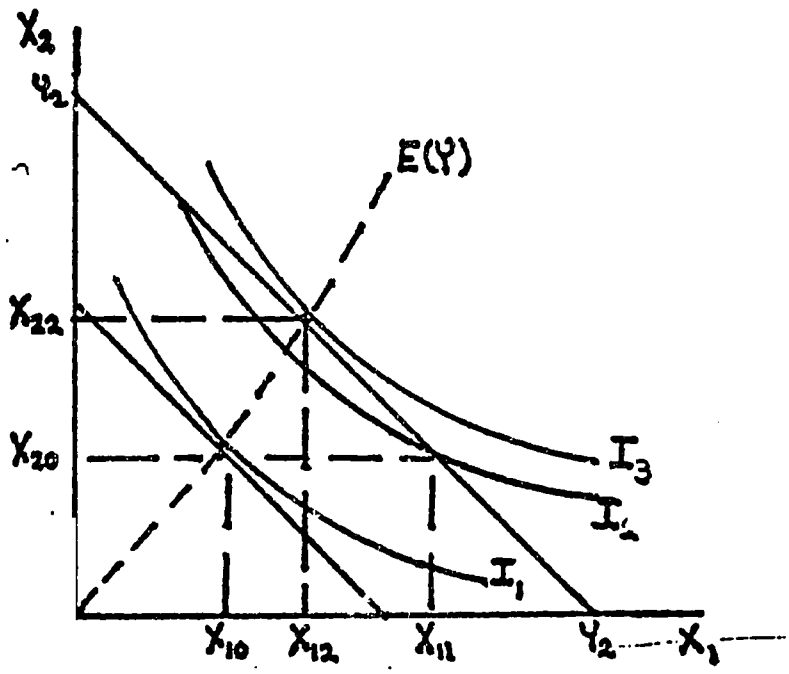
---

(2) Although this point emerges clearly in discussions of consumer reaction to transfers, it has not prevented economists from considering the comparative efficiency of needs-based subsidy programs under the assumption that arbitrage will not occur. For a discussion of the conditions under which goods deliveries and price policies have superior characteristics, see Weitzman (1974).

(3) In a very real sense, such goods lotteries are equivalent to the selective distribution of ration tickets, with the transfer of tickets

At first glance, the logic of the standard neoclassical critique of goods-based transfer programs seems unassailable. Consider the case of a poor individual who has been selected as the regular recipient of some subsidized commodity. For the sake of analytical convenience, we will suppose that the individual consumes only X{1}, the "basic" commodity, and X{2}, the "non-basic" commodity.

Figure 1




---

effected through parallel or legal markets depending upon the legal system in place. For a comprehensive discussion of the economics of alternative rationing modes, see Tobin (1970).

In Figure 1, the initial consumption point [X{10},X{20}] is determined by the individual's preference pattern I{1}, income Y, and the existing price ratio [P{2}/P{1}]. At the same price ratio, the individual's consumption expansion path is given by E{Y}. Since the non-basic good is taken to be superior and conversely for the basic good, the expansion path bends upward.

Using this graphical apparatus, we can show the essential futility of any attempt by the state to promote a discontinuous jump in the individual's basic consumption through a subsidized delivery policy. Suppose, for example, that the state delivers an additional increment of X{1} to the individual free of charge, in the hope that the new consumption combination will be maintained at the "needs-intensive" level [X{11},X{20}]. As a result of this transfer, the individual with income Y{2} is able to move to I{2}, and is unambiguously better off. The question is, will the individual choose to remain at this point?

Note that the individual has benefitted from a transfer of income equal to [X{11} - X{10} - T], where T is the transactions cost of monetizing the goods delivered. When we draw in the corresponding budget line [Y{2}Y{2}] in Figure 1, we can see why the individual is not likely to remain at [X{11},X{20}] for very long. Given the set of indifference curves derived from the individual's utility function, it is clear that the point [X{11},X{20}] does not represent the point of maximum attainable welfare for the individual. Rather, this point is given by [X{12},X{22}], the tangency of the new budget line and I{3}. For the state the outcome must be a great disappointment, since the resulting basic needs increment (X{12} - X{10}) is considerably less than the anticipated increment (X{11} - X{10}).

Inevitably, this kind of apparatus can be employed to discourage attempts at changing the behavior of the poor through direct deliveries of marketable goods and services. Before closing the case, however, we may find it worthwhile to reflect on the evidence which is available concerning people's actual behavior under such circumstances.

Few television viewers can have escaped the spectacle of "the lucky contestants" -- people of relatively modest means who, for more or less gratuitous reasons, are given wildly luxurious consumption items which they themselves could never have afforded. Generally, a subdued voice accompanies the laughter and tears with the announcement that the winners can, if they choose, have an equivalent amount in cash. Although the subsequent decisions are never announced on the air, a wealth of anecdotal evidence suggests that many contestants choose the luxury goods. Some argument basing itself on the "prestige" or "memorial" value of the commodities in question can no doubt be made, but it must be admitted that this is a somewhat unsatisfactory explanation for the phenomenon. It certainly offers little encouragement for the view of human decision-making which is portrayed in Figure 1.

Additional evidence is provided by a recent USDA study of the behavior of poor people receiving food stamps in the U.S.. While calculating that about 20 percent of the average stamp-value increment to income should be spent on food, the study finds that the real proportion is quite close to 50 (4). It is extremely difficult for the

---

(4) See USDA (1974). An ideal food stamp program could generate a ful substitution toward food by selling stamps equal in value to the consumer's current food budget. Since income variation among the poor makes such an approach impossible to administer, stamps are priced at a

standard theory to digest this result -- so difficult, in fact, that it suggests another look at the theory. Perhaps the simple calculus underlying consumer analysis is simply a poor approximation to the truth.

It is curious that neoclassical demand analysis has been so little questioned during a period in which conventional producer theory has been undergoing a furious assault. The critics of received producer theory have been led by R. Nelson and S. Winter, whose evolutionary model of firm behavior combines elements of Leibenstein and Schumpeter in a new approach to the dynamics of growth in a capitalist society (5). Fundamental to the theory are the effects of satisficing and inertia: Left to their own devices, most firms will stay with familiar production rules as long as profits remain above some satisfactory minimum. In this view managers are not constantly searching the available production set, but tend to wait until profits drop below a critical minimum before attempting some adaptation. Thus, "innovating entrepreneurs" are crucial to system dynamics, since their stochastic introduction of superior technologies drives managers to move toward more efficient techniques by eroding their profit margins.

Note that in the evolutionary view inertia and satisficing are

---

"average" level and some of the "wealthier" poor receive an effective increment to income. At the observed income elasticity of demand for food, about 20 percent of this increment should be spent on food instead of the 50 percent which has been observed. For a complete discussion of such food transfer programs, see Reutlinger and Selowsky (1975).

(5) A good summary of the Nelson-Winter hypothesis, along with a critical discussion of the received theory, is to be found in Nelson and Winter (1974). Leibenstein's critique of conventional microeconomic theory is most fully developed in Leibenstein (1976). Schumpeter's classic discussion of the role of innovating entrepreneurs can be found in Schumpeter (1934).

ordinary and innovation is extraordinary. The salvation of the system lies in the behavior of the dynamic few. Although this world view does not seem to have inspired any parallel critique of consumer theory, it is easy to argue that consumers should be even more prone to inertia an satisficing than producers. Producers, after all, have an immediately-readable index of relative performance (the rate of return to invested capital), and are subject to pressure from owners who can compare rates across firms and punish the laggards. Consumer satisfaction, on the other hand, has no such quality of comparability. Since one main force encouraging broad-ranging search is thus removed for consumers, it would be natural to suppose some presence of the evolutionary weakness in this corner as well.

Here, of course, it is tempting to marshal the support of thousands of cliches -- "We're all creatures of habit," "You can lead a horse to water but you can't make him drink," etc.. The underlying message reflects the Leibenstein-Nelson notion that people generally like to settle rather quickly into fixed, comfortable routines, and tha they are very little disposed to budge themselves from those routines unless extreme circumstances intervene. In this view, consumers are no more disposed to search exhaustively across feasible activity sets than are producers. The conventional notion of utility-maximizing choice ca therefore be given no more (and probably less) credence than the conventional notion of profit-maximizing choice.

Is all of this a prelude to some callow dismissal of the work of myriad well-informed and penetrating minds in the domain of microeconomic theory? This is far from our intention. It does seem inportant, however, to consider the evolutionary view as an alternative

It may be particularly relevant for societies where education levels are low, communications are poor, and chronic problems of health and nutrition lead to a state of relative lethargy for many poor people. From the perspective of BHN policy, the evolutionary view is significant because it provides a rationale for government imposition of public preferences on the economic behavior of the poor. It should be borne in mind, however, that the appeal of the evolutionary approach is mainly pragmatic. When the poor are either too passive or too uncertain to express their preferences, it is easier for the government to impose its own.

Thus, the phenomenon which we need to explain is the apparent readiness of many people to acquiesce in maintaining externally-imposed patterns of consumption when prior analysis of their revealed preferences would suggest arbitrage as a much more likely response. In attempting an appropriate modification of consumer theory, we must confine ourselves to a set of simplifications while keeping the main theme of the argument before us. Suppose, then, that we continue the case of the individual with current income Y who chooses to consume two goods, X{1} (basic) and X{2} (non-basic) in the quantities [X{10},X{20}]. The standard analysis would identify [X{10},X{20}] as the point of maximum utility, given current prices.

By implication, this point has been reached through an exhaustive search of the available combinations. Since it is precisely this view of behavior which seems so dubious, the process of modification should begin here. An attractive alternative reveals itself if we simply abandon the assumption that the consumer is sure that his choice is the correct one. It seems more realistic to associate consumer equilibrium

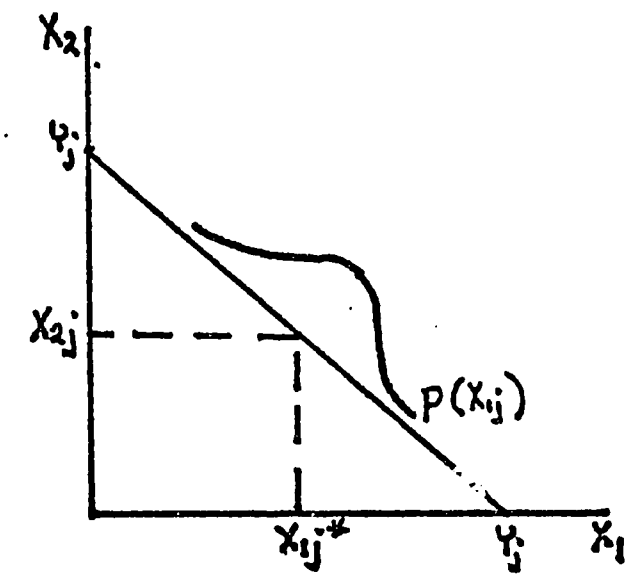
at Y with the consumption pattern which maximizes expected utility, given Y. Formally, we can specify the individual's problem as:

Choose X{1j} such that E[U(Y{j})] is maximized, given that

$$E[U] = [p(X\{1j\})]*[U'(Y\{j\})]$$

Since two goods are involved in this analysis and the individual' income is exhausted by his purchases, we can specify all functions in terms of X{1}. Here p(X{1j}) is the probability that a particular consumption level (X{ij}) maximizes utility, and U'(Y{j}) is to be thought of as some appropriately-scaled measure of attainable utility. Graphically, we are specifying a situation of the following sort:

Figure 2



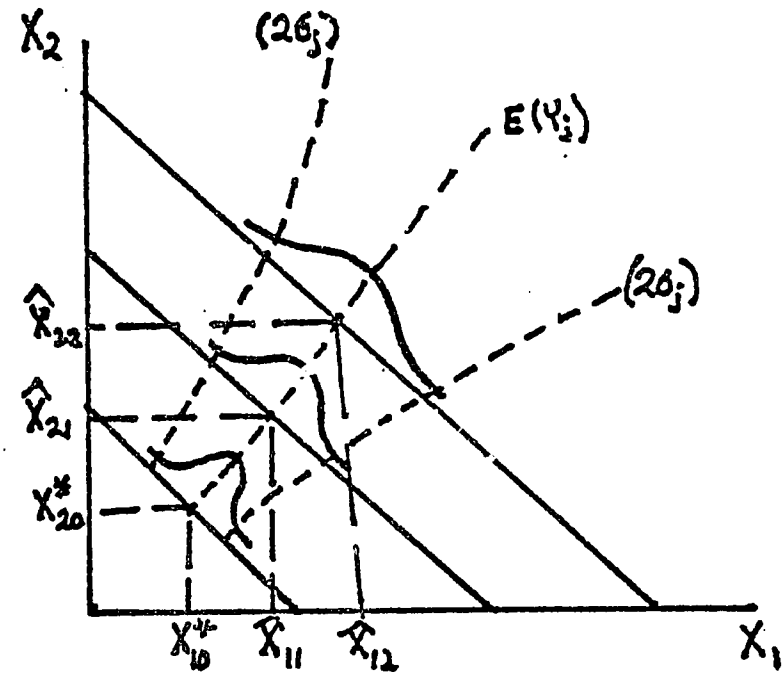
Thus far, there seems to be nothing which really differentiates this model from the standard approach. The individual freely chooses to consume X{1j}, which is the point of tangency between the budget line and the relevant indifference curve. However, the conventional treatment of indifference is done away with in this formulation. The individual in question is only relatively certain that his current consumption pattern is the best one. Although the existence of a normal density function is sufficient to guarantee transitivity ex ante, we are viewing the individual as a Bayesian who admits that the current position of the function might be modified by further experimentation.

All this seems like a theoretical nicety until we consider the sort of problem which has motivated this discussion. What happens when the individual is boosted to an entirely new domain of potential consumption experience by a significant transfer of income in money or in kind? Suppose, for example, that the individual at income level Y{j} is asked to specify the increases in X{1}-consumption which would characterize a steady growth in his income. Now, there is little question that the individual will be able to specify this expansion path, but once we admit the psychological principle which generated the probability distribution at Y{j}, we are left with an undeniable complication.

In asking the individual to project an expansion path with no prior experience of higher income, we are asking him to forecast his own behavior in a domain where no actual observations are available for consideration. In econometric forecasting, forecast variance grows with the distance from the region of greatest experience (i.e. the vector of mean values). It does not seem unreasonable to suppose that the same

thing should be true of individuals. Thus, we can think of the individual's expansion path as the locus of maximum likelihood consumption levels [$\bar{X}'\{1j\}$], with the variances of the associated probability distributions increasing with the distance from the current income level:

Figure 3




Again, the relevance of this approach may seem suspect. As the individual becomes richer, he will simply follow the expansion path X{1j} (Y{j}), and all off-path points seem to have no operational significance. This is quite true as long as the individual's income

growth is experienced in cash. A growth in kind may be quite another matter, however.

It is entirely possible to imagine levels of uncertainty so great that moderately risk-averse individuals would fail to arbitrage even with a very low transactions cost or probability of being caught and punished (6). Thus, commodity lotteries which deliver substantial increments in real income equivalence to poor families may attain the preferred consumption pattern in a large number of cases. Again, the fact that such lotteries might be run successfully does not imply that they should be run. At least two arguments must be convincingly defended before such a conclusion can be drawn. First, it must be shown that the government can run effective lotteries by choosing appropriate goods, services, and delivery mechanisms; secondly, it must be argued that the state can really determine and reinforce the self-interest of poor families better than they can themselves. In the following sections, these two claims will be discussed in detail.

## B. Fixed Yardsticks in Basic Needs Planning

As we have just noted, the evolutionary view of economic behavior leads to a planning approach which emphasizes the imposition of publically-determined values on the consumption patterns of the poor. Paul Streeten (among others) has provided the rationale for such an approach (7):

---

(6) This subject will be considered in more detail in Chapter IV.
(7) See Streeten (1979).

> The essence of the case for basic needs is that the gap between requirements and actual living levels can be filled sooner, and with fewer resources, than by alternative routes...

Among the reasons given for this superiority, three bear directly on th social desirability of current consumption patterns (8):

(1). There is some evidence that the poor are not always efficient optimizers, especially concerning nutrition and health...

(2). The manner in which additional income is earned may affect nutrition adversely...

(3). There is maldistribution within households as well as betwee households; women and children tend to be neglected in favor of adult males.

If we are to accept this judgment and presume that the values of the poor are misplaced, then the state may be able to improve on the results of an income promotion policy by setting its own standards. A.C. Pigou considered this problem in his classic Economics of Welfare (9):

> ...the minimum standard...must be conceived, not as a subjective minimum of satisfaction, but as an objective minimum of conditions... Thus, the minimum includes some defined quantity and quality of house accommodation, of medical care, of education, of food, of leisure, of the apparatus of sanitary convenience and safety where work is carried on, and so on...Furthermore, the minimum is absolute. If a citizen can afford to maintain it in all departments, th State cares nothing that he would prefer to fail in one. It will not allow him, for example, to save money for a carouse at the cost of living in a room unfit for human habitation.

Thus, Pigou adds explicit detail to Streeten's position. If the

---

(8) Streeten (1979).
(9) See Pigou (1919).

government is going to attempt to improve on the results of income promotion on the demand side, then it must define appropriate standards and enforce them. We can see a strong reflection of this approach in some of the most recent thinking about basic needs. A recent discussion of the problem by D.P. Ghai and T. Alfthan for the ILO (1977) provides a good case in point. Their work is largely concerned with identifying expenses in the "core" (food, clothing, shelter (including sanitation), health, education, water supply) and the "periphery" (transport, fuel and oil, contraception, household furniture, etc.) of the poor consumer's budget. Because their discussion is general in nature, Ghai and Alfthan make no attempt to hammer out the hard definitional problems in a particular context. In addition, they assert that basic needs targets should not be "subsistence minima," but "decent standards" as defined by some combination of national and international norms.

This general theme has also been picked up by some international donor agencies, including USAID. In a recent official policy paper, for example, we find (10):

> International discussions of basic human needs have tended to define needs in terms of the minimum requirements of a family for productive survival. These include requirements for...food of sufficient quantity and quality to satisfy basic nutritional requirements, adequate housing and clothing,... safe drinking water, sanitation, and basic health, family planning, and education facilities and services.
>
> Targets for meeting these requirements can be established in several ways...; in practice, owing to conceptual and data limitations, the mix will vary from place to place. Increased effectiveness in meeting basic needs objectives will nonetheless require increased emphasis on setting measurable goals targets in terms of levels of well-being or status, including associated improvements in data and analysis.

---

(10) See USAID (1978).

Obviously, the idea of targeting on physically-defined consumption standards is becoming very popular. Because we can see other possible approaches to BHN planning, and we think that the adoption of a major strategy has important consequences for the recurrent cost burden of the government, we think that this targeting strategy warrants close critical scrutiny. We will begin at the practical level, by considering this kind of approach as a planning problem.

The first task is to define basic needs for a country in a way which is sensible by local criteria and sufficiently disaggregative to be useful in practice. This is an extremely tall order, as a glance at a realistic case will suffice to demonstrate. Among the factors which will undoubtedly affect judgments about the level of needs, we can cite age, sex, work activity level, location, climatic variation, and culture as a minimum set. Suppose that we control only for 4 age categories, 2 sexes, 3 activity levels, 6 location types, 3 culture groups, and 2 levels of climatic variation in determining "objective" need. We are saddled with the need for enumerating 864 targets for each kind of basic consumption. Suppose that we allow for 2 indices for each of 6 specific categories (nutrition, water, shelter, clothing, health, education). We now have the necessity for 10,368 target numbers.

Unfortunately, more details must be considered. We must be concerned with the circumstances under which delivery to paticular individuals would take place. All available evidence shows that joint consumption by families has its own dynamics. Thus, not only the individual's characteristics but the characteristics of his joint consumption group must be taken into account in setting goals. Suppose

that we preserve simplicity by allowing for only 6 group types. The target number now escalates to 62,208, and we begin to perceive in the targeting approach a full employment policy for statisticians.

This is obviously a reduction of the argument to absurdity, and yet the individual measures of variation are modest enough in number. Now a further problem intrudes -- In order to plan, we have to worry about specific goods and services. Each of these, in turn, will be valued only insofar as it makes a fundamental contribution to the basic welfare of the individual. Although opinions on its exact nature will, of course, differ, we may usefully think about a "meta-production" function which traces the effect of each kind of consumption on productivity, longevity, and "life circumstances" (including the organizational and political circumstances under which the individual's needs are provided).

Thus, we arrive at the need to specify a series of functions which relate both kinds of inputs and the ways in which their combinations are organized to ultimate outcomes. If we could do this, we could arrive at some notion of marginal benefits associated with the additional consumption of particular kinds of goods and services by particular kinds of poor people. In addition, as always, we would have to worry about the measurement of costs using shadow prices.

Now other problems arise. We are attempting to set up a planning problem in a context where the production function is highly nonlinear, its parameters are unknown, relative value weights must be determined through some process, and many different kinds of individuals, goods and services, and production possibilities are involved. Worse, we cannot even define demand and supply conditions with the usual precision. In

the case of demand, a belief in "sub-optimal" consumption by the poor throws the whole basis for estimating demand functions out the window and we are left with uncertainty. On the supply side, the introduction of social desirability into the problem of technical choice leaves us in a similar state.

In brief, the targeting approach involves attempting to solve for the optimum resurce allocation in a system with:

(a). Tremendous (and unfamiliar) nonlinearity
(b). Unknown parameter values, most of which will inevitably remain speculative
(c). No apparent exogeneities, either on the demand or supply sides

Optimism about implementing such a system would seem premature, to put it mildly. None of the individual components of the problem is anything other than a distillation from the general approach, but we obviously have to forget about anything so complicated. But how, then, can it be usefully simplified? One suggestion would be the specification of much simpler targets and far fewer aggregative categories. Then, however, the Pigovian maxim would come to bear: When subsidized individual consumption patterns are subsequently observed to depart from those which are socially sanctioned, should legal sanctions be applied? Obviously, a negative answer would do away with one pillar of the targeting approach. A positive answer, on the other hand, would yield the follow-on question: What are allowable exceptions? And it turns out, of course, that a legally-acceptable definition of exception would lead immediately back to the thousands of particular categories which we had hoped to avoid.

Thus, we begin our scrutiny of targeting by judging that the implied planning schema is hopelessly complex. At best, any public

policy which builds on this organizing principle will be hopeful and exploratory. It therefore seems appropriate to us to attempt a movement back toward first principles. Since the assumption that the poor don't consume "efficiently" is certainly a contributor to complexity in planning, let us consider it directly.

### C. Do the Poor Know What They Are Doing?

Any assertion that the poor consume "irrationally" must be based on a system of knowledge and values which purports to be different than that of the poor themselves. Even strong advocates of paternalist welfare policies would presumably not argue that the poor were in fact pathological. Thus, a statement by some observer that the poor are "irrational" in some context must stem from a divergence along one of three dimensions: (1) The valuation of the expected results of current consumption; (2) Expectations concerning the results of current consumption; (3) The relative valuation of future consumption as opposed to current consumption.

These dimensions represent a mixture of the verifiable and the purely value-laden. Thus, the selection of the desirable end results of consumption for a particular poor family represents a value judgment, as does the weighting of each end result for each family member. The degree to which each unit good or service consumed contributes to each end result is verifiable, at least in principle, for most kinds of consumption (This is not to say, of course, that any such measurements have ever been successfully undertaken). Finally, the choice of a rate of time preference is clearly a value judgment.

Which divergence serves as the source for most judgments concerning the irrationality of the poor? It is probably not unfair to say that the most common source is the second one. There seems to be a common feeling that poor, largely illerate people mean as well as anyone else, but that they are not sufficiently well informed about expected outcomes to make good choices. Among examples of this kind of judgment, we might cite the notion that poor families shortchange infants at mealtime because they do not understand the long-run consequences of malnutrition in early childhood. The tendency of many rural families to keep school-age children at home to help with the chores provides another case in point. Further examples are easy to come by.

Now, it is always easy to find anecdotes whose selective presentation will reinforce a particular point of view. For our purposes, a more general approach would be desirable. Three questions need to be answered before the notion of "irrationality" on the part of the poor can really be adopted as a basis for public policy: (1) Does the general pattern of consumption by the poor suggest any marked divergences from behavior which seems "desirable" from the BHN perspective? (2) For poor countries, is there systematic evidence that representative poor families consume in ways which are relatively injurious, once their whole life circumstances are taken into account? (3) Even if such unfortunate behavior is observable, is it likely that the total results of direct intervention by public planners will be more desirable than the existing situation?

It might be supposed that the answer to the first question would be affirmative, in view of the attention which has been given to this problem. In fact, available budget studies for countries in Asia, Latin

America, and Africa invariably show that a huge proportion of expenditures by the poor are allocated to basic needs categories. The most comprehensive work to date is undoubtedly the Brookings-ECIEL (1978) study of consumer behavior in ten Andean cities. Notional estimates of basic needs expenditures following the Ghai-Alfthan decomposition are as indicated in Table 1.

Table 1

Basic Needs Expenditures as Percent of Family Budgets for the Lowest Income Quartiles (Ten Andean Cities) (a)

| City | Mean Income of Lowest Quartile | Percent (b) |
|---|---|---|
| Bogota | 1037 | 89.7 |
| Barranquila | 1100 | 88.7 |
| Cali | 1029 | 88.6 |
| Medellin | 1055 | 87.8 |
| Santiago | 942 | 85.7 |
| Quito | 845 | 90.5 |
| Quayaquil | 896 | 91.4 |
| Lima | 1161 | 80.8 |
| Caracas | 1602 | 80.4 |
| Maracaibo | 1452 | 82.0 |

(a) Relevant definitions are as follows:

"Basic" expenditure categories: Food and beverages; housing; clothing; medical; education

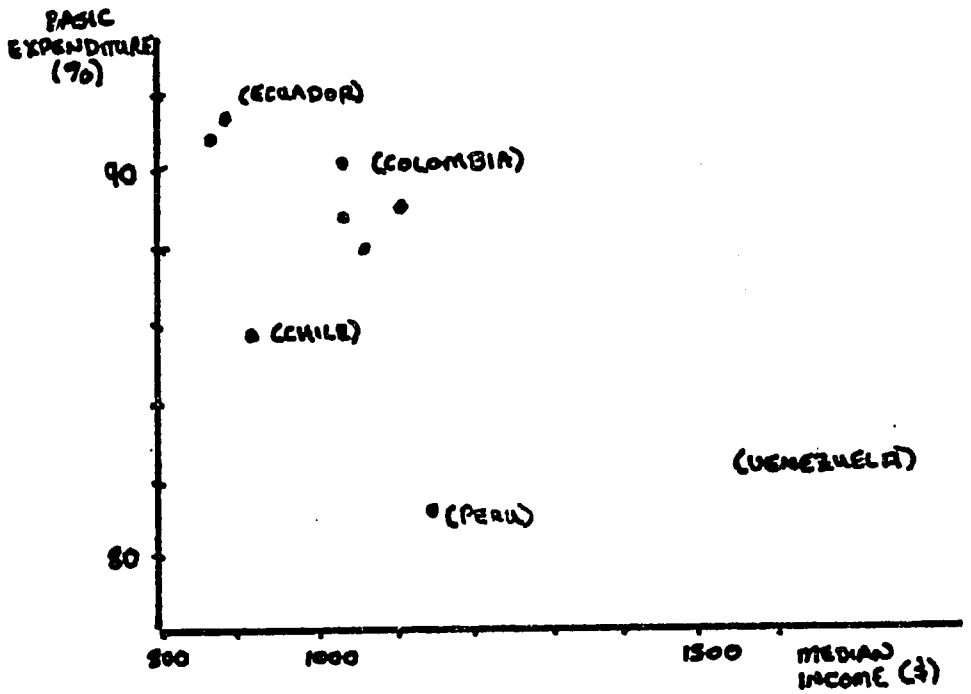
"Non-Basic" expenditure categories: Furnishings and operations; recreation and culture; vehicle operation; public transportation; communication; other consumption (tobacco, personal care, ceremonies); insurance; gifts and transfers; other nonconsumption

(b) From the perspective of the "full income" approach which is developed later in this report, all of these figures may well be underestimates of true budget proportions.

There are obviously some problems with this kind of categorization (Are ceremonies never basic needs? Are all furnishings "non-basic"?).

Nevertheless, it reveals a persistent tendency by poor families to spend most of their money on categories which are closely identified with basic needs. In addition, there is a tendency for this proportion to expand as the median income of the lowest quartile gets lower across cities, as evidenced by Figure 4.

Figure 4




The observations for the North Andean countries in this set are very tightly distributed along a line, while the two South Andean observations determine a lower, nearly parallel line. With this kind of evidence before us, it is very hard to argue that the poor (at least th

urban poor) are not spending their money wisely by BHN standards. Existing budget surveys for Africa and Asia have not been done as consistently as the Brookings-ECIEL surveys, but they exhibit the same disproportionate allocation to BHN expenditures by the poor.

If the first basis for judging poor consumers to be irrational does not hold up well, then perhaps the other two possible arguments are credible. Do systematic studies exist which show obviously injurious behavior by poor families, once their whole life circumstances have been taken into account? In fact, the latter proviso rules out almost all existing efforts, since the great majority of studies have been done by sector specialists on subjects of immediate interest to themselves. In the absence of a total consideration of life circumstances, very few kinds of behavior can be dismissed as aberrant. Is disproportionate food allocation to adult males a sign of pure discrimination, or do the men of the family simply need a fairly high proportion of a grossly inadequate family diet to maintain strength sufficient to keep their jobs (11)? Are poor rural families who keep their school-age children at home acting unreasonably, or do they observe a very high proportion of local primary graduates living in circumstances which have not changed as a result of schooling?

As we have mentioned, little or no research on total family behavior has been done. Existing studies must yield ambiguous conclusions for the most part, because they are partial. We can conclude that very little systematic and comprehensive evidence exists

---

(11) In fact, the common feeling that such disproportionate allocation occurs is not backed by much systematic evidence. For one case (Kinshasa, Zaire) in which the children of the poor seem to be disproportionately favored, see Wheeler (1978).

to support the assertion that the poor behave in injurious ways by BHN standards.

This leads us finally to the third question, which we pose again in a form modified by the apparent answers to (1) and (2). Is there an reason to believe that public planners can define and promote the self-interest of the poor bettter than they can themselves? First, it should be noted that policy planners are generally working in agencies which are sector-specific. To suppose that they are more likely to understand the whole life circumstances of the poor better than their academic counterparts seems very unlikely.

Let us suppose for a moment, however, that public planners are really endowed with such perception. Then the assumption that their choices will improve on those of the poor implies that they can analyze the associated risks and returns better than the people who are directl affected. If an accumulating body of evidence from the industrial economies is relevant, the opposite conclusion appears warranted. Studies of the ability of "experts" to outperform the guesses of masses of "uninformed" players in risk-return analysis are almost always embarrassing for the experts. By the same token, it is difficult to believe that a relatively small group of planners in some public agency could understand the full life situations of a great variety of poor people and guess the expected returns to various activities better than they could themselves

In summary, we see no reason to suppose that the poor are not trustworthy as judges of their own self-interest. The benefits of direct intervention on the demand side, if any, are likely to be so small that they would be considerably outweighed by the resource costs

of project administration. Thus, we conclude that the revealed preferences of the poor generally provide an excellent guide to the measurement of relative benefits for BHN planning. Where this does not seem to be the case, resources spent on public education rather than coercive intervention are likely to yield higher returns.

# II. FULL INCOME ACCOUNTING AND REVEALED PREFERENCE ANALYSIS

## A. An Augmented View of Family Income

In Chapter I, we have seen that very little of the available evidence supports the view that the government should attempt to dictate consumption choices to the poor. Through supply-side intervention, it can certainly offer incentives for increasing BHN consumption, both by raising real incomes and adjusting relative prices. By respecting the revealed preferences of the poor whenever possible, this approach will maximize the degree to which local willingness to pay can ease the recurrent subsidy burden. Our research has, however, led to the conclusion that the standard approach to deducing the preferences of the poor suffers from a failure to define the concepts "income" and "price" appropriately. In this section, we suggest augmenting the standard definitions by incorporating the value of time. In so doing, we are lead to a significantly revised view of consumption, savings, and investment behavior by poor families.

The Western microeconomics of Marshall and his successors developed during an era in which the industrial revolution had essentially solved the survival problem for many North Atlantic populations. The resulting theory of consumer behavior is therefore preoccupied with alternative "tastes" for goods and services which are not directly linked to survival needs. It is perhaps this same unconcern with sheer survival which has focused the theory on individual choices, although it is perfectly apparent that joint consumption by nuclear families has retained a strong role in most Western economies.

In the standard one-period analysis, the consumption of goods and services determines an individual's welfare level. Present consumption is implicitly enjoyment, and deferred consumption (i.e. proportionately more consumption in the future through savings-investment) is determined by interaction between individual "impatience" (the "rate of time preference") and available production possibilities.

Thus, the only bridge between consumption now and in the future in textbook microtheory is the simple act of "saving." Precise definitions are wanting, probably (as has been noted) because microeconomics implicitly assumes a level of well-being sufficiently advanced that the link between current consumption and current productivity (i.e. contributions to future consumption) is broken.

This is patently not the case in many poor societies. When family income is sufficiently low that physical productivity is a variable function of consumption composition, the story becomes more complicated. To the extent that workers perform better and longer with a nutritious diet and superior access to medical care, and to the extent that these input levels have long-run consequences, the general term "consumption" is revealed to have a very large component which might with considerable justice be regarded as "investment." As a first order of business, then, it makes sense to define family economic behavior a little more carefully.

A sensible way to begin is to think about the business accounting model as a possible alternative to the usual consumer model. Although it may seem a little artificial, this way of characterizing the operations of a poor family may be more accurate than its competitors. Thus, we can think of the goal of the family as maximizing "profit" for

all its members (12). In procuring "profit" for itself, the household must sell some product or service in the external market. In exchange, it can purchase goods or services. Some of these are used as inputs to current production --e.g. land rent, materials, labor services, capital services. Additional resources must be put aside to assure higher future operating levels (or simply the neutralizing of depreciation) through the accumulation of capital.

Thus, the "family" enterprise has revenue from sales (either implicit or explicit) and costs on both current and capital account. The difference is "operating profit," which the family distributes amon its members and is presumably free to enjoy as it wishes. Admittedly, this set of definitions is tricky. Goods and services may be purchased as "inputs," or as profit items; family labor can be used either internally or externally, and the problem of implicit payment to internal labor services clearly arises. The division of outflows into "profit-related," "input-related," and "capital-related" components may seem abstract and overly complicated. It is important to note, however, that the standard distinction between savings and consumption (where on kind of consumption is "leisure" -- not working) is also an abstraction and not a very helpful one in this context.

In the re-definition of expenses, profit-related outflows are ver much like the standard notion of consumption. Some current income is spent on goods which are only tenuously linked with productivity or survival. Examples might include alcohol, soft drinks, snack foods, dry

---

(12) "Profit" and enjoyment (or utility) are obviously linked notions here but for the moment it seems better to leave the basic notion in imprecise form and pass on to a full description of the accounts system

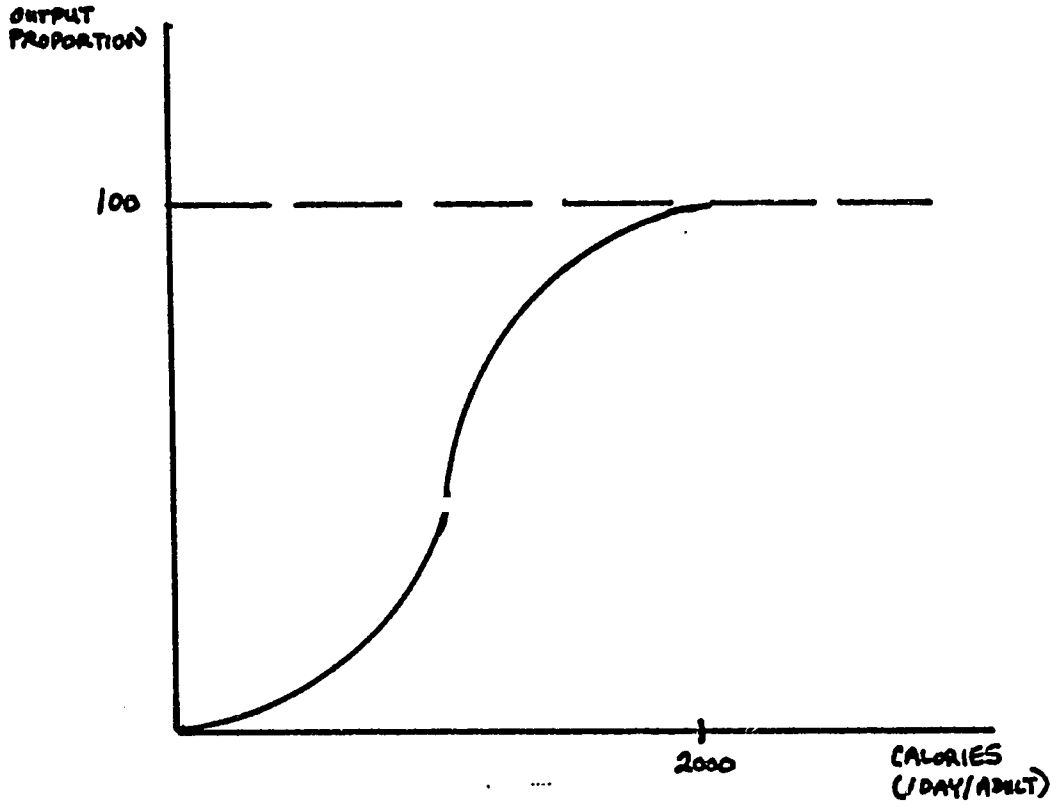
cleaning, comic books, attendance at movies or sports matches, stylish clothes, etc.. All studies of the behavior of the poor show that even those who are regarded as desperate cases by the international aid community spend some of their current income on such things. Richer people always spend proportionately more of their income in these ways, of course, and this fact gives us a basis for defining "pure profit" allocation in the family accounts. Pure profit items are those whose purchase elasticity with respect to gross family revenue is small at low gross revenue levels but rises as revenue increases. Thus, it is perfectly possible to define specific items as pure profit items in particular empirical contexts.

We are left with operating expenses and saving-investment as two categories which remain to be distinguished. One component of their definition is clear: They are all items whose gross-revenue elasticity is large at low income levels and non-increasing at higher levels. Both categories are necessary for family operations, but they differ rather drastically in their productivity impacts. Operating inputs can be roughly divided into three kinds -- direct production materials (i.e. those which are combined by family members to produce outputs for sale), indirect production materials (i.e. those which are consumed by family members as a necessary complement to activity -- food, water, etc.), and the services of current stocks of human and physical capital.

Generally, we may define productivity in the following way: Direct -- a fixed-coefficient relationship with marketed output (this is not strictly correct, undoubtedly, but it will certainly simplify the analysis at a point which is not crucial); Indirect -- a relationship between any of the indirect materials inputs and output for the family

which is exemplified by the following graph.

Figure 5




In Figure 5, we see the proportion of potential output (determine by available stocks of capital, labor, and land) which can be hypothetically reached at different levels of consumption of calories. The graph is intended principally to demonstrate that indirect material inputs are distinguished by the principle of satiation from the perspective of production, and the satiation level is well-nigh automatically reached once average family output (/income) has been reached. Factor service inputs are also characterized by diminishing

marginal productivity, of course, but they are not as obviously subject to the asymptotic limitation which characterizes many materials inputs at realistic operating levels (13).

Finally, we come to the capital account component of expenditure. One of the most familiar notions in microeconomics is that capital accumulation must occur to maintain productivity as gross operating levels rise. Part of current revenues must be set aside for this accumulation (and, of course, to compensate for depreciation). The family faces the same problem -- How should current revenues be allocated among profit, operating expenses, and investment? In the microeconomic model of business operation, the relevant goal would be defined as maximization of the discounted flow of expected profits, and there is no reason why the same general principle could not be applied to the operations of the poor family. Thus, spendable revenues should be allocated so that the marginal long-run expected consumption value of expenditure across uses is the same.

In practice, the family's decision will be determined by a few crucial values: The discount rate, dictated by the relative preference for the present over the future; the current productivity profile of various inputs; the expected productivity of investments; the expected rate of real growth in the rest of the economy; and the availability of desirable consumer goods and services. Given its perception of these

---

(13) It has been pointed out that this conclusion should be modified to account for the role of complementary factors -- most notably land. In Indonesia, the omnipresence and richness of agricultural land leads to a situation in which the marginal productivity of female and child labor at low capital levels is considerable, even at fairly high use rates. In arid lands such as Botswana, on the other hand, this seems not to be the case. The evidence comes from Boston University research projects currently underway in the two countries mentioned.

values, the family will shift its allocation around until the marginal condition is approximately met.

Since the consumption decision will always be subject to some finite discount rate, some gross revenue will be set aside for "profit consumption," even by the poorest families. The maximum efficiency increment from operating expenditures should be attainable at fairly low gross revenue levels, so that total expenditure (at constant prices) on such inputs should stabilize near the asymptotic maximal contribution level. There is no reason, on the other hand, to suppose that the rate of return to investment should decline with its level for individual families. Generally, in the trade between present and future profit-consumption, the shift should be toward a higher propensity to save as gross revenue expands. Thus, a definite pattern should emerge as gross revenue expands: The proportion devoted to operating expenditures should drop, and the proportions devoted to profit-consumption and investment should rise (different relative rates of change can be imagined).

Among consumption goods and services, it is fairly easy to differentiate between "operating" and "profit" purchases. The former should exhibit a purchase elasticity with respect to gross revenue which is high at low revenue levels and the latter should exhibit an elasticity pattern which is precisely the converse. The pattern for saving-investment expenditures should be very similar to that exhibited by "profit" operations.

Now we arrive at a definitional problem which obviously causes some ambiguity in our perception of the difference between operating and investment expenditures. Many necessary operating expenditures at very

low gross revenue levels have productivity implications which span more than one year. Extra food for hungry infants, for example, can have lifelong effects on strength, intelligence, and health. To a certain extent, the same thing is true of clean water and, perhaps, shelter. Thus, some accounting confusion is inevitable. For precise numerical exercises this might cause problems, but for analytical purposes this limited source of confusion is not really very important.

On the expenditure side, then, we have decomposed consumption into two parts and pointed out some discrepancies in the distinction between operating expenses and investment for poor families at some points. By now, it should be clear that all operating expenses and at least some investment expenses have generally been termed "basic needs" expenditures in recent years. The tripartite definition of expenditure which has been proposed here may be helpful in giving precise meaning to basic needs and their evolution in particular contexts. We will return to this problem in the last part of this section.

Having examined the outflow side of the family ledger at a rather general level, let us now turn our attention to the income side for a moment. In thinking about the allocation problem of the poor, we are immediately confronted with the need to re-define income. For an accurate analysis of behavior, we need to look at income as opportunity value for a poor family.

The poor really have two important exchange media -- money and time. In analyzing the economic behavior of poor families, then, we need to consider as income their money income; money-equivalent income from own-production; and the value-equivalent of the time which they spend in the procurement of goods and services (other than own-production) or in

the realization of savings as investment. Similarly, we need to define the "price" of any particular good or service purchased on profit, current operating, or capital account as the full opportunity cost associated with the purchase.

As an illustration of these principles, consider the case of a three-person family consisting of a man in his thirties, his wife, and one twelve-year-old son. We will suppose that the man has a factory jo which pays 30/month and the wife keeps house, markets, and works part-time as a seamstress, for which she can expect to receive approximately .10/hour. The son could find employment in the informal sector for about the same wage, if the family decided to terminate his education, and the father could "moonlight" for about .15/hour if he chose to do so. Now, suppose that the maximum sustainable working day is 12 hours. By our definition, the full opportunity income of the family in question is 48 (father) + 36 (mother) + 36 (son), or 120/month. This is the income which they could obtain by allocating their time fully to work each month. On the expenditure side, a brief accounting of time and money allotted (per day) yields:

Table 2

| | Father | | Mother | | Son | |
|---|---|---|---|---|---|---|
| | Money | Time | Money | Time | Money | Time |
| Profit | .10 | 2 | 0 | 2 | 0 | 1 |
| Operating | .90 | 2 | .30 | 7 | .20 | 1 |
| Investment | 0 | 0 | 0 | 0 | .10 | 7 |

This yields the following full income calculation for expenditure:

Table 3

| Category | Money | Time | | | Total Value |
|---|---|---|---|---|---|
| | | F | M | S | |
| Profit | .10 | 2 | 2 | 1 | .70 |
| Operating | 1.40 | 2 | 7 | 1 | 2.50 |
| Investment | .10 | 0 | 0 | 7 | .80 |
| Income | 1.60 | .60 | .90 | .90 | 4.00 |

We therefore have the following comparison of expenditure proportions:

Table 4

| | Standard Budget Survey | | Full Income Accounting | |
|---|---|---|---|---|
| | Value | Percent | Value | Percent |
| Profit | .10 | 6 | .70 | 17 |
| Operating | 1.40 | 88 | 2.50 | 63 |
| Investment | .10 | 6 | .80 | 20 |
| Total | 1.60 | | 4.00 | |

In the full income accounting, the expenditure proportion for profit rises because the opportunity value of leisure time during the twelve-hour "workable" day is considered. Similarly, the proportion allocated to investment rises substantially because attention is paid to the opportunity value of the son's education time. In re-defining income, we have generated a very large shift in estimated expenditure proportions away from operations and toward profit and investment. There can be little doubt that the resulting impression of family economic behavior is quite different and (at least to us) more revealing.

Besides the insights into family behavior generated by this re-definition of income, there is an additional gain from the associate re-definition of prices to account for the value of time spent in acquisition. For BHN analysis, this re-definition is particularly important because it throws into strong relief one crucial point -- whe the opportunity cost of time is accounted for, even "free" goods are no longer free. In the simple example which we have considered, education is an obvious case in point. Although classroom instruction may have n direct monetary charge attached to it, it is far from free for the family in question. If allowed to chose voluntarily, then, the family is very likely to treat the education of the son purely as an investment. The present opportunity costs will be borne only if the present value of the future earnings stream appears large enough.

Although revised notions of income and price have been introduced as aids to understanding economic behavior, we have not yet arrived at definition of the economic objective of poor families which is appropriate. Much of the discussion of basic needs has focused on the idea that contributions of goods and services to the physical quality o life (generally associated with health and longevity) should be valued in considering the well-being of the poor. Of course, it seems perfectly sensible for the poor, as normal human beings, to value these things as well. Therefore, we may suppose that for any poor family, th objective of economic activity is the maximization of the present valu of the total profit-consumption stream for all living (or about-to-be living) family members.

A close look at this definition will reveal that it incorporates all of the BHN components which have been discussed. A family which

gives equal weight to the discounted lifetime consumption of its youngest members will obviously have a strong interest in infant nutrition and investment in primary education. Similarly, adult members can be expected to give some value to expected profit-consumption in their future years, so they must pay attention to their own health and nutrition. The independent valuation of literacy is not explicitly incorporated in this definition, but to the extent that consumer activities requiring literacy exhibit a revenue elasticity which rises with income, the role of literacy is implicitly included.

Thus, we adopt as our presumed goal for poor families the maximization of the present value of profit consumption (PVPC) for all living family members, according to the best information available to them. If this definition is acceptable, we can move toward an analytical framework which will greatly aid us in drawing inferences about the behavior of the poor.

According to our definition of family accounts and the goals of economic activity, functionally-equivalent poor families with equal access to resources in a competitive bidding environment should be at essentially equivalent welfare levels by their own estimation. In a particular area, then, our view of the family leads rather directly to the acceptance of utility-equivalence for equivalent families. What does "equivalence" mean in this context? Here our notion of potential income comes into play again. Families should be regarded as equivalent if their opportunity income is actuarially identical.

In constructing valid comparison groups, we can draw on the fact that a very high proportion of variation in earnings can be explained by age-sex composition, education, ethnicity, and a few other family

characteristics. When families are actually observed, of course, they will be seen to have different expenditure patterns. Under the assumption of welfare-equivalence, this information can be used to infer the trade-off ratios which the poor apply across expenditure categories. Such information can in turn be extremely valuable for determining appropriate public projects from the viewpoint of the poor themselves. Under some circumstances, this utility-equivalence approach could be criticized for observing trade-offs at only one set of relative prices. This is a weakness which is common to many attempts to infer valuation from cross-sectional evidence. Since one of the main objectives of public policy is to expand the supply of valuable goods and services to the poor, thus lowering their prices, information deduced from the current set of trade-offs may well be insufficient.

It is precisely in this context that the re-definition of family accounts becomes most useful. All goods and services, after all, are marketed under conditions which incorporate considerations of access. In confronting the problem of variable access to the same commodities, therefore, the poor are in effect confronting consumption possibility sets at different relative prices. Rather than being restricted to an unvarying set of price ratios, the careful cross-sectional observer is in possession of a continuum of prices and quantities for most of the relevant goods and services.

How can this information be used in practice? Here it is necessary to go one theoretical step further in the analysis. We are modeling the poor family as a maximizing unit whose revenue potential is basically determined by the opportunity cost of time for each of its members. Family workers are substitutable in some activities and not in

others. In any case, the family should adjust the time inputs of its members until the marginal objective gain associated with a unit of opportunity expenditure is the same in all uses. As previously noted, it is impossible to deduce the marginal valuation put on alternative kinds of expenditure for most outputs (with some strategic exceptions, which will be mentioned). Fortunately, however, it is generally possible to deduce the marginal costs associated with particular expenditure types at particular places.

For poor families (as for all families), payment for any good or service has three components: The market price of the commodity itself, the opportunity value of the time spent in procurement, and the part of ground rent which is attributable to differential access to the source of the commodity. The "price" of the commodity in question, then, will be the sum of all three components. Since most kinds of expenditure fall into regular patterns, it is perfectly possible to standardize all price measurements on some time unit which is convenient for aggregation (rents, for example, are generally collected weekly or monthly, so other prices could be calibrated in the same way).

From the econometric point of view, of course, the imputation of ground rent to several access potentials and the imputation of value to the time of family members present a challenge. Both purchase prices and procurement times can be observed directly. Under the assumption of equivalent utility, the econometric problem can be resolved if sufficient observations are available. The exact way in which this would be done involves a rather technical discussion of utility function specification and econometric estimation (14).

Thus, the re-definition of family accounts seems to lead us down a

promising path. Enough information is available to yield fairly complete marginal benefit estimates for alternative goods and services from the perspective of poor families, which should provide valuable information to project planners.

## B. Full Income Accounts: A Detailed Look

In order to discuss the ways in which the revised definition of family accounts can affect public financial planning, a more detailed system will be tentatively proposed here. The main themes will be the role of the opportunity cost of time and its interaction with the problem of location in particular areas where residential sites and sources of goods and services are separated. If the urban land market is working normally, the weighted sum of the present values of time-cos savings and access advantages in particular contexts (employment, marketing, water, etc.) for all income classes should be observable as bid-rent surface. This complex interaction will result in one price -- land rent -- which represents a quasi-equilibrium arrangement under all the given conditions of access, preference, and income. In the account system which has been drawn up, these conditions will be strongly reflected.

A full tableau of the relevant price components for different "needs" and "non-needs" categories is presented in Table 5

---

(14) A more detailed discussion of the econometric approach will be included in the next section.

Table 5

Proposed Accounting System - Family Budget

| Category | Use Payments | | Access Payments | | |
|---|---|---|---|---|---|
| | Money | Time Value | Money | Time Value | |
| | | | Rent | Travel | |
| **Operating-Investment ("Basic Needs")** | | | | | |
| "High-Value Foods (a) | x(b) | x | x | x | x |
| Employment | | | x | x | x |
| Welfare Services | | | | | |
| Health | x | x | | x | x |
| Education | x | x | x | x | x |
| "Basic Materials" (c) | x | x | x | | |
| Density | | | x | | |
| Water | x | | x | | x |
| Sanitation | x | x | x | | |
| **Profit Consumption ("Non-Basic")** | | | | | |
| "Low-Value" Foods (a) | x | x | x | x | x |
| Other Consumer Goods | x | x | x | x | x |
| Clothing | x | x | x | x | x |
| Recreation (including (leisure time) | x | x | x | x | x |
| Shelter-Related | | | | | |
| "Non-Basic Materials" (c) | x | x | x | x | x |
| Furniture | x | x | x | x | x |
| Monetary Savings | x | | | | |

---

(a) In accord with the scheme suggested earlier, "basic" foods and beverages could be categorized as those whose gross revenue elasticity is high for the poor and falling; conversely for "non-basic" foods and beverages. Both these definitions depend, of course, on the principle of consumer sovereignty. An alternative would be the use of nutritional benefit/cost ratios.

(b) The "x's" refer to points in the table where survey information would have to be collected.

(c) "Basic materials" could well be defined in the same manner as "basic foods" -- high and declining gross revenue elasticity.

As previously noted, the primary criterion for discriminating between basic and non-basic items lies in a distinction between profit-consumption and operations-investment expenditures. Where ambiguity prevails, a further distinction can be made on the basis of revenue elasticities.

Food

As noted in the preceding chapter, food consumption can have all three characteristics (profit, maintenance, investment). Basic foods should be those in which the latter two qualities dominate, and such foods should logically be those with a high benefit/cost ratio as measured by nutrients/dollar of expenditure. Under the assumption that the poor are aware of nutritional value, comparative patterns of revenue elasticity change work well for discriminating between the "Basic" and "Non-Basic" categories. The available evidence suggests, fact, that there is no great difference between "optimum" diets for the poor (nutritionally defined) and those which the poor choose for themselves. Where divergences are apparent, many nutritionists would probably be reluctant to reject local practices in the absence of a careful look at local conditions. In rural areas, particularly, diets have evolved over long time periods through a process which should not easily be dismissed.

For each kind of food, as for many other goods and services, several kinds of payment are associated with the final procurement price. If the food is purchased in a market, some money will obviousl change hands. In addition, time must be spent in marketing and food

preparation by some member(s) of the family. The direct calculation of market-related costs must be supplemented by a consideration of access payments. Part of the rental of land in a particular location is undoubtedly associated with the discounted stream of travel time savings in marketing. In addition, of course, there will be money and time costs for transportation to and from the market to consider. At first, the imputation of a particular component of travel and time costs to particular commodities may seem impossible, since many trips are used for joint purchases. In such cases, however, it is perfectly possible to apportion costs among commodities on a value-proportion basis even if direct, detailed observation of unit time-allocations is unnecessarily complicated.

In the case of own-production, the situation is somewhat more complicated. Here the relevant notion is the income-opportunity value of the food, which could be defined either as its market value or the value of the family resources which go into its production, whichever is highest. To suppose that family resource costs are higher is to suppose a degree of irrationality among family members which is quite unlikely to be present, so that the appropriate opportunity cost will undoubtedly be the market value of the foods produced.

In an accounting sense, the addition of own-production to the budget can be accomplished by adding the value of the food as part of income and as one component of cost. Care should be taken to avoid double-counting here, which would be the case if the market value of the food and the value of the family resources used up in its production were both counted as "costs" in the accounts. The only legitimate cost in this context is the time cost associated with food preparation and

time required for transporting the food from field to home (15). Since family labor has to be used up in own-production, the value-equivalent has to be subtracted from potential income, as well.

Employment

Generally, employment is regarded as a source of income rather than as a cost. It is both, of course, since few families can locate next to sites where all of their members can find employment. Thus, resources are inevitably used up in journeys to work, and the true income from employment is income net of transport-related costs. In an accounting sense, there are two ways to deal with this problem. One is simply to enter work-related income net of transport costs; the other is to enter full work-related income and then to enter transport costs separately in the accounts. The latter option seems preferable, since access to work is one component of the location problem and is therefore one of the determinants of ground rent.

The inclusion of cost detail associated with employment-related income certainly does no harm, and such information may in fact be quite helpful in illuminating the trade-offs which are made by poor families. For example, an extension of econometric analysis to the inference of value-related weights on employment travel times may have important basic needs implications. Because transport connections in many poor cities are subject to high variance due to maintenance problems (with the attendant risk of firing in the face of disastrous tardiness),

---

(15) The latter is most analogous to family marketing cost, which is no absorbed by producers in the regular market.

families may choose co-location with the employment site of the main breadwinner as a form of insurance. The result could be tremendous crowding and sanitation problems in adjacent areas, coupled with the welfare-opportunity loss associated with very high rents. In such cases, a very defensible basic needs policy would be an investment in systematic maintenance for public transport (or some policy of inspection-related subsidy to encourage maintenance in private transport vehicles).

Hopefully, this example points again to the overriding importance of the location problem in a systematic consideration of basic needs strategies. With an appropriately general view of the trade-offs actually made by the poor, it may be easier to realize large gains in welfare at relatively low social cost.

Welfare Services (Health)

It is in the analysis of "free" or "low-priced" public service provision that the opportunity-flow analysis of family economic activity really comes into its own. Health services are generally offered at fixed sites (except in the case of mobile vaccination units), with the usual problems of differential access. To the extent that health services are seen as contributing to the specified goal (PVPC for all family members), then, we would expect some willingness to incur cost in procuring them. Indeed, the problem of access guarantees that the poor will always pay for health services.

Once this fundamental notion has been accepted, certain conclusions must follow. First, existing patterns of behavior by poor

families should result from the same sort of balancing of marginal costs and benefits which exists in other economic domains. Secondly, the existence of varying benefits and costs associated with public health facilities undoubtedly puts them into competition with private facilities in many cases. Such private services will run the gamut from moonlighting surgeons to full-time traditional herbalists and shamen. The latter may well be sought out (at a fee) by desperately poor families, even when modern-sector "free" alternatives are available.

In the case of health, as in the case of food, a whole set of cost factors needs to be considered in determining the full price paid by the poor. Some money payment is generally required for services, even in the public sector. Substantial waiting time is often involved as well, and this will be particularly true for low-priced mass public services. Again time has value, and the combination of rental increment, travel payment, and travel time value shows up in the calculation. Medical service demands are generally linked to evident need, so that their distribution within time units is likely to be random. Nevertheless, their incidence across longer intervals will be reasonably predictable (and higher, undoubtedly, for the poor than for the rich).

Thus, it would certainly be a mistake to assume that the poor make little or no provision for medical needs in their full set of economic arrangements. The costs of untreated illness are all too evident to the poor themselves as personal suffering, work income lost, and shortened life expectancy. Again, a full analysis of opportunity flows is required to reveal the true willingness of the poor to pay for medical services at different quality levels.

Welfare Services (Education)

While medical expenditures can justifiably be viewed as some combination of maintenance and investment costs, educational expenditures are undoubtedly the purest of investments undertaken by many poor families. Again, a full-flows analysis is necessary to determine the true savings propensity of poor families, since the opportunity cost of an adolescent's time may be relaively significant in many rural and urban areas. If education really is regarded as an investment by the family, then it should be undertaken only when the discounted flow of expected returns exceeds the discounted flow of opportunity costs for the family as a whole. In fact, the reported behavior of poor families lends substantial credence to this view (16). High absentee rates in many Third World schools suggest a conscious decision by parents to maintain their perception of an economic allocation of their children's time, as does strong seasonal absenteeism in agricultural areas.

Because schooling always costs poor families in an opportunity sense, it is again reasonable to suppose that private alternatives to public education should exist. This is of course the case in every country where such activities are legal (and in most, undoubtedly, where they are not). Private schools, school-business arrangements, and apprenticeships of many kinds exist in poor cities. In addition,

---

(16) The presence of investor-like behavior is evident, even if some would claim that the poor calculate costs and benefits inappropriately. As we have already mentioned, there is little evidence to support the notion that the risk-return analysis of the experts is likely to be better than that of the poor themselves.

current research on entry-level employment arrangements in large enterprises suggests an important training component, since there seems to be some willingness on the part of new employees to part with current opportunity income in exchange for higher compensating payments as training and experience increase.

In countries where private arrangements for education seem to flourish in competition with the public-sector arrangements, it seems futile to attempt a basic needs policy in education without a thorough analysis of the perceived willingness of the poor to pay for educational alternatives.

### Shelter (Basic Materials)

The problem of separating basic needs expenditures from others in the provision of shelter is similar to the problem for food. The easiest basis for discrimination, as usual, is separation by gross-expenditure elasticity. Building materials with initially high (and declining) elasticities would be regarded as basic in this dichotomy, with the converse true for their counterparts. It is always the case that poor families are most concerned with having a roof (of some kind) over their heads, and it is unquestionably the transition from "non-shelter" to "shelter" which has the most impact on physical welfare, provided that densities are not extreme. Basic materials may, of course, be priced implicitly or explicitly in the income-expenditure accounts of poor families. In addition, families may be renting one shelter while investing in another. In doing the accounts properly, it is again necessary to take opportunity costs into consideration.

According to the criterion adopted, only declining-elasticity materials should be regarded as basic. In some cases, the market value of these materials will be deducible from shelter rentals along with access components. In other cases, they will be directly observable as families purchase (or produce) materials for their own homes. In all cases several components of cost will again be present. For renters, materials values (along with the resource cost of combining them) can be inferred from rental data, as previously explained. For purchasers the situation is a little more complex, since the full cost of new housing also includes labor costs (own- and hired) and transport costs, among others.

In any case, a full accounting should be possible along the lines suggested by the analysis of basic food consumption. The simplest accounting technique would undoubtedly be evaluation at market rates for rental units and at resource opportunity costs for units being built. Although one kind of expenditure is generally termed "consumption" and the other "investment," nothing is lost in this case by aggregating them together as basic needs outlays. By the same principle, any observed expenditures (in opportunity terms, again) on increasing-elasticity building materials (along with some proportional allocation of labor time in the case of own-production) should be relegated to the non-basic category.

### Shelter-Related (Density)

Once basic shelter from the elements has been accounted for, the only other residence-specific factor which is likely to be of concern

from the perspective of basic welfare is density of habitation (and even this has not been proven in any systematic way. Some evidence suggests that the incidence rates for many contagious diseases, mental disorders, crime, and children's learning rates are all related to the level of stress associated with living at densities above some "tolerance threshold" which is culturally defined. Thus, the purchase of residential density by the poor is likely to affect fundamental welfare patterns.

Density is usually lumped together with structure quality and accessibility in the determination of residential rental rates, so that econometric inference is necessary to determine its separate value in the market. It is assumed here that few if any truly poor families will be above the threshold density level previously mentioned, so that all density-related payments can be defined as basic needs expenditures.

Shelter-Related (Water)

Of all the goods and services which have been discussed, water is the most important in a fundamental sense. At the same time, water is a commodity whose true cost to the poor can be quite difficult to measure in many circumstances. Generally, water is available under differing conditions of quality and access. The full price of available water is often included in housing rental, so that individual components of value in the market must again be deduced econometrically. Even "free" water is by no means free, since access costs are likely to be very high if it is clean and plentiful. The evaluation of water expenditure may also include a monetary charge if it is purchased from a vendor or on some

unit-fee basis. The unit monetary value will presumably vary with quality type, and this must be defined (all water types can safely be termed "basic" for our purposes). Frequently the procurement of water involves queuing, and the associated opportunity costs must be evaluated.

Again, as in the case of education, the full-income approach seems necessary for a sensible analysis of public water policy for the poor. In many cases the observable money expenditure for water may be small, and a failure to incorporate the access-based rental increment and the opportunity value of time spent in hauling water will lead to a tremendous underestimate of the value of water to the poor. In addition, the full analysis of costs allows for a much more comprehensive consideration of alternative approaches to the financing of water-supply expansion.

### Shelter-Related (Sanitation)

The final component of shelter-related welfare expenditure which has a strong basic needs interpretation is the purchase of services for the disposal of household wastes. The problem is in some respects similar to that of water, although it is principally the quality of service (rather than any question of quantity) which is at issue here. In some cases (e.g. night-soil men), there may be a direct price associated with the provision of sanitary services. When they are provided on the premises (at different quality levels, which must be defined), the phenomenon of queuing may again be a problem (particularly for children, and for adults at rush hous points in the morning and

evening), so that time costs may be a factor. Finally, rent increments are again likely to be charged for access to particular sanitary arrangements, and these have to be taken into account.

Profit-Consumption (Non-Basic) Expenditures

It has seemed eminently worthwhile to describe the accounting components for needs-based expenditures at some length, since they will figure in the public finance discussion and it will be possible to introduce them in the revised accounting framework at that point without much further trouble. Profit-consumption, on the other hand, is less important here. Although an accounting of this expenditure is necessary, it does not need to be presented in any detail at this point. Thus, it can briefly be said that all of the listed components of current profit consumption have five cost-indices, since some component of access value is operating in each case.

## C. Planning Strategy: Project Analysis and Net Subsidy Accounting

As we have already seen, basic needs planning has several features which differ substantially from those assumed by conventional benefit-cost analysis. The productivity of social investments is largely unknown, and the relative valuationn of social goods such as literacy, health, and longevity remains problematic and politically explosive. Recurrent costs join early-period capital costs as a dominant feature, and the whole emphasis is on the incidence of net benefits. For all these reasons, a methodology for the effective operation of basic needs programs must focus on efficient on-going administration as well as efficient prior selection. Because the prior selection problem is more familiar, we will begin with a consideration of the relevant aspects of project analysis. In the second part of this section, we will propose the use of "net subsidy accounting" as a tool for effective administration. This would enable public planners to adjust net subsidy levels across sectors in response to changes in recurrent cost flows or the levels and social distributions of basic welfare indicators. The methodology of social benefit-cost analysis will serve as the conceptual basis for our approach.

### (1). Project Analysis - Benefit Estimation

If we attempted to choose BHN projects in the standard way, we would have to value project outputs in one of two ways. First, we could adopt a "top-down" approach to valuation, which would require:

-Some hedonic welfare function to be maximized (i.e. some linear weighted sum of components such as productivity, literacy, longevity, etc.).

-A model of hedonic "production," with quantities of goods and services consumed as inputs.

This is simple enough to say, but impossible to implement. A quick enumeration of the major objections should prove persuasive. It is not at all clear what rational basis exists for specifying unit trade-off weights for the social objective function (which is more important, an extra year of average longevity, or a 5 percent reduction in the rate of illiteracy?). The choice of components for such a function will always remain controversial. In addition, it is doubtful that the social welfare function should be linear (implying the utility-independence of hedonic attributes). In addition, the specification and estimation of an adequate hedonic production model is at present only possible at a very aggregate level (17). In particular national cases, even the most commonly-addressed questions (such as the climate-variable relationship between basic nutrients and health) are not satisfactorily resolved.

Thus, we can probably dismiss "top-down" valuation as a reasonable approach to BHN benefit estimation. As a second alternative, we might consider the approach which is usually taken in project analysis: The evaluation of benefits at current market prices. Only a moment's reflection suffices for a rejection of this approach, as well. Given local supply curves, local market prices are determined by aggregate

---

(17) The growth-welfare model presented in Appendix B provides an example of this kind of approach.

demand curves which are heavily weighted toward the preferences of the rich. In poor countries with extreme income inequality, this source of bias is compounded. Obviously, we have to look elsewhere for a solution to the problem of valuation.

As we have argued in Chapter I, there are strong reasons for supposing that the revealed preferences of the poor themselves provide a sound guide for benefit estimation. We have argued that the ability of public planners to arrive at relative valuations which are superior to those of the poor themselves is highly questionable. A full economic analysis of the behavior of the poor also suggests another reason for using their revealed preferences whenever possible: If the valuation of particular goods and services by the poor is relatively low, then either huge recurrent subsidies or (expensive) coercion will be needed to promote their consumption. Since this whole study has been inspired by the obvious scarcity of recurrent subsidy resources, it does not seem very sensible to adopt valuations which are very different than those of the poor themselves.

The trick, of course, is to determine what these valuations actually are. Two ways of identifying the preferences of the poor can be imagined: First, direct surveys which confronted the poor with alternative possibilities and costs might be utilized, so that some sense of desirable allocations at current prices could be established. Secondly, indirect estimation of preferences from comprehensive household survey information could be done. Such an approach would have one obvious advantage. It would indicate the trade-offs which the poor really choose to make at current price ratios, as opposed to the choices which they say they would make. In addition, such a survey would

provide information on budgetary allocations by the "nearly poor," whose behavior is highly likely to serve as a model for that of the poor.

After giving rather detailed consideration to the econometric approach, we will present a brief decription of experiences with direct preference surveys. Once both approaches have been introduced, we will present an evaluation of their relative merits.

Econometric Estimation: The Welfare-Equivalence Approach

The main econometric task in this context is the design of a model of behavior which is plausible and which allows for inferences to be drawn about price-responsiveness in the demand for BHN goods and services by poor families. As previously noted, a promising possibility is the use of the full income-expenditure approach in a model which takes spatial access into account.

The key to inference is the assumption of equivalent welfare at equal potential income levels, and the notion of potential income is an actuarial one. Allowing for normal distributions of advantageous personal attributes and luck, this approach regards family income as the sum of incomes which family members could earn in a particular area if they chose to work full time. The income level is in turn predicted from observations on individual age, sex, education, ethnicity, urban residency status, and other variables.

Although incomes are assigned actuarial equivalence, measured incomes will of course not be equal for two reasons. First, randomly-distributed personal characteristics and luck will intervene, as previously mentioned. Secondly, two families may behave very

differently in their allocation of time and money, so that the level of monetization in their exchange transactions (and their observable money income) is quite different. For econometric purposes, the use of acutarial equivalence is the key to inference because it allows us to impose a specific model of welfare determination. If families have the same income-earning resources (in this case, human resources -- people as poor as those under consideration generally don't have any physical capital), then they are taken to be equally well off, regardless of the way in which they choose to use their resources.

Is this equal-welfare assumption plausible? If people at the same potential income level have good information about alternative consumption patterns and a substantial willingness to move, they should adjust their circumstances until they are at equivalent levels of satisfaction. Information and degree of inertia are obviously two relevant variables here, and as we have noted, the neoclassical and evolutionary models make very different assumptions about their role in economic behavior.

Its imposing of the equal-welfare criterion makes the econometric approach inherently neoclassical, so it is worthwhile to ask whether there seems to be any observational justification for such an approach. Only anecdotal evidence is available on this subject, and we conclude that it certainly doe not discredit the neoclassical assumption. In Lagos, Nigeria, for example, as many as 1/3 of all poor families shift locations in any given year. This suggests a very substantial willingness to adapt to newly-perceived opportunities and to experiment with new circumstances.

Under such conditions of high mobility, the assumption that

welfare levels are the same up to some level of randomly-distributed error does not seem to be unreasonable. If it is accepted, then families with equivalent potential incomes can be lumped together in subsamples and the condition of welfare equality imposed in the presence of observable variations in economic behavior to yield inferences concerning relative utility-weightings and price-responsiveness. In the absence of information on time allocation, marginal rates of substitution at average time-money price ratios can be inferred (18). When time allocations are known as well, price-responsiveness can also be calculated.

While information on marginal rates of substitution can be quite useful as a guide to direction in public investment policy, the existence of information on price responsiveness can yield actual estimates of willingness to pay and the consequent level of recurrent subsidy necessary to particular activities, under the assumption that appropriate recovery mechanisms can be established. In order to illustrate the possibilities, one fairly simple (and therefore probably implementable) version of this approach will be presented here, along with a discussion of the way in which it could be used as an aid to planning. The treatment will be general, because data appropriate for such an exercise have simply been unavailable.

We begin with the problem of selecting an appropriate subsample of family observations. The families in question would not have to be precisely identical, although they should be closely matched. Thus, equivalent status could be assigned to two families of five with the

---

(18) For pioneering work in this area, see Wheaton (1977).

same ethnic and urban residency status, each containing a man in his thirties with a primary education, an illiterate woman in her twenties, a late-adolescent boy, an early-adolescent girl, and an infant. For all such families, we could define equivalent utility in the following way (19):

$$(1) \quad \ln U\{o\} = \ln (M + R) + Sm\{i,n\}[B\{ij\} * \ln X\{ij\}]$$

where

M = Total money expenditure (on goods and services for the family

R = Ground rent paid by the family

X{ij} = Time spent on acquiring the ith good or service by a family member in the jth age-sex group

The imposition of welfare equivalence aloows for the use of constant utility on the left-hand side of the equation. The two log-additive components are taken to be money expenditures (including ground rent, which should adjust significantly to compensate for differential time use in a space-scarce urban market) and time spent in pursuits other than direct money-earning. The form of separability imposed on the utility function seems defensible in this case, under the assumption that money income (which equals M+R -- money expenditure) is freely transferable among competing uses.

It is possible to re-formulate (1) as a stochastic equation in the following way:

---

(19) In this report, Sm{1,n} will be used to denote summation from i=1 to i=n.

$$(2)\quad \ln (M+R) = \ln U\{o\} - Sm\{i,n\}[\ B\{ij\}*X\{ij\}\ ] + e\{ij\}$$

When the equation is specified in this form, it is possible to use observations on total money expenditure and the allocation of time by family members to obtain etimates for the B's, as well as a value-equivalent estimate for ln U{o}. In this simple specification, the assumption is maintained that the stochastic error term (e) is randomly (and normally) distributed, reflecting the presence of unobservable personal traits and sheer luck in the determination of departures from precise equivalence in welfare levels. There are no apparent failures of the classical error assumptions in this case, so that ordinary least squares should yield unbiased, efficient estimates (20).

The resulting estimates of equation parameters allow for the estimation of the money value imputed to time in particular uses by family members (generally reflecting rent adjustments as proximity changes) as well as marginal rates of substitution between time allocations within the family. Once this approach has been used to obtain a set of time-value estimates, these can be combined with prices (where the latter are charged) to obtain full-price demand schedules for

---

(20) The imposition of a log-linear utility function, of course, represents several strong assumptions. Nothing in this way of formulating the problem precludes the imposition of a more complex functional form such as the generalized CES:

$$(3a)\quad U\{o\} = (M+R)**B0 + Sm\{i,n\}[\ B1\{ij\} * (X\{ij\}**B2\{ij\}\ ]$$

$$(3b)\quad M+R = [\ U\{o\} - Sm\{i,n\}(\ B\{ij\}*(X\{ij\}**B2\{ij\}\ )\ ]**(1/B0)$$

Here, of course, non-linear regression is necessary to obtain parameter estimates.

particular goods and services. It is these schedules, in turn, which should allow for the estimation of the demand for the services of particular basic needs facilities (education, health, water, etc.) under alternative assumptions about pricing and siting. The equation in form (2) can be used as a means of estimating the total willingness to pay for particular goods and services, and this in turn can serve as the basis for calculating the general subsidy level necessary for a particular project.

## Econometric Estimation: The Problem of Discrete Choice

As noted many times in this report, it is logical to suppose that the recurrent cost burden will be minimized when net subsidy flows are in accord with the preferences of the poor whenever possible. Full income analysis has been suggested as a means of improving the planner's ability to deduce the true allocation of resources across profit, maintenance, and investment activities. Undoubtedly, this way of measuring consumer preferences can be very useful in the determination of a strategy for the allocation of subsidies across sectors. At the same time, however, additional guidance is needed for the choice of precise activities which will be subsidized.

In analyzing the activities of the poor, the analyst confronts the set of general issues which must always be faced by the applied microeconomist. In the set of implicit and explicit markets for goods and services which serve as the foci for economic activities, many goods are divisible and can be analyzed using statistical techniques which rely on the principle of continuity. A good example is provided by the

measurement of valuation of residential attributes which was discussed previously. The use of continuous econometric models based on the principles of substitutability and equivalent utility is to be recommended in such cases.

Unfortunately, many of the choices faced by public planners cannot be informed by this sort of continuous analysis. Whenever basic needs activities are strongly associated with public goods, deliveries must be organized through public institutions which take on particular characteristics. Education, for example, can be organized in a finite number of ways, and the government must make explicit choices for institutional development. Similarly, medical care can be organized in a variety of ways. In deciding what kinds of schools and medical facilities to build, and what kinds of personnel to train, the government is forced to make discrete choices. When the institutions themselves are in place, poor consumers will themselves then have to make decisions governing the incidence and frequency of use.

Such problems are not exactly unknown in microeconomic analysis, of course. Many of the classic consumer choice problems involve the choice among commodities which are "lumpy," so that the standard model of continuous substitution does not readily apply. Until the present decade such discrete choice problems were largely ignored by applied microeconomists, largely because because existing methodologies were simply inadequate for proper analysis.

Recently, however, this situation has changed. Beginning in the 1960's, an interest in urban mass transit led to a considerable expansion of the empirical literature on "modal choices" by consumers. In thinking about the problem of alternative investments in transit

modes, policy analysts were motivated by the obvious need to determine the preferences of consumers. While preserving the basic structure of revealed preference theory which underlies all of microeconomic analysis, they had to promote a rather drastic revision of the standard econometric approach to measurement in order to take discrete choices into account.

Although this modeling technique was therefore originally designed to handle a particular public policy problem in the United States, the utility of the approach goes considerably beyond that which may have been envisioned by its original designers. In fact, it is precisely the technique which is required for a consistent analysis of the preferences of the poor when they are confronted by alternative institutional modes for the provision of basic needs services such as health and education.

This section will include a consideration of the general approach to modeling which is appropriate in the context of discrete choice, and a relatively brief description of the modeling technique. The applicability of the technique will be illustrated with a description of a project currently underway in rural Kenya which was inspired by the student-faculty seminars at Boston University which were conducted as part of this Basic Needs Project, and by the background research on discrete choice analysis whose findings are summarized in this chapter.

The section will conclude with a voluminous Appendix which provides a comprehensive, detailed explanation of the state of the art in discrete choice modeling. As far as we know, no such exposition is currently available from any other source. Much of the pioneering work in this field has only been published in the professional journals during the past two years, and several of the techniques discussed in

the Appendix have only now reached the stage of application in on-going research at MIT and elsewhere. Obviously, we would not have devoted so much time and attention to this subject if we did not judge it to be of fundamental importance to BHN analysis. It is difficult to see how further progress in the analysis of service mode choices by poor consumers can be made unless the family of techniques described in this section is brought to bear. Although the Appendix is highly technical in parts, we hope that it will be useful to USAID specialists in the application of econometric techniques to policy problems of this type.

a) Discrete Choices by the Poor and the Public Investment Problem

As noted in the introduction to this report, any comprehensive analysis of the consumption of the poor should take existing modes for the provision of goods and services into account. In both rural and urban areas, poor people generally have at least some choice in determining what kinds of provision they will enjoy. In the field of health care, for example, the poor can avail themselves of numerous alternative modes. At the most basic level, they can engage in self-diagnosis and cure with the aid of medicines bought over the counter in local market areas or traditional medicines produced from local materials. In the case of the latter, traditional healers with various specialties may perform a role in the process. Curative medical assistance can generally be obtained from private practitioners in market towns or urban areas, although the cost may be prohibitive for the very poor. In many areas, a secondary market in medical services has sprung up, centering on the provision of care by nursing graduates

who have experience but are not formally licensed to practice medicine. At the same time, public authorities may operate proximate dispensaries or hospitals with out-patient services, and mission societies may offer complementary facilities.

In the field of health care, then, it would be a mistake to characterize the choice environment of the poor too simplistically. Education provides another example of this kind of multifaceted phenomenon. Two educational choices available to the poor are obvious: Primary education in some form is now available in many locations throughout the Third World, and those families which judge the opportunity cost of such education to be too high can always choose to introduce their young to traditional modes of agriculture or animal husbandry. Some reflection, however, reveals that other educational modes are also commonly available to the poor.

The analysis of rural-urban migration patterns frequently reveals the importance of apprenticeship arrangements as the first stage in the movement of young workers from rural agriculture to full-time urban employment. Apprentice arrangements are often made through extended families, with labor provided in return for room, board, and training in a particular trade or light production activity. In addition, public disillusion with poorly-supported public primary education has now progressed to the point where commercial alternatives have sprung up to provide training in services or the industrial arts. Youthful labor is the primary currency in this system, and many commercial schools produce goods and services for profit. Although exploitative arrangements undoubtedly exist, the persistence of commercial training programs suggests that they fill a genuine need which has not been well satisfied

by alternative public arrangements (21).

Again, then, it must be noted that the educational choices of poor families are frequently made in a multimodal context. In determining policies for the expansion of education, it would seem appropriate for public planners to look at the revealed preferences of the poor before deciding on the shape of future institutional offerings. This said, it must be admitted that the problem of measurement is not a simple one. In the prior discussion of measurement in the context of

---

(21) In this connection, it is interesting to note that the fundamental notion underlying commercial training programs has been appropriated by the Cubans as a part of their educational system. The Cuban government has long emphasized its commitment to "practice" as well as theory in the education of the nation's youth. In rural areas, pupils from the primary grades on up spend part of each day in agricultural production. In urban secondary schools it is not uncommon for students to spend part of each day in light industrial activities, such as the assembly of transistor radios. The difference in the Cuban case, of course, is that the "profit" goes to the state rather than to educational entrepreneurs.

From one perspective, the Cuban approach can be seen as an imaginative partial solution to the recurrent cost problem in the provision of educational services. Skeptics from other Third World countries have questioned the general utility of such a substantial allocation of time to non-scholastic activities in publically-sponsored education. In Jamaica, by contrast, the schools have been devoted to more traditional instruction, but the rate of absenteeism is quite high. Many poor Jamaican families have chosen to allocate only part of their children's time to education, with the rest devoted to assistance with family production and marketing

This juxtaposition of the Cuban and Jamaican cases serves once again to illustrate the two conflicting philosophies which can underlie basic needs policies. The Cubans have decided that the externalities associated wth education are so large that they have enforced relatively faithful attendance at educational institutions. At the same time, they have in effect expropriated a significant proportion of the labor time of poor families in order to support these educational activities. In the Jamaican case, laissez-faire is more prevalent. As could be expected, poor families left to their own devices choose to retain some of their children's labor time as a contribution to the acquisition of goods and services other than education. Those who judge the Cuban system to be better must certainly count themselves among those who believe in tremendous externalities associated with education, unless they simply support the notion that the state knows the self-interest of the poor better than they know it themselves.

residential choices, it was proposed that the statistical approach to revealed preferences is likely to yield more reliable results than the kinds of direct consumer surveys which are feasible in practice. The

b) Modeling Discrete Choices: The Fundamental Issues

At the most basic level, there is little difference between the theory underlying discrete choice analysis and that which serves as the foundation for the microeconomic analysis of consumer behavior in the continuous case. With all consumption possibilities available at positive opportunity cost (measured in money- or time-equivalents), and with goods and services substitutable for one another as sources of satisfaction, rational consumers in the continuous case can be expected to adjust their resource expenditures until they have equated (marginal utility/price) ratios for all goods and services. In some sense, this same logic should also apply to cases in which available goods and services are "lumpy."

In the analysis of consumer choice among competing "lumpy" forms of a good or service, we can adopt the usual near-tautology by asserting that the choice made will be superior in the perception of the consumer in question, given the relative characteristics of the competing modes, their prices, and the preferences of the consumer. For the postulate of rationality to hold, we would expect identical consumers to make the same choice among identical modes.

In the case of continuous choices, it is possible to use well-established econometric methodologies to estimate the relative valuation accorded incremental units of alternative goods and services

by the "representative" poor family. In the section on residential choice, an approach originally due to Wheaton (1977) was used to illustrate the measurement of relative valuation in his context.

Here, the problem is somewhat more difficult. When the consumer chooses among a set of goods or services which are complete, mutually exclusive competitors, all possibilities save one will be excluded in the result. The classical analysis seems to break down in such situations, and the notion of "hedonic" choice becomes fundamental to an appropriate re-structuring of the analysis.

For half a decade now, the theory of consumer behavior has moved progressively toward Kelvin Lancaster's hedonic specification of choice. According to this notion, very few commodities have direct value to consumers. Rather, almost all goods and services are actually bundles of attributes which themselves enter into consumer utility functions. Different foods, for example, can be regarded as alternative combinations of nutrients and tastes. It is possible to make sensible decompositions of almost all imaginable goods and services in this hedonic sense.

Where consumption possibilities are continuously divisible, this Lancastrian view of consumer behavior can aid in simplifying the choice set which is being modeled. In cases of multimodal choice, the use of the hedonic specification turns out to be even more helpful. Rather than regarding alternative modes as discrete entities, we can regard them as alternative combinations of valued attributes. It is then possible to regard the ultimate choice of one mode in a way which is consistent with the theory of revealed preference -- the combination chosen must represent a vector sum of hedonic valuations which is superior to that

associated with any other available combination.

This hedonic specification is the real key to the generation of an effective measurement methodology in the discrete choice case. Suppose that we regard alternative means of providing health and educational services, not as distinct entities, but as alternative combinations of valued attributes. From the perspective of public policy, it is clearly the attributes themselves which are of fundamental importance. Once alternative patterns of valuation are understood, then the minimization of recurrent support cost would imply a strong preference for the provision of modes whose attribute "scores" are superior for particular target groups. At the same time, the use of clever measurement should make it possible to discuss the provision of modes which, while they may not currently exist, would be superior to any existing modes in the eyes of interested poor consumers.

Obviously, we have not yet arrived at a concrete methodology which would actually permit the kind of measurement mentioned above. By combining hedonic specification with an appropriate understanding of the term "representative consumer" in microeconomics, however, we can move to an understanding of the general statistical approach to measurement which is required.

As in the standard microeconomic case, the key to modeling discrete choices lies in the notion of a utility function which characterizes the "representative consumer." Since individuals or families are never perfectly identical, of course, this is an idealized construct. Since social planning must be directed toward perceived central tendencies in target groups, however, this construct remains an essential one. For the planner, the problem is to specify the

components of the consumer's utility function which have relevance for public policy, to adopt a specification of the utility function which makes sense behaviorally, and to use an appropriate econometric technique with data on consumer choices to deduce the relative valuations placed on releva components by the "representative consumer." These measured valuations can in turn be used as a guide for policy in cases where the values of consumers are taken to be valuable guides for allocation.

In the discrete choice case, the first step is the specification of an appropriate utility function and the hedonic attribute components which characterize the choice modes under consideration. Once this has been accomplished, appropriate survey data have to be gathered for analysis. Obviously, the values of all hedonic attributes have to be indexed for available alternative modes. At the same time, the characteristics of consumers which may bear on the formation of preferences must be recorded. Finally, the modal choices actually made by the consumers surveyed must be registered.

It is at this stage that the major econometric puzzle presents itself. Rather than observing some level of consumption of each available modal offering, we are confronted with a very discontinuous set of observations for each consumer. One mode will have been chosen; the rest will have been ignored. How should such a result be recorded for effective statistical work? The answer lies in the theory of probability, which provides the final building block for the measurement of utility in discrete choice cases.

The most important underlying notion in this context can undoubtedly be communicated with the aid of a simple example. Suppose

that we are observing the behavior of several consumers who can choose between two alternative modes of consumption. In a pure random sample, we observe that in 2000 cases, mode A is chosen 500 times. When asked to predict the behavior of another consumer drawn at random when confronted with the same choices, we would naturally answer in a contingent mode: The probability that the consumer will choose mode A is approximately .25. We would expect mode B to be chosen by approximately 75 percent of the consumers in the population.

Our prediction is an example of an unconditional probability calculation. Knowing nothing of the characteristics of the consumers or of modes A and B, we can simply take advantage of random sampling and the law of large numbers to produce a simple prediction of behavior in a bimodal context. The same basic principle can be applied to our consumer choice problem, except that conditional probability becomes the operative principle. If we know the hedonic attribute values for particular modes and we have complementary information on the characteristics of consumers whose choices have been observed, we should be able to improve our prediction. In producing the statistics necessary for conditional prediction, we will also have produced the information on relative hedonic valuation which is useful for policy planning.

At the risk of heroic simplification, the basis for econometric work will be briefly sketched here. Although we have all learned simple probability so well that the process of calculation is automatic, it is apparent on reflection that the unconditional probabilities mentioned above were produced as follows: For mode A, a choice by some consumer in the sample results in the awarding of a score of "1" for A and "0" for B. The converse results when B is chosen. The unconditional

probability that A will be chosen by a consumer selected at random is simply the mean score for in the sample (that is, the sum of the 1's and 0's, divided by the sample size), and similarly for B.

Now, consider the modal choice case. Again, a mode is either chosen or not chosen. If it is chosen, it awarded a score of 1; if not, a 0 is awarded. For consumers with particular attributes, there will be certain characteristics which will make one mode more valuable than the others. Here we do not base econometric inference on measured quantities of goods or services consumed. Rather, we use a score of 1 ("chosen") or 0("not chosen") as the unit of observation of consumption. In econometric work with this kind of binary data, estimated coefficients have an interpretation which is different in form but not in substance from those which characterize standard utility analysis. Rather than measuring relative contribution to utility in the sense examined for residential analysis, coefficients in this context can be interpreted as measuring the relative contribution of one increment in a particular attribute value to the probability that a particular mode will be chosen over its competitors.

Once a full set of probability coefficients has been estimated for an appropriate set of consumers, then prediction becomes possible. Given a set of competitive modes with known attribute values, it should be possible to use the estimated parameters to determine the mode which has the highest probability of selection by the representative consumer. In fact, the calculation of relative conditional probabilities allows us to stand the original unconditional calculation on its head: Once we have calculated selection probabilities for each of the competitive modes, then we can also take a randomly-selected groups of consumers of

arbitrary size and predict their distribution across the available modal choices. For public planners interested in potential rates of utilization of service modes by representative populations in particular areas, the value of such predictions is obviously far from negligible.

The preceding paragraphs have incorporated an attempt to summarize a relatively complex, sophisticated approach to measurement. Much has fallen by the wayside in the attempt, and an illustration drawn from a typical basic needs planning problem will be used as a means for further amplification of the approach. Before passing to the illustration, however, it should be pointed out to the interested reader that a careful perusal of the lengthy Appendix to this chapter may amply repay itself at this point. The Appendix sets out in great detail the theoretical and probabilistic underpinnings of the econometric approach to valuation in discrete choice cases, along with a detailed explanation of the major competitive techniques which are currently available. is hoped that this detailed explanation will be of interest to those who are concerned with the effective use of survey data for facilities planning in the context of basic needs.

c) An Illustration: Health Mode Choices by the Rural Poor

A good example of the utility of discrete choice analysis in the context of basic needs planning is provided by the problem of siting and design of health facilities for the poor in rural areas (22). As has

---

(22) This section summarizes a measurement strategy which has been derived by the authors of this report and Germanu Mwabu, a doctoral student at Boston University, for thesis work by the latter in Kenya. This thesis, along with others currently in progress at Boston

been stressed repeatedly in this report, the problem of willingness to pay cannot be escaped in analyzing the relative desirability of BHN projects from the perspective of recurrent cost. The poor will always have to pay for access to any scarce good or service, whether in direct payment, time, or ground rent. The rapid expansion of BHN provision at supportable standards depends crucially upon the ability of public planners to identify the ways in which the poor are already paying for basic needs, and to translate these implicit or explicit payment flows into forms which can absorb a substantial portion of the recurrent costs associated with BHN provision.

The planning of rural health facilities provides one concrete example of this sort of policy problem. As in most other BHN categories which depend on recurrent government support, rural health care systems in many poor countries are in a state of disarray. Siting is frequently done arbitrarily; personnel go unpaid for long periods, with consequent impacts on morale and service efficiency; crucial medicines are frequently unavailable; and even those services which are available are often underutilized by those for whom they were intended.

Under these circumstances, rural health planners face a multifaceted problem. They must provide service facilities and

---

University, was proposed by Mr. Mwabu as a result of his participation in the workshop which formed one focus of the Basic Needs Project. In Mr. Mwabu's case the inspiration proved quite fruitful, since he is currently on-site in rural Kenya collecting sample information with the full support of the Rockefeller Foundation. Upon his return to Boston University, he will collaborate with the authors in an analysis of rural health choices along the lines sketched in this chapter. It is hoped that the results will be of assistance to the Kenyan Ministry of Health in its planning for the future provision of appropriate rural health facilities which minimize the recurrent subsidy necessary for attracting the poor.

personnel which are simultaneously appropriate for local conditions, supportable under local financial constraints, accessible to target populations, and acceptable to the people whom they are intended to serve. Under the best of circumstances, such a planning problem would not be an easy one. In the case of rural health care, all problems are compounded by substantial ignorance concerning the preferences of the client population itself. Although health facilities are intended to serve the rural poor, planning is frequently done in almost total ignorance of the preferences of the poor themselves, the environmental context in which they would choose rural clinic facilities over alternative modes, and the degree to which they might have a willingness to pay for medical services which they found to be relatively desirable.

An appropriate information base would obviously represent a desirable starting point under such conditions, and discrete choice modeling seems to provide a good vehicle for organizing a large quantity of potentially relevant information in a consistent, usable form. The first step in the analysis must be the identification of health care modes which are actually available to the rural poor. In fact, the number turns out to be surprisingly large. People in rural areas can simply self-treat, either with traditional medicines or with drugs purchased over the counter in nearby markets; they can consult traditional healers, who may well be specialized; they can visit private clinics run by doctors or nurse practitioners in nearby towns; they can go to mission facilities; or they can visit government-run clinics or hospitals. Each mode will be available to a particular rural family at a cost, of course. For privately procured drugs, treatment by traditional healers, or visits to private clinics, the cost will be in

time and money. The use of government facilities, on the other hand, may simply involve the expenditure of time which could be employed in other activities. Since government health services are generally not rationed by price, waiting times are frequently quite lengthy in such facilities.

Coupled with the problem of cost in the choice of health care mode is the problem of perceived quality. Alternative modes will generally differ along several quality dimensions. For different kinds of symptoms, the probability of successful treatment will undoubtedly differ. Differential availability of medicines will be a factor, as will be the speed, efficiency, and courtesy of the services offered by alternative modes. In addition, the credentials of the service personnel themselves will weigh in many decisions concerning modal choice when particular symptoms arise.

Modal characteristics will be joined by personal characteristics in influencing the final choice of service. Education and prior urban residency will have a bearing on the perceived desirability of traditional modes, ceteris paribus. Undoubtedly, family income will have an important effect on the responsiveness of decisions to the perceived costs of alternative modes.

Such a problem is made to order for discrete choice analysis. representative sample of families in a particular rural area can be drawn and observed. All relevant modes can be registered and modal choices in response to various symptoms recorded. Once data on modal choices, modal characteristics, and family characteristics have been obtained, they can be used to estimate the parameters of a utility-based choice model of the type previously discussed. The results of such an

estimation exercise are potentially of great assistance to rural health planners who are attempting to maximize coverage at minimum cost. In a particular area, for example, analysis of the data may well reveal that the poor are extremely sensitive to the opportunity cost associated with time spent in traveling and waiting for "free" medical care. A steeply-rising marginal distance-disutility curve may produce as its dual a substantial willingness to pay for proximate private services, even when the unit charges are substantial by local standards.

Since private facilities may well maintain a more dependable inventory of medicines, among other things, it would be necessary for health planners to be capable of distinguishing between proximity and dependability effects in the choices of rural health consumers. It is in this context that multivariate analysis comes into its own, of course. Given sufficient data, it is possible to deduce the separate effects of these and all other relevant variables on relative modal valuations by representative consumers. When the inevitable trade-offs in facilities planning are made, the information provided by these relative valuation measurements should prove invaluable in arriving at sensible allocations.

From the perspective of recurrent cost analysis, the sort of econometric approach outlined here seems to offer the possibility of relevant measurements which are both invaluable and unobtainable by other means. When opportunity cost factors, incomes, family characteristics, and modal quality characteristics are combined in an appropriately-specified multivariate analysis, it becomes possible to obtain a relatively accurate perception of the willingness of the rural poor to pay for service alternatives and the likelihood that they will

take advantage of alternative modes once they are made available. Since cost and coverage are the two essential criteria in this context, it seems that the argument for this sort of modeling approach is quite strong from a practical perspective.

## Direct Survey Techniques

Although much useful information could obviously be gotten from econometric inference, it seems easier at first glance to approach the preference problem more directly. It is always possible to survey the opinions of those who will be the immediate beneficiaries of BHN programs. Of course, such surveys must be truly economic in nature to be relevant. Not only the project options, but their direct cost to the client group must be specified. In fact, information which is truly comparable to that supplied by the econometric approach can be obtained only if people are asked about their choices at varying prices. It is possible to design games which elicit client choices under hypothetical price-variable conditions, and some experiments have been run in West Africa and Mexico (23). Because of the rapid escalation in price-quantity combinations as the choice set is enlarged, this approach seems to be feasible only under certain relatively limited conditions. In both cases cited, choices were elicited for basic residential service projects (variable-quality housing, water, and sanitation arrangements).

---

(23) Good descriptions of these exercises have not been published, but we have obtained much relevant information in the course of conversations with Michael Cohen of the World Bank (West Africa) and Tomasz Sudra, an independent consultant-planner, who has been responsible for much of the work in Mexico.

A Comparison

It is possible to define circumstances under which each approach makes sense. When a particular type of BHN program (e.g. residential, health,education) has already been chosen and the client population is small and spatially stable, the direct survey approach is reasonable. The most ideal situation for a direct survey is that of one rural village in its dealings with a particular ministry. When any of the three basic conditions -- sectoral focus, small group size, and group stability -- is violated, however, the appeal of direct surveys begins to erode very rapidly.

The relative advantage of sample-based inference is most immediately apparent in the opposite case. If strategic BHN choices are to be made across sectoral categories for the urban poor, sample-based econometric inference is probably the only feasible way to approach the problem. Only actual observation can yield reliable evidence concerning economic behavior when the available choice set is so large and varied.

In addition, the absence of locational stability in urban areas can cause serious difficulties for the direct survey approach. If a particular area is chosen as the focus for a BHN investment, the state has effectively chosen the current residents of that area as lottery winners. Under these conditions, poor clients are apt to push survey results in directions which will maximize their potential windfall. Thus, a poor family head in a favored area may indicate a strong (and unfelt) preference for elaborate toilet facilities, knowing full well that these are highly prized by the middle class and can yield a high

value in resale once they have been provided. Hopefully, it is not unduly cynical to suppose that such an emphasis on careful calculation may dominate simple gratitude on the part of the poor (24).

(2). Project Analysis - Cost Evaluation

Although benefits in the BHN context are more difficult to estimate than is usually the case, the estimation of costs can draw almost totally on existing methodologies for project evaluation (25). From the government's point of view, the true cost of the project in each period is the social value of the resources which it uses up. In the social equivalent of cash flow analysis, costs are generally divided into capital (project construction) and recurrent (project operation) categories, subtracted from estimated yearly benefits, and the present value of the net benefit flow to the end of the project is calculated.

In principle, comparative present value calculations provide a good basis for choosing worthwhile projects. In reality, of course, few governments in poor countries have made consistent use of cost-benefit analysis as a tool for social evaluation. This is partly because skilled analysts have been quite scarce and partly because broad allocations across BHN sectors have been determined politically at the

---

(24) Again, some World Bank experience with housing projects in West Africa supports this view. In one case known to staff members, rent levels almost immediately doubled in a squatter settlement after the Bank announced its intention to support an upgrading project in the area.

(25) Any of the standard approaches is relevant here. See for example Sen-Marglin (1972), Little-Mirrlees (1974), Squire-Van der Tak (1975), or Schydlowsky (1973). A good discussion of the underlying differences can be found in Lal (1974).

highest levels, with subsequent programs for BHN delivery replicated in roughly standardized form.

In practice, this somewhat haphazard approach to program design and implementation has made at least two major contributions to the recurrent cost problem. The foremost contribution has been a pronounced tendency to underestimate or simply ignore the recurrent cost component of BHN projects in long-range budgetary planning. The unfortunate association of "development" with capital budgeting alone has resulted from a failure to pay much attention to true production functions for BHN goods and services. As a result, yearly non-defense allocations have tended to be disproportionately devoted to the capital budget in the name of rapid progress.

The second major recurrent cost problem which has been created by inattention to prior project analysis is somewhat more deeply buried but no less real. Even BHN planners who have attempted to make correct recurrent cost projections have tended to treat them only as cash flow problems. To reiterate, however, the only resource valuation which is really meaningful in this context is the social valuation. Generaly, recurrent cost elements in prior project planning can be subdivided into three broad cost categories: Labor, materials, and land. The first two categories may have both domestic-source components and foreign exchange components. Thus, at least four general shadow values must be included in a social recurrent cost accounting -- domestic labor, domestic materials, foreign exchange, and land. In the absence of such an accounting, the true cost to society of any program cannot even be approximately known.

There are really two aspects of this second problem which are of

particular interest here. The first is the social cost calculation which has just been discussed. The second is tied to the problem of land. As we have repeatedly asserted, a critical component of basic needs satisfaction for the poor is the problem of access to appropriate delivery sites, and the basic needs delivery problem is critically related to the distribution of activities over space. In urban areas, it may be particularlly important to take opportunity cost into account when locating some BHN delivery site. From the government's perspective, the opportunity value of the land is generally the potential tax revenue foregone in the most valuable alternative use. Unless this value is considered as part of the project planning account, one true but invisible recurrent cost component for a BHN delivery site may be very large indeed.

Thus far, we have been considering only the prior planning problems associated with BHN projects. We have recommended that a heavy emphasis on valuation by the poor be coupled with the social valuation of recurrent (as well as capital) costs in thinking about the true net subsidy burden which will have to be borne by the state. In the spirit of benefit-cost analysis, one further point concerning the evaluation of social benefits must be added. The relevant basis for valuing the true social value of a project is the net contribution to delivery capacity which it makes. As we noted previously, no service is really free to the poor, and well-developed markets for all BHN goods and services will exist even if the government does nothing. In most cases, the implementation of a new project will result in some diversion of demand from existing suppliers. In cases (e.g. education) where competitors are declared to be illegal or patronage of government-sponsored programs is

required, the true welfare gain for the poor may be a good deal less than anticipated.

(3). Project Monitoring - Net Subsidy Accounting

It is important to organize BHN investment activities in ways which will yield the highest expected net present value when social benefits and costs are taken into account. In the previous sketch of a recurrent-cost minimizing approach to this problem, we have emphasized the importance of input opportunity values as the basis for calculating true anticipated recurrent costs. Unfortunately, few projects turn out wholly as expected. In poor countries, even well-elaborated project plans admit to very substantial degrees of uncertainty, and subsequent events often overturn the original calculations entirely (26). Thus, it is not only to be feared but expected that many social projects will have surprising outcomes.

As noted in USAID (1974), program planners are unavoidably cast in an experimental role (27). Although they can initiate projects with

---

(26) Few treatments of this subject surpass that of Hirschman (1967). His discussion of the Pakistan Karnaphuli paper mill case is particularly entertaining.
(27) The discussion in this section (and in the sector-specific chapters) is intended as a complement to the standard USAID Evaluation Handbook (1974) rather than as a critique. The Handbook is designed to introduce field workers to statistically-based methods for running experiments, and its discussion of ways to evaluate the relationship between project intentions and results, with attendant costs, is simple and clear. In this report, we are focusing on general recurrent cost problems, so our approach is somewhat different. Our main emphasis is on observation of the simultaneous evolution of net subsidies (current-account drains on the government treasury) and the apparent results of projects to which the subsidies are tied. Thus, we focus more on payment flows. In addition, we are more concerned with the measurement of opportunity values as an adjunct to more conventional

certain models of social and economic behavior in mind, it is not at all clear that either the models themselves will turn out to be valid or that variables which they hve assumed to be exogenous will remain so. As we have seen, BHN projects have particular problems because functions in which BHN inputs combine to produce observable welfare outputs are complex and generally unknown. Another large degree of uncertainty attaches to the question of ultimate benefit incidence in such projects. In Chapter I, some attention was devoted to the problem of arbitrage in goals-oriented BHN projects with "lottery-like" characteristics. The net result of such activities may be the capturing of long-run subsidy flows by income groups who were never intended to be beneficiaries. Thus, a useful approach to BHN project monitoring must have two basic characteristics: It must be capable of revealing the annual flow of true net subsidies, and of determining the apparent pay-off to this flow with respect to both impacts on the welfare of target groups and diversions to other groups.

The challenge is to develop and implement an information and evaluation system which will aid public planners in shifting toward "preferred" allocational patterns where the incidence of costs and benefits shifts continually and in unexpected ways. The organizing theme for the sort of analysis which will be proposed is again provided by the methodology of social benefit-cost analysis. However, we are interested in recurrent rather than capital cost, so that the focus is on the pricing and disposition of on-going public welfare activities.

---

financial flows data. Basically, however, we are still concerned with the on-going evaluation of relative costs and benefits, and the design and evaluation questions considered by the Handbook are certainly important elements in this same process.

Discounting should therefore not be a major component of the analysis and little effort should be devoted to the calculation of net benefit measures in comparable value terms. Rather, an appropriate system should allow public planners to monitor the annual changes in key social indicators and the evolution of net subsidies across the basic needs activities which are obviously linked to those indicators. The approach should be explicitly multi-objective, with relative weightings remaining the task of the ultimate decision-makers. The system should incorporate an implicit "reduced form" model of behavior, since it would allow planners to observe the covariation of public inputs (net subsidies) and outputs (basic welfare indicators) over time.

What sort of accounting system would be appropriate in this context? The following elements are obviously crucial:

(1). Key indicators must be held to some minimal essential set (28).

(2). These indicators must be reliable, so effective sample survey design and execution are central to the process.

(3). The accounting system must be set up to provide at least three kinds of information:

-Financial flows, so that yearly net financial subsidies can be calculated across activity types.

-Opportunity flows, where all relevant inputs and outputs are appropriately shadow-priced.

-Incidence calculations, which show yearly net income effects for the public sector and by income class for alternative project types.

Although details would have to be situation-specific, the

---

(28) Again, many of the suggestions in USAID (1974) are useful in this regard.

following sketch of the accounting system should at least provide some notion of major system components.

Welfare Indicators

The major indicies of project benefits in a social accounting system must remain suggestive, since their true relation to underlying components of welfare can only be guessed. Nevertheless, the availability of project (or area) specific data on a relatively small set of social indicators would certainly be helpful. If a relatively comprehensive accounting system is to be established, the prevalent skill constraints in LDC's dictate a very small number of indicators and a willingness to accept indicator approximations from "windshield surveys" and other rough sampling methods in many cases. Since the main purpose of the whole system is to suggest trends in benefits and net subsidies at the level of first approximations, such an approach may be perfectly adequate. For the BHN issues which have been of concern to us in this report, a representative small set for a particular area is

presented in Table 6.

Table 6

Indices

Residence

Density

Health

Infant Mortality Rate
Chronic Illness Rate

Education

Competency Scores
Unemployment of Graduates by Curriculum Type

Water

Daily Consumption Rates
Cleanliness

Sanitation

Sewerate Arrangements
Garbage Disposal

Recipients

Families in project area, by full income class

There can be little doubt that the availability of frequently-collected area-specific statistics like those above would be invaluable to planners who are concerned with yearly budgetary allocations. In the absence of such information, it is difficult to see

how any reasonable discussion of alternatives would be possible. It is particularly important to note that the suggested approach is area specific rather than family specific. By comparison with family-specific experimental approaches, it gains in convenience but loses in validity. Since the main interest here lies in adjusting net subsidy levels across areas as local circumstances change, the area approach seems preferable.

### The Net Subsidy Accounts

As previously mentioned, three kinds of accounts are suggested in this context. They will be presented here in order.

#### Financial Subsidy Accounts

Although the ultimate concern of public planners should be with opportunity flows, it is impractical to administer development projects without specific attention to cash flow accounts as well. A representative combined (capital-recurrent) net subsidy accounting system is presented in Table 7 (29).

---

(29) The best approach here seems to be to regard all investment allocations as untied, so that investment costs are incurred during the period of project construction in all the accounting tables.

Table 7

REVENUES

Priced Services

Source

Rich
Middle
Poor

Taxes

- RECURRENT COSTS

Operations

Domestic

Labor (by Recipient Income Group)

Rich
Middle
Poor

Materials

Imported

(Same Breakdown)

Overhead

(Same Breakdown)

- CAPITAL COSTS

Domestic

Labor (by Recipient Income Group)

Equipment
Materials

Imported

(Same Breakdown)

= NET SUBSIDY

If such accounts could be integrated across areas and across sectors, they could be combined with area-specific distribution and

welfare indicators to provide a good sense of use efficiency for recurrent subsidies in different contexts. Another advantage of this system would be its ability to highlight both the results of government investments and the adjustable policy instruments at the government's disposal. Central among these would be service prices, property and sales taxes, and the current rate of investment in new projects. If sufficient fiscal control could be maintained, available taxes and prices could be adjusted in ways which yielded an ongoing investment fund. If such a system were instituted, the job of government financial managers and policy planners would be greatly eased, since informed discussion of alternatives would be possible.

With on-going monitoring of benefit distribution, it should be possible for the government to identify projects whose benefits have been essentially captured by more affluent groups with the passage of time. Such projects could be "unloaded" into the private market through auctions, yielding the addition of capitalized benefits to the investment fund and the elimination of the public subsidy burden. Given the scarcity of administrative resources, such an approach would probably make sense even if captured projects were yielding some profit for the government. Instant capitalization and reallocation to projects directly benefitting the poor would seem to be called for.

## Opportunity Accounts

Although the financial accounts are important, they should be supplemented by a parallel set of opportunity accounts. These accounts would provide a way of monitoring the yearly net flow of opportunity values in the basic needs system. Several differences between this and the financial accounts should be emphasized.

First (and this is particularly true for residential development projects), the opportunity tax flow should be incorporated as a cost. Secondly, labor should be appropriately shadow-priced by type. Finally, an appropriate shadow value should be attached to foreign exchange. An obvious advantage of this approach would be the adjustment of shadow rates as the external situation changed. Thus, it would be possible for planners to identify trends in the opportunity value of net subsidies to particular activities. Again, this might allow administrators to adjust the application levels of policy instruments (prices, tax rates, auction sales, new investment rates).

## Net Incidence Accounts

Finally, a set of accounts should be developed which provides explicit information on net benefit distribution. Essentially, this kind of information would represent a re-arrangement of the elements in the first set of accounts, with some important differences. For example, an additional obvious benefit of residential development projects is the increment in income (net of taxes and rentals) provided to project proprietors through market-clearing arrangements. This

increment can be divided into flows to three income classes, so that an accounting of net flows in the project context would look like the following:

(1). Government (already known): Rents + Taxes - Domestic Wages - Materials - Foreign Exchange

(2). Income Classes: Extra Rent + Wages (this is under the assumption that "willingness to pay" is operative, so that the prevailing rent level in the parallel market is an index of perceived benefits).

Although it would always be somewhat incomplete, this "net flows" information would provide a very desirable complement to information about the distribution of benefits as indexed by welfare indices.

## III. FULL INCOME ANALYSIS AND RESOURCE MOBILIZATION: STRATEGIC ISSUES

The preceding chapters have been designed to establish a general framework within which our approach to recurrent cost analysis can be consistently developed. Much attention has been paid to the basic planning approach because most program recommendations must follow loically from choices made at this higher level. Thus, our advocacy of full income analysis as a basis for BHN policy formulation follows from our conclusion that the best guide for BHN intervention in most cases is provided by the economic behavior of poor families themselves. The full income approach seems particularly important to us because it focuses attention on the inevitability of payment for goods and services by the poor. Once this inevitability has been accepted, the discussion can usefully be shifted to the practical level-- In what forms can this payment be most effectively mobilized as a source of support for current and future BHN efforts?

As we hope to have made clear in the preceding sections, our emphasis on payment by the poor does not mean that we advocate increasing their burdens. Rather, we would argue that a full analysis of income and expenditure simply reveals the magnitude of the effort which the poor themselves are making. Under existing fiscal constraints, their continuing assumption of much of the cost of BHN delivery is inevitable. The challenge for policy research in this area is to identify ways in which government subsidies can most usefully complement payments by the poor in assuring a growing and improving flow of BHN goods and services.

This chapter will be devoted to identifying the crucial policy issues at a very general level. All stem from one central problem: Under existing arrangements, payments provided by the poor cannot effectively utilized. The general source of this problem, in turn, is the inappropriate structure of existing assets whose services must combine with direct payments by the poor to support the delivery of BHN products. Its primary reflection in urban areas is the private ownership of land and shelter services, for which the poor must bid in the attempt to gain access to employment and sites where essential goods and services are supplied. In rural areas, on the other hand, its primary reflection is the difference between optimal settlement patterns dictated by relatively high-level BHN service provision and land-tenure patterns associated with low-technology agricultural production. In both urban and rural areas, another major barrier to rapid improvement in BHN flows is the absence of capital assets which can be combined with local resources to provide higher levels of income and recurrent cost support.

### A. The Urban Problem: Net Subsidy Flows and the Capitalization Process

This report was originally inspired by the monotonous regularity of recurrent cost problems in public projects intended to help the poor Although investments in utilities (e.g. water, sanitation, housing) are generally characterized by a lower relative recurrent cost component than human services (e.g. health, education), all share the problem of serious shortfalls in the ability to pay for on-going maintenance, personnel, and materials. In urban settings, the problems of water

provision provide a classic case in point. Public-water authorities are frequently given quasi-independent financial status, which means that their rate of service expansion is directly linked to their ability to recover costs. In spite of effective monopoly power and the ability to enforce cross subsidies through differential prices, public water corporations are repeatedly forced to expand in poor residential areas at very slow rates because they perceive that the payments which they can extract from the poor for water provision will not even cover long-run average cost.

Under this constraint and the requirement of financial independence, public water providers are forced to devote substantial resources to the extension of upgraded services to relatively wealthy families whose payments will refill the cross-subsidy pool. Thus, the most basic of needs -- clean water -- is provided to the poor at a rate which is frequently disappointing. It is true, of course, that other fiscal mechanisms could be brought to bear and the shortfall to the water authority made good by the central treasury. The problem with this approach is the familiar one: The funds provided would come at substantial opportunity cost, and in many cases they would simply be subtracted from other BHN projects.

Thus, the situation is made to seem nearly hopeless. However, when the circumstances under which the poor cannot "afford" to cover at least long-run average cost for water provision are examined a little more closely, some paradoxes begin to emerge. It has frequently been observed that poor families who have no access to piped water will pay water vendors for inferior deliveries at a price which would be sufficient to cover the average costs of installing piped service (30).

At the same time, water authorities also report that the poor who live near their newly-installed lines suffer from a manifest inability to pay for the service.

The paradox is, of course, no paradox at all. The poor who live near water lines cannot afford to pay the water authority because they have already paid for the water. The collectors, however, are their landlords. Existing empirical work on the contribution of water valuation to rental payments by the poor are scarce, although existing econometric studies suggest that a large component of shelter rent is paid by the poor for proximity to good water. If these results can be generalized (and observed water-vendor payments in cities as varied as Lagos, Nigeria, and Lima, Peru, certainly suggest that they can be), then the recurrent cost problem in urban water provision may not be as formidable as it appears.

The case of water has been chosen because it provides an extreme example of the phenomenon in question, but we think that the underlying principle is of general importance: To the degree that the enjoyment of public subsidy flows by the poor must occur jointly with their use of assets which are in private hands, some proportion of the total subsidy will be transferred to assets owners through a process of bidding. It is in the interest of public policy, therefore, to give specific attention to the distribution of assets ownership in determining appropriate financial policies for the sectors which are in question.

This principle has relevance for all three kinds of productive assets -- physical capital, human capital, and land. In poor urban

---

(30) This observation has become rather common for all three poor continents. See for example Grimes (1977).

\\\

areas, however, the recurrent cost problem seems most inescapably tied to the question of land ownership. Land is of major importance because most needs-based welfare problems for the poor focus on the problem of physical access to the relevant services: Children must get to school and back; the sick must have access to medical facilities; the family must be fairly near a water source; and employees must have access to the workplace.

Insofar as they value their time, the poor will be willing to pay a premium for access to valuable service sites. If the family is considered as the relevant decision unit, the value ascribed to particular kinds of access will vary with anticipated net benefits associated with differential productivities, among other things (31). Thus, the poor will undoubtedly play an important role in determining the familiar urban bid-rent surface. Given the fact that ground rents are determined by a competitive bidding process and that most urban land is privately-owned, it is clear that incremental benefits to improved access or service quality which are not captured by user fees will be partly captured by shelter owners. The degree of capture will depend, of course, on the prevailing price elasticities of demand and supply. In the case of water, demand is relatively price inelastic, for example. In poor cities where one-story building is the only practical alternative, interior usable space at relevant locations is quite price-inelastic in supply. Thus, much of the intended benefits from water are likely to be transferred to landlords in the absence of additional public intervention.

---

(31) This subject has already been discussed in the section on full income analysis.

What is true of land is also true of human and physical capital i other contexts. The use of scarce public resources in the training of doctor, for example, allows a private person to acquire an income-earning asset at essentially no cost. In the subse distribution of returns among the state, the poor, and the doctor, the state and the poor will unquestionably emerge as the losers. Since the state will not be able to recover its cost, supply expansion will be relatively slow and monopoly pricing arrangements will be likely to prevail.

It is one thing to identify a general problem and another thing t solve it, of course. In wealthy nations, the mind of the policy analys turns naturally to valorization charges as a way of recapturing the capitalized value of net government subsidies. This issue will be explored in depth in the next section of the report. For the moment, w will continue the identification and discussion of fundamental policy problems at a general level.

B. The Urban-Rural Split

Given the fact that most people in any country reside in communities of some size, and the existence of an approximate rank-size ordering for these communities, any division between urban and rural is somewhat arbitrary. There is, however, a relationship between the aggregate stock of available community resources and the stochastic determination of demand for different goods and services which determines community "size" in the sense which is important for this analysis.

Communities which are physically small, scattered, and poor will be considered rural for our purposes. Large rural settlements will be classed as "urban" if land owrership is non-communal and aggregate income is sufficient to support specialized health and education services at prevailing opportunity cost levels. The reasons for this peculiar distinction lie in our specification of the kinds of problems which characterize urban and rural areas. Two particularly visible differences are important for the analysis which will underlie our recurrent cost approach.

First, the spatial continuity of urban settlement patterns is disrupted in rural areas. Villages are distributed in response to random environmental features, and the relatively large distances between them neutralize the kind of substitution which characterizes access bidding in large urban areas. This discontinuity means that the poor in rural areas simply cannot substitute time for money in the ways which are open to their urban counterparts. Given some desirable facility at a particular location, rural people face a discontinuous decision: Co-locate, or do without the service.

The second characteristic which distinguishes rural areas is the difference in ownership patterns. The relevant ancillary resources in rural areas may or may not be in the hands of the same poor people who will benefit from the installation of BHN projects. Here it is impossible to generalize, because rural land tenure patterns differ strongly across regions (the latifundia are still common in Latin America, for example, while small, semi-communal villages dominate the landscape of rural Africa). In areas where the poor are tenant farmers or landless laborers, much of the analysis which has been applied to

urban cases may be relevant. Landlords will again have control of an important ancillary resource, and they can be expected to extract a proportionate share of the net benefit flow generated by projects. In semi-communal situations where villages are widely distributed, the situation is different, of course. The net subsidy flows from projects will still be present, but the beneficiary may well be the village as a collective group. Scarcity bidding will not be as prevalent in small rural villages where ethnic and family ties are likely to play the majo role in determining the right of settlement. Redistribution within families through movement between villages may occur, but this is of no great consequence since the people helped are likely to be of the same class as those who were originally benefitted.

### C. The Rural Problem: Spatial Discontinuity and Opportunity Cost in BHN Provision

With at least a rough distinction between residential groups at our disposal, we can now turn our attention to the "village," which wil serve as a convenient term for rural groupings under our definition. A mentioned earlier, our examination of BHN programs in poor rural areas has led us to the conclusion that a major source of financial failure i the difference between the optimal distributions of agricultural land holdings and BHN delivery sites. We will begin with a consideration of the simple economics underlying this difference, and illustrate the principles involved with a discussion of some recent attempts to implement integrated rural development plans.

While BHN advocates generally list the same set of basic needs in urban and rural areas, a few differences seem to have emerged by

consensus. Most fundamentally, the "housing problem" for the poor is generally not considered in a rural context, although shelter-related problems of water provision and sanitation are certainly on everyone's list. Since the rural poor are often more wretchedly housed than their urban counterparts, this omission seems curious. Although it is never discussed, the obvious implication is that only residential density is of fundamental interest (32). In most cases, the relative dispersion of small living groups in rural villages is sufficient to suppress the density problem. With housing eliminated from the list and adequate rural space available for rudimentary waste disposal, the primary subjects for BHN analysis (besides food, whose production economics are the other facet of the problem under discussion) become water, health, and education services. In analyzing the structure of demand for these three kinds of service, we are immediately led to a problem of varying critical scale which is stochastically determined.

Within any appreciable unit of time, all the people in a particular area will demand water. The demand for educational services will involve fewer people (but on a very regular basis), while the demand for health services will come from even fewer people per unit time. Since each expected demand level is sensitive to the "full price" (in the sense which we have defined) of acquisition, it is necessary to consider the supply side as well. If village residents wish to augment traditional sources of water, education, and health care they will confront an unavoidable relationship between levels of delivery system responsiveness and cost. As a general proposition, it is reasonable to

---

(32) If this is correct, then many suggestions for housing the urban poor as a contribution to "basic needs" are simply inappropriate.

assert that the recurrent cost of a delivery system will rise as system responsiveness increases (33). A well needs no built-in responsiveness once it has been dug, and its operational cost is therefore negligible. Modern health and education services, on the other hand, are costly at any level of responsiveness. It is well- established tha cost rises faster than effective resonsiveness.

Thus, the basic economics of supply and demand would dictate very different rural service availabilities for villages in the absence of recurrent subsidies. The combination of universal demand, a high-price alternative, and negligible recurrent cost would guarantee a strong tendency toward the installation of wells. Since all variable values g the other way for health and education, however, the opposite is to be expected. Significant individual scale economies for health and education are augmented by interactive effects which make the joint provision of services highly desirable. This kind of connectivity greatly enhances the appeal of integrated rural development programs which attempt to collect scattered rural villages into larger serviced settlements. In theory, the social desirability of such projects is great because large groups will exhibit a stochastically-determined willingness to pay for basic services which will significantly reduce the level of recurrent subsidy necessary for providing them.

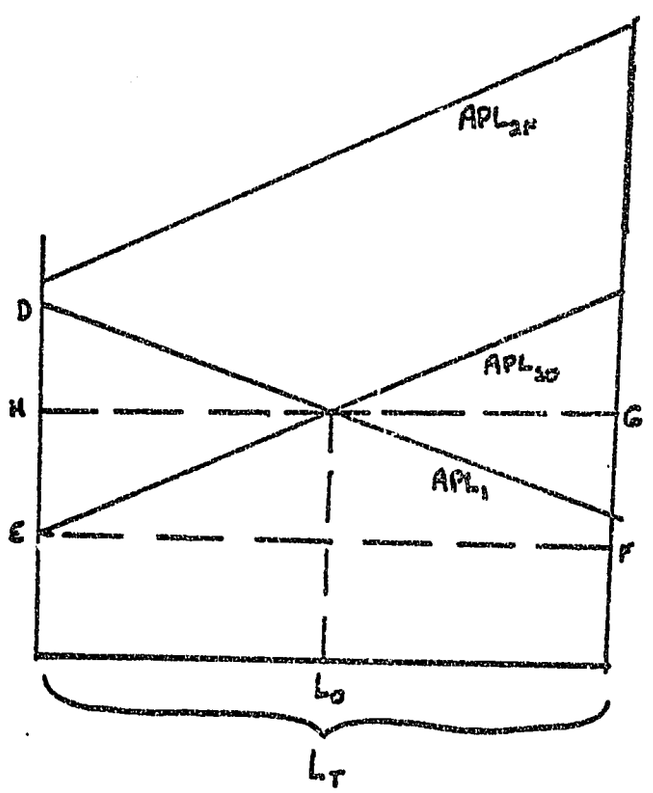
If the analysis could truly be concluded at this point, then a broadly acceptable approach to the recurrent cost problem in rural area would be available. However, the analysis is faulty because it has pai insufficient attention to opportunity costs. It is this failure, in

---

(33) An obvious illustration of this principle is the classic adage about running a successful restaurant -- keep the menu limited.

fact, which seems to have led to many of the problems common to rural development schemes. Unfortunately, the discontinuity associated with fixed-site provision in rural areas leaves farmers with a discontinuous choice. Either they relocate to the service site, abandoning the fields which were previously cultivated, or they remain and fail to take advantage of the new service. Thus, the true cost of a rural project which attracts sufficient clients to perform at an appropriate scale may be quite high, at least in the short run. Some idea about this can be gleaned from Figure 6, which describes the allocation of a fixed labor supply between two villages and the resulting productivity pattern.

Figure 6



Let the total labor supply for the two villages be measured by the length of the horizontal axis in this diagram. Under current production techniques, each village has a declining schedule for average labor productivity which reflects the application of more labor to land whose quality diminishes with expansion and the time loss involved in farming land which is farther and farther from the village. Under the assumption that village ethics are redistributional and that no irreconcilable differences (ethnic or otherwise) divide the two groups, the long-run allocation between the two villages should be around L{0}, the point of equal average labor productivity. (This is not necessarily the output-maximizing allocation, since we are not assuming marginal-productivity-based payments).

As previously noted, this is an expected long-run allocation which should occur slowly enough to keep migration-based conflicts to a minimum. Now consider the problem which must result if the government attempts to set up a rural needs program which focuses on centers large enough to capture major scale economies in the relevant service provision. Suppose village {2} is chosen as the site for investment. In the long run, the goal is to raise the APL curve from APL{20} to APL{2F}. This result is expected to come from the cumulative effects of better health and education services, along with the benefits provided by superior water access and enhanced food production from the use of better techniques.

In theory, the contribution to prouctivity made by all of these investments could be at least partly recaptured through appropriate institutional arrangements, and a substantial component of recurrent cost along with fixed costs could be provided from the flow of project

benefits in the long run. The problem, of course, is that long-run prospects will do little to solve the financial problem in the short run. If the villagers from V{2} arrive in V{1}, it is reasonable to suppose that their reception will very cold if any reduction in current income is anticipated. The V{2} people, on the other hand, will be very difficult to uproot if their medium-term expectation in the new area is below the same in their former village. In the short run, the government will face an agricultural output loss of EFGH which will have to be compensated if the transfer is to be made voluntarily. In fact, some villagers will remain in Village {2} until compensation is raised to at least DCGH, since the average productivity of labor in village {2} will rise to this limit as the population begins moving to village {1}.

Obviously, a voluntary transfer in cases where public and private discount rates and (what is probably more important) public and private expectations concerning average payoffs to the project differ significantly is likely to be a very expensive proposition. Difficulties and the need for additional compensation will undoubtedly arise as a result of: (1). Unwillingness on the part of V{2} residents to abandon agricultural land; (2). Political and social problems associated with combining two village structures into one; (3). The sudden assertion of property rights by current villagers in the face of a sudden influx and the attempt to extract rents from the newcomers.

The last problem could be approached by establishing a completely new village, but most of the other problems would simply be compounded. With all this said, it should not be imagined that the evident difficulty of implementing such an integrated approach has prevented various countries from attempting it. A tremendous number have in fact

tried this approach at some level. It is of interest here simply to give a brief review of some of the results for countries where very ambitious attempts along these lines have been made.

D. Integrated Rural Development: Some Illustrative Experiences

Among non-communist poor nations, the rural development efforts of Tanzania are perhaps the best known (34). In underlying rationale, the ujamaa program was established for reasons which are almost exactly identical to those previously mentioned. The ujamaa vilages were to be communal operations run at sufficient scale to capture significant economies while allowing for productivity advances through the promotion of infrastructure, extension services, and marketing. Although a strongly-expressed commitment to this approach was evident in the late 1960's, by 1975 the program had broken down so badly that it was abandoned. In its place, the government initiated a "villagization" program which concentrated on the assembly of small villages into large ones in order to capture service scale economies. The present status of the latter program is not entirely clear, although it is obvious that difficulties are being encountered.

General discussions of the failure of ujamaa tend to focus on three issues: (1). The reluctance of villagers to move; (2). The failure of agricultural productivity to rise as anticipated; and (3). The imposition of a recurrent cost load which was simply unsupportable. While climatic problems and fluctuations in foreign exchange earnings

---

(34) For a detailed treatment of recent Tanzanian experience, see Nyerere (1975), Sabot (1975), Clark (1975), and Cliffe (1975).

certainly added to the difficulty during this period, there can be little doubt that the system as originally planned could not have met its recurrent cost burden anyway.

The original analysis of the economics of this kind of program may reveal part of the underlying problem. At the most fundamental level, the government was asking Tanzanian peasants to change their consumption-savings allocation drastically in fulfillment of a plan tuned to the government's expectation of investment pay-off and the government's discount rate. The graphical apparatus in Figure 6 attempts to analyze the result. From the peasant's point of view, the anticipated effect was a drop in current average income in exchange for government reassurances about long-run growth and welfare dividends.

Not surprisingly, many peasants proved to be reluctant participants in ujamaa, and this has continued to be the case for the villagization program. In spite of persistent instructions from President Nyerere to avoid the use of force, many breaches of discipline at local levels have been reported. In some cases, peasants have reportedly been burned out of their existing villages.

The Tanzanian government has also proved unable to muster the administrative, financial, and technical resources necessary for a rapid up-shift of labor productivity in project areas. The result has been unfortunate in many cases: A deadweight loss in agricultural output, borne mostly by the peasants, continued rapid rural-urban migration by people attempting to escape from these situations, and, undoubtedly, increased hostility on the part of the peasants to the whole notion of nontraditional programs and policies.

An interesting parallel to the Tanzanian case is provided by the

Cuban new communities program, which has apparently been successful by some criteria during the past half-decade (35). In spirit, the Cuban program is like ujamaa, but at a much higher material level. As part of the program, rural new towns are constructed with many amenities (new apartments, well-staffed schools, clinics, etc.). Farmers in the areas surrounding these developments are given the option of surrendering their land in exchange for housing and employment in the new communities. The long-run plan is to combine small plots into large areas where agricultural production can be "rationalized" (i.e. where critical scale can be achieved for the application of fertilizers, equipment, etc.), while excess labor will be absorbed in new industries which will co-locate in the new towns.

Most of the available evidence suggests that little or no coercion has been used in the Cuban program, but the new communities are not lacking for in-migrants. If older farmers in the countryside choose to remain on their private plots at a level of living substantially below that offered by the new communities, their children very frequently make the opposite decision.

One obvious conclusion in the Cuban case is that the short-run subsidy to such operations must be huge. Under the assumption that the extremely generous subsidies from the U.S.S.R. are essentially untied

---

(35) Part of the information summarized in this discussion (as well as that included in the housing sector discussion of the Cuban experience which follows in section IV) was provided by participants in a summer field study in Cuba sponsored jointly by the M.I.T. Department of Urban Studies and Planning and the Cuban Ministry of Construction. The Cuban rural new towns program was one of the primary subjects of the study. Particularly useful information was provided by Katrinka Ebbe, a student participant, and paticipating M.I.T. Professors Bennett Harrison and Karen Polenske. Useful background material was obtained from Dominguez (1978), Morawetz (1977), and Leontief (1971).

specific projects, therefore, the opportunity cost of the rural new communities operation must be very high. The long-run benefits are still unproven, since productivity growth in Cuba has not been very noticeable until recently. The Cubans' gamble (aside from the ideological appeal of rural communal arrangements as opposed to private farms) is that the long-run boost in the productivity and welfare of the rural population will validate the whole operation. In this, they may be correct. Two questions, however, remain: (1). Could some other investment program have made everyone (including the target rural population) better off? (2). Could the recurrent and capital cost of this operation during the critical first decade of its operation have been borne by the government in the absence of such generous Soviet subsidies?

Such questions seem relevant for BHN planning for several reasons. First, most Third World countries are much poorer than Cuba was in 1960, and the base of life expectancy and literacy on which they are building is much lower. In addition, many poor countries have no hope of attracting subsidies at the level which the Soviets have provided for Cuba. Finally, most poor countries cannot hope to match the levels of infrastructure and technical expertise with which the Cubans began their operation. Certainly, the Cuban case shows that there is some recurrent subsidy level at which it is possible to attract rural people (even those who have lived under an individualist tradition) into centers which can provide services at appropriate scale. It is also certain, however, that very poor countries cannot afford such subsidy levels (36).

In the African context, some further support for the above

assertion is provided in a rather perverse way by the recent experience of the village development program in Botswana (37). Because Botswana largely arid, villages have naturally tended to grow to substantial siz near good water sources. The political evolution of Tswana society has reflected this phenomenon, and power in the large villages has been firmly in the hands of a group of major chiefs. From the perspective o government planners this situation was fortuitous, since the rural people were already collected in groups large enough to allow for the provision of relatively high-quality social services at appropriate scale. The simultaneous appearance of these services in villages and the advent of a program for providing new tube wells at points widely scattered over the landscape, however, have led to a new situation.

At the well sites, the expected return to agriculture and herding is obviously a lot higher than formerly. At the same time, the provision of education and health services in the large villages provides conditions which are particularly advantageous for children. The result has been a rapid out-migration of men in response to higher potential earnings near new well sites, while women and children are disproportionately left in the serviced village sites. The fact that families have already coexisted in large villages for long periods of

---

(36) For evidence on some more typical (and less successful) efforts a new community development in Latin America, see Smith (1969), Nelson (1974), and Dozier (1969). One exception which may prove the rule -- a private settlement project in Northern Parana, Brazil -- is discussed i Katzman (1978). In his discussion of the Northern Parana experience, Katzman makes it clear that the calculation of service delivery levels on the basis of willingness to pay was a major reason for the long-run solvency of the project.

(37) Boston University is currently collaborating with the Republic of Botswana in a study of certain aspects of this program under USAID auspices.

time seems to make it possible for such an arrangement to persist.

## Conclusion

In this overview of the problems associated with integrated BHN efforts in the countryside, certain themes seem to dominate. It is undoubtedly worthwhile to recapitulate them briefly. At low levels of capital accumulation, there is a marked dissonance between the distribution of land holdings which is optimal for agricultural production and the distribution which is optimal for basic needs investments. To the extent that production is in the hands of adults and the most important needs-based investments are focused on children of primary school age or below, the dissonance could be overcome, of course, by breaking up families. Any non-coercive alternative involves a substantial direct recurrent cost in the form of subsidies to compensate farmers for current income loss in the face of expectations, tastes, and discount rates which differ from those of public planners.

Coercive policies will involve a large output loss with no short-run compensation to the affected peasants. In addition, such policies will exacerbate the problem of rural-urban migration and are quite likely to result in an atmosphere of cynicism or bitterness which makes further advances very difficult (38). Coercion can also be expensive, of course, since a police force of some kind has to be paid

---

(38) Even the Chinese have now been forced to admit this problem. Recent dispatches from Peking make frequent references to a growing informal sector dominated by young people who have fled from rural communes. For good current background, see Eberstadt (1979) and Morawetz (1977).

to maintain it. Frequently, the travel pass under these circumstances becomes simply another ancillary resource required by the peasant family to return to what it considers to be a superior arrangement. Like all such resources, it allows the proprietors (the police) to extract part of the resulting flow of net benefits as rent.

## IV. RECURRENT COSTS AND PROGRAM IMPLEMENTATION

The central issue in this analysis of the recurrent cost problem is the need for increased annual support levels for basic human needs projects. As always, there are two ways of attaining this goal: Either the current required level of net subsidy can be increased through improved utilization of available resources and cost reductions, or the total level of available resources can be increased. A complete analysis of the role of public finance in BHN policy implementation requires that attention be paid to both options. The first possible route to amelioration of the recurrent cost problem lies through efficiency improvements. All the evidence from project experiences suggests the need for the use of project analysis in the valuation of benefits and costs. This subject has explored in detail in Chapter II. For monitoring on-gong projects, it is very desirable to elaborate a system for observing project impacts and the evolution of project-specific net subsidy levels to particular income groups through time. We have concluded that the evaluation of benefits should be strongly guided by the preferences of the poor, as revealed by statistical inference or direct preference surveys. The evaluation of costs must depend upon opportunity values, and it will be argued that recurrent cost problems can originate in the failure to define the set of scarce resources broadly enough.

It is this problem of true opportunity cost which leads to the second major organizational issue in recurrent cost analysis, the question of institutional design. We have concluded that two major issues must be resolved in this context to assure the efficient use of

scarce recurrent support resources. Appropriate organization for basic needs delivery depends upon proper identification of the domain of organizational responsibility and appropriate institutional design. After examining the available evidence, we have concluded that the appropriate domain for organization is generally territorial rather tha sectoral, and that both the national planning environment and the observed behavior of project clients create a strong argument for small scale in fundamental resource allocation units.

Along with the basic planning and institutional strategies, cost-effective BHN program design must include a well-articulated delivery strategy. In our view, two approaches to delivery are possible. If it adopts the Pigovian approach, the government can defin "minimum decent standards," and exhaust its recurrent budget by awardin entitlements to these standards to some group of poor families. In the discussion which follows, we have chosen to continue calling this the "lottery" approach. As an alternative, the government can preserve horizontal equity by promoting maximum expansion of relevant supplies under fairly general conditions of access. Although lotteries have bee quite common in BHN-type programs, we conclude that the argument for maximal supply expansion is stronger.

The three questions of organizational design which have just beer outlined are not only important for the efficient operation of BHN projects. They are also partially interdependent with strategies for increasing the total level of recurrent resources available for the support of BHN projects. Two basic resource mobilization schemes are possible. First, the government can implement policies for redistribution, either through the central fiscal system or through th

use of cross-subsidy schemes within sectorally- or territorially-defined organizations. Secondly, the government can aid in the building of organizations which enable the poor themselves to assure higher levels of recurrent support.

In considering these two problems jointly, we have concluded that three principles are important to their solution. First, in many cases the government could redefine the legal nature of its recurrent support obligation to particular projects by assigning project clients the ownership of appropriate capital assets as a steady support source. Secondly, BHN delivery organizations should minimize the recurrent subsidy burden and administrative overheads wherever possible through one-shot recovery of the capitalized value of net subsidy flows. Finally, any implementation of cross-subsidy policies by BHN delivery organizations should be based on the classic principle -- tax where demand is inelastic and subsidize where it is elastic -- but under the revised "full price" definition of elasticity for poor families which has been developed in the preceding sections.

## A. Organizational Strategy: Planning Flexibility and Local Control

In Chapter II, we discussed what is in effect the information gathering and processing apparatus which would be appropriate for a BHN delivery system. The efficiency with which this information system can promote the accumulation and allocation of resources is not, of course, independent of the organizational structure within which it operates. The organization is simultaneously the arbitor of BHN flows and the lin with a broader developmental environment, and this dual role suggests one broad principle which underlies the question of appropriate organization in this context. Both the external and internal linkages of BHN organizations are inescapably subject to high levels of uncertainty. When generalized uncertainty is tied to a high shadow value of skilled management, two characteristics of effective delivery systems seem apparent: Their domain of responsiblity should generally be territorial rather than sectoral, and the lowest units in the decision hierarchy should be granted a substantial proportion of the total resource collection and spending powers of the system.

### (1). Domain of Responsibility

As we have noted, one substantial element of undertainty in the operation of BHN systems lies in the poorly-understood links between particular BHN inputs and welfare results. In any particular environment, however, it is possible to accumulate useful experience a to act on it effectively if both the information and resource-flow

systems are appropriately organized. In this context, the primary response to uncertainty must be an organization which is capable of learning and which retains broad allocational flexibility. Sector-specific bureaucracies are almost certain to be inappropriate in such cases. Inevitably, they become increasingly myopic and self-serving in their information use and demands for resources with the passage of time. Given the need for learning and relative flexibility, then, the appropriate domain of responsibility for BHN organizations should go across sectors.

If organizations are to have cross-sectoral decision authority, then what should be the fundamental unit of organization? An important key is provided by the inherently spatial nature of many basic needs problems. The BHN problem for the poor is frequently a problem of access to appropriate delivery sites, and the opportunity cost of co-location in both urban and rural contexts is often high. Thus, the unit in which many BHn-related values are created and distributed is territorial, and territory naturally suggests itself as a substitute for sector as the basis for BHN organization.

(2). Structure of the Decision Hierarchy

Since large variations in life circumstances within and across territorial units are inescapable in poor countries, it is clear that an effective BHN organization must have the capacity to accumulate and transfer resources both interpersonally and interregionally. The question, as always, is the distribution of power over resource flows within the decision hierarchy. In order to determine the appropriate

scale for decision-making in the BHN context, we have only to examine the relationship between a few simple principles of organization theory and the environment which is characteristic of LDC's. Our conclusions are reinforced by some general observations about willingness to be taxed and the accountability of the taxing authority.

In any system of organization which has more than one hierarchica level, the question of optimum size and complexity is inevitably tied t a rough calculus of costs and benefits (39). The benefits to large organizations may include the internalization of various externalities; the potential for processing a large amount of information efficiently; the provision of incentives for honest communication which may not exis in a system in which short-term contracts are drawn between anonymous units; and resource allocation under conditions of relatively full information.

On the other hand, large organizations suffer from certain unavoidable costs. Primary among these are the rising rate of transmission error as organization size increases and increasing response time. Even in industrial economies with highly developed communications and information-processing facilities, many organization have experienced declining marginal net benefits beyond a certain size. In poor countries, two factors combine to produce an organizational cos structure which is quite disadvantageous. Obviously, existing comunications links are frequently very poor. At the same time, the financial environment is generally subject to considerably greater uncertainty. Large fluctuations in foreign exchange earnings provide

---

(39) A full discussions of the costs and benefits associated with organization size can be found in Williamson (1975).

one example of the problem.

Both of these factors, coupled with a general lack of experience with large-scale management techniques, dictate relatively small-scale organizations as the fundamental instruments for resource transfers. These arguments take on more strength when they are joined with some propositions concerning the willingness of poor communities to be taxed. Generally, experience with development projects has suggested several reasons for supporting locally-oriented fiscal arrangements.

When taxation (either implicit or explicit) is in question, there appears to be a relationship between the local willingness to be taxed and the local visibility of results. In a society with poor communications and very unequal representation of separate community interests, the link between taxes lost to the central treasury and return payments is quite easily broken. Partly, this is undoubtedly due to the direct experience of costs without any concommitant ability to gauge the level of benefits. Another contribution to resentment of centralized taxing-spending mechanisms is made by the perception that much of the money is simply absorbed at higher levels (40).

The problem of linkage between taxing and spending is undoubtedly compounded by the high rate of corruption characteristic of many poor

---

(40) In Africa, the problem of central absorption has led to results which are paradoxical in many cases. Governments whose announced policies have been explicitly socialist have frequently created large allocative-administrative bureaucracies charged with overseeing welfare programs for the poor. Consequently, much of the money allocated to "basic needs" expenditures has been routed into welfare expenditures for members of the bureucracy itself. The public housing programs of states such as Mali and Tanzania provide good examples of this phenomenon. At the same time, states such as Upper Volta which are much less socialist in orientation have not experienced this pattern of bureaucratic growth. Less money may therefore be allocated to "basic needs" projects, but a higher portion of the benefits often reaches the truly poor.

societies. Some combination of occupational insecurity for public officials and their responsibility to central organizations rather than local communities undoubtedly contributes to this phenomenon. In any case, it can generally be said that the monetization of resource flows always creates opportunities for corruption, and that flows which are not monetized are often preferable when they are possible.

Finally, the problem of state-community transfers on capital account needs to be addressed. In the case of general taxation, the fiscal system at least has the advantage of familiarity with the principle and a more or less passive acceptance of payment at some level. In many cases, however, (and particularly for BHN projects), the phenomenon may be reversed. Rather than delivering current financial support in exchange for current taxation, the state may deliver a capital asset to a community under some arrangement for repayment. Field experience with willingness to pay under such circumstances suggests very strongly that the terms of repayment must be clearly worked out and agreed to by the community in question before the asset is delivered. Attempts to impose payment schedules after asset services have become regarded as "free" by poor clients have not met with much success (41).

In summary, then, the determinants of appropriate organizational forms seem to point to a particular set of characteristics which would be very desirable in the BHN context. Territory provides a better basi

---

(41) Harold Lubell of USAID supplied an excellent example of this phenomenon in a recent exchange with the authors. Lubell recounted the experience of the Oxfam tubewell drilling rig team in Bihar, India, who discovered that they had great difficulty in collecting payments once a well had been drilled but that they could invariably collect the full cost in advance of drilling.

than function for organization when trade-offs across welfare categories are important. At the same time, relatively small scale, major local resource responsiblity, and non-monetization wherever possible have substantial appeal. Finally, experience suggests that fiscal relations between regional/national and local organizations be based on clear agreements in advance concerning the repayment of capital allocations. The latter arrangement seems approprite because of the apparent way in which expectations-formation affects willingness to pay.

Many of the features just cited have in fact been characteristic of organizations which have undertaken at least some BHN-type activities in very different contexts. The provision of education, sanitation, health, and housing services by independent communities in the United States is certainly one example. Another is provided by the Chinese communes. In both cases, territorially-organized units have served as the basis for taxing and spending on needs-type activities. A principle role of the central authority in both cases has been that of redistribution between communities and provider of high-level technical services. In the Chinese case, alternatives to money flows (such as work points) have also been widely employed.

## B. Delivery Strategy: Horizontal Equity and Arbitrage Problems

Both of the preceding sections have been built around the premise that some kind of targeting strategy is important to BHN deliveries. Whether the emphasis is on the supply side or the demand side, the whole approach is based on the notion that strtegic intervention will yield greater returns for the poor than attempts to mobilize the fiscal system for pure income redistribution. Thus far, the emphasis has been on the design of a planning strategy and an organizational format withi which targeted deliveries can take place. Before proceeding further, however, it is important to determine the kind of targeting strategy which seems most sensible.

As previously mentioned, two basic approaches can be identified, and each stems from a particular set of assumptions about the behaviora basis for family economic activity. The neoclassical approach gives considerable weight to the revealed preferences of the poor in determining strategy and emphasizes changes in the relative (full-) prices of relevant goods and services in influencing behavior where thi seems appropriate. The logical goal of this approach is therefore a policy of generalized supply expansion at the maximal rate for BHN good and services.

To the extent that BHN demand by the poor is price- and income-elastic (which is generally the case) and conversely for the rich, such general supply-side policies will have an incidence which is highly progressive. A key component of the approach is horizontal equity. Because geographical unevenness is inevitable with limited

resources, a strong fiscal principle must be the recapturing of a substantial portion of the net benefit flow generated for poor recipients.

The evolutionary approach to family behavior suggests a very different targeting strategy. Because of its assumptions concerning the role of inertia and information gaps in determining the behavior of consumers, this view leads to the design of "appropriate standards" for the poor and the design of specific consumption targets as a means of assuring the attainment of those standards. Evolutionary strategies therefore frequently become "lotteries," in that some relatively small group of poor families is selected for the enjoyment of "appropriate standards" at a level far above what they formerly could have afforded. The exhaustion of budgetary resources inevitably entails substantial horizontal inequity, since only lottery winners receive any direct benefits from the approach.

In practice, the lottery approach to BHN allocation has been quite common. Public housing for relatively few poor families at high resource opportunity cost is frequently pointed to as an example. High-quality clinics and schools provide others. Although horizontal equity problems are obvious, a common preference for lotteries seems generally to stem from two sources. The first is a notion of decent standards on the part of key administrators who frequently look to wealthy nations for examples. Since much of the financing for BHN-type projects is arranged by representatives of these nations, the problem of local sensitivity compounds the desire to build at rich-country standards (42). A second source of the appeal of provision at "decent standards" is undeniably political. In many communities, the awarding

of elaborate entitlements to a relatively small group of recipients may be sufficient to retain the support of large (and hopeful) masses of poor families.

Although the appeal of visible and relatively glamorous projects is obvious, the lottery approach has been strongly criticized by supply-expansion advocates on two grounds. The first is essentially philosophical -- the attendant horizontal inequity is simply rejected a inappropriate by assumption. The second objection is practical. It holds that BHN lotteries will fail, even on their own terms, because of the general phenomenon of arbitrage. This objection and a partial evolutionary response to it have aready been mentioned. Where BHN deliveries are primarily in the form of goods (housing, food, medicine) they can always be resold. Where delivery comes in the form of access to superior services (particularly in urban areas), the right of access can also be arbitraged.

In either case, the practical critique of the BHN lottery strategies is that arbitraging will inevitably make them inefficient income transfer strategies. The poor will move to the same allocation which they would have enjoyed had they simply received a money supplement, and both the horizontal equity principle and BHN goals will

---

(42) In a recent World Bank housing mission to Nigeria, for example, proposals for housing construction at low cost frequently generated hostile resistance from Nigerian administrators on the grounds that suc proposals were condescending-- the mission experts, it was asserted (an accurately, of course), would not have recommended the same kind of housing in their own countries. It should be pointed out, however, that hostility to the recommendations of foreign experts is not the only source of this phenomenon. The housing literature is full of examples o slum clearance projects which have replaced dynamic but unsightly squatter settlements with housing designed at "decent" (and extremely costly) standards.

be violated in the process.

If lotteries must result in the elimination of any unique BHN component in redistributional strategies, then they are obviously not worthy of consideration as BHN programs. Since so much follows from the adoption of a basic delivery strategy, the point is a crucial one. Unfortunately, very little detailed evidence can be brought to bear on this question. Some level of arbitrage is probably inevitable, but it is not clear whether this is a major problem (43). Neoclassical critics generally maintain that massive arbitrage is inevitable unless a large surveillance force is employed to ensure that poor clients retain what they have been allotted. If this were the case, any benefits to a particular BHN allocation wuld be heavily offset by the costs of maintaining it. An evolutionist response might well be that the assumption of utility-certainty on the part of poor families is fallacious, and that the costs of enforcing a particular consumption pattern would therefore be quite small.

---

(43) Anecdotal evidence abounds, of course, and its message is generally that arbitrage is common. For housing examples, see Harris (1969) and Wheaton (1979). Nutrition examples can be found in Austin (1976). Again, M. Cohen's experience with land speculation in anticipation of World Bank squatter-upgrading investments is obviously germane.

C. Recurrent Financing and Local Resource Mobilization

Having considered the general strategy of BHN delivery from the perspective of recurrent cost savings, we are now ready to examine appropriate forms for programs at the local level. Our approach to thi problem has been generated in response to an apparent contradiction in BHN budgeting procedures. During the past two decades, the general approach to recurrent BHN support by national governments has been part of the annual budgeting process. Support is generally by specific ministries (e.g. health, education, housing, public works), and annual allocations are simply parceled out from total ministry budgets. Unfortunately, social expenditures are tied to growth in national incom at constant revenue-effort levels, while desired growth in the nutrition, health, and education levels of the population is frequently higher than the growth rate of income.

International donors have frequently proven willing to underwrite the capital costs of BHN projects, but have shown no concommitant willingness to assure recurrent costs. There is a certain mathematical inevitability to the result. With BHN recurrent expenditures tied to the general growth rate of national income and BHN capital expenditures heavily underwritten by foreign donors, the number of personnel and quantities of materials demanding support from every unit of ministeria recurrent budget is bound to grow. Three possibilities follow: 1). Cheaper personnel and materials can be employed; 2). Some BHN projects can be abandoned as underpaid personnel begin drifting away; or 3). There can be some shift in the proportion of recurrent support provided by recipient communities, either by contract or through the

intermediation of corruption (44).

Neither abandonment nor corruption is very desirable, and our research suggests that greater central revenue effort does not provide a very hopeful prospect. Thus, the two most promising options seem to be cheaper projects and some easing of the recurrent burden through better resource mobilization by recipient groups. We will consider these options in reverse order.

The key to ourapproach to resource mobilization lies in our interpretation of the basic economics of consumer behavior. Whether urban or rural areas are being considered, a central element of the analysis is the use of time by poor families. When BHN projects are established, the primary element in the resulting flow of net benefits to the poor is time -- time saved in traveling or waiting, or time saved for health-related reasons. In urban areas, much of the value of these time savings will be capitalized in ground rents. In largely non-monetized rural economic settings, the time saved will be reallocated among other activities -- own-production, household work, education, and leisure -- as preferences dictate.

Thus, the first key element in thinking about enhancing support

---

(44) In many poor countries, corruption seems to be a common and unfortunate response to this problem. In the Republic of Zaire, for example, recurrent cost problems are particularly acute in the provision of education and medical services outside Kinshasa and the mining areas of Shaba Province. In one "free" provincial hospital known to the authors, helpless hospital patients must provide fees for service from doctors and nurses on a very regular basis. This corruption is attributed by some to a general lack of professional conscience. The fact that salary payments to these personnel are wildly erratic and steadily eroding in real terms may also have a good deal to do with the problem. The same salary loss is observable in rural Zairian schools, and many rural teachers are known to demand bribes from their students in return for favorable consideration in the classroom.

for BHN projects is the element of resource translation. Unquestionably, BHN projects yield net benefits for the poor which they themselves acknowledge. In principle project clients would be willing to part with a very substantial portion of the realized net benefit flo as a means for assuring continued support. The degree to which poor families can actually provide direct support, however, depends on local circumstances.

Generally, it can be said that the provision of some degree of direct support in urban areas isfeasible, while the form taken by the flow in rural areas is inappropriate. In both cases the primary savings to the poor are in time. The scarcity of space and the importance of access to BHN delivery sites in urban areas combine to produce land markets which largely capitalize the value of these time savings. In principle, at least, this capitalized value can be recovered through the use of appropriate policy instruments. Rural village economies, on the other hand, are non-monetized to a much greater extent. Time savings for the poor are therefore not freely exchangeable for the labor and materials expenses which comprise the bulk of BHN recurrent costs. From the perspective of poor communities, many of these expenses are effectively imports. Thus, the primary problem for rural areas is the conversion of a net benefit flow in loca exchange units (labor time) into a monetized flow which can be used for imports.

## Urban Resource Mobilization

In an urban area, any public project whose output is valued by

consumers will generate a shift in the bid-rent surface which reflects the value of access to this output or the value of the enhanced access to other sites which it provides. Roads and public transit lines are most commonly thought of in this context, but the same reasoning applies to water lines, sewerage lines, schools, and clinics. Because it is impossible to build projects which discriminate by income class in fluid urban markets, bid-rent shifts will reflect the sum of valuations by all groups in the urban market. Nevertheless, poor families represent a very high proportion of the population in poor cities, and the value which they place on the outputs of particular projects will have an important impact on the resulting bid-rent shifts.

At first glance, the whole notion that poor families could have such an impact on urban land values seems implausible. The people in question are, after all, extremely poor, and it might be claimed that they will simply have no disposable income above the subsistence margin to pay additional rents. This sort of reasoning is fundamentally incorrect because it neglects two kinds of adjustments which the poor are generally capable of making. In the first place, part of the time savings generated by a BHN project will be used for additional labor and converted to income by poor families. Secondly, the poor have another means of payment at their disposal. When no additional time and money are available, they can pay with density.

If a particular urban land parcel shifts to twice its former density of occupation by identical poor families who spend the same proportion of their money incomes on rent, then the market value of the parcel will obviously double. Such a fixed-ratio outcome is implausible because density is a source of disutility, but there can be

no doubt that higher densities in areas accessible to new BHN sites will lead to substantial increases in land values.

For BHN projects which are operating under potentially heavy recurrent cost burdens, this shift in bid-rents seems to offer a promising surce of additional support. Lest this prescription appear unnecessarily cruel to financially-strapped poor families, we should reiterate: As long as urban land is largely in private hands and access is valuable to poor families, they are going to pay anyway. The only relevant question is: Who will receive the payments? Here it is necessary to point again to the fundamental difference between lottery and supply expansion strategies in providing BHN goods and services. Neither strategy would, of course, advocate leaving these payments in the hands of rich landlords. To the degree that poor families have entitlements to fixed areas or unique access to BHN deliveries, however, the lottery approach would leave the resulting payment flow (explicit or implicit) in their hands. In so doing, it would add another relatively small group to the urban middle class.

The supply expansion approach, on the other hand, would attempt to recapture the payment flow as a source of recurrent support. Three basic options are available in this regard. At the newly-installed site, BHn deliveries can simply be provided for unit fees. Alternatively, the increase in local bid-rents can be recaptured through a regular levy on the property whose value has been increased by the project. Finally, an attempt can be made to recapture the whole capitalized value of the net benefit stream from local landlords through the imposition of a one-time valorization charge.

Each of these options may have merit in particular circumstances.

Some obvious disadvantages attach to the first two, however. If the poor can pay mostly with time and density, then the levying of unit fees for BHW deliveries is a form of pricing which discriminates heavily in favor of richer families (It is true, of course, that some arrangement for payment in time through public labor might be arranged, and this option will be discussed in the concluding part of this chapter). The collection of periodic levies on adjacent property is, of course, the American system (generally organized at the town level through property taxes), but its applicability is doubtful in cities where administrative systems are very poorly developed. Decentralized tax collection of any kind is notoriously difficult under such circumstances.

This brings us to the third option, which has considerable appeal. When administrative capacity is very limited, the convenience of valorization charges is undeniable. If the public authority can at one swoop recapture the capitalized value of the expected net benefit flow from a BHW project, this lump sum becomes available as a source of recurrent support. In order to assure an uninterrupted flow of payments to the project in question, the valorization collection should undoubtedly be invested in some interest-bearing debt certificate (public or private), so that a multitude of payments from landowners is effectively replaced by one regular payment from an institutional borrower.

It may justifiably be objected, of course, that this suggestion flies in the face of imperfections in Third World credit markets. Few landholders will be able to supply a large valorization payment from their own pockets, and in the absence of well-developed capital markets they may be forced into a position of de facto expropriation. If the

landlords are relatively wealthy, the fact of expropriation may not trouble public authorities, but the result certainly would. The state as landlord would probably be in a worse position than the state as tax collector, because it would be necessary to collect rents more frequently and from more than one family on each parcel of land. In addition, many "landlords" are poor families who rent part of their meager space to even poorer people. It is certainly not in the interest of anyone to expropriate such people.

We are therefore left with a paradox. If credit markets are poorly developed and de facto expropriation is undesirable, the only remaining option is for the state to grant loans to landowners. It may seem justifiable to wonder whether the state as creditor is really in a better position than the state as tax collector or landlord. The substitution of one kind of valorization payment for another may seem more like a conceptual sleight of hand than an effective policy. Whether this is true or not will depend largely on the legal system which is in place.

In many cases, the foreclosure laws are more highly developed and readily enforceable than the tax delinquency laws. If this is the case it will be administratively convenient to adopt a system of legal obligations for loan repayment rather than a system of periodic tax collection. An additional appeal of the former is that the state can simply refuse to install projects in areas where landholders are not willing to make the loan agreement in advance.

Obviously, valorization charges do not represent any kind of panacea. Besides the nearly-inevitable problem of credit adminstration, public authorities would be saddled with the necessity for determining

appropriate charges for particular BHN delivery sites. The latter problem seems particularly difficult, but two defenses may be offered. First, it may be possible in many cases to let market processes themselves determine appropriate valuations. Public authorities could simply announce their intention to construct and maintain delivery systems particularly suited to the needs of the poor, and then allow a competitive bidding process to determine the location.

Even if this expedient is not available, the problem of value estimation is not insurmountable. A system of valorization charges has been in place in at least one Third World city -- Bogota, Colombia -- for quite some time now, and extensive documentation by the World Bank suggests that it works reasonably well. In Bogota, valorization charges have been primarily used in association with water line installations. A staff of trained evaluators is used to estimate the incremental rent contour surrounding a new project, and landholders are charged proportionately. Capital market problems have necessitated the extension of loans to many landholders, but no great resistance to the program has been encountered and repayments have apparently been adequate (45).

Thus, a system of valorization charges seems feasible, and it might prove to be a valuable part of BHN policy in poor urban areas. In some cases, of course, technical and administrative problems might simply be too burdensome, so that user charges would have to be

---

(45) One reason for this seeming passivity is undoubtedly the fact that charges in Bogota are only intended to cover the present value of project capital and operating costs. Any remaining differential is captured by landholders as rent, and it is not unreasonable to suppose that this differential is large in many cases.

utilized. If so, it would seem appropriate to levy these charges in a form which the poor find it possible to pay. Where BHN projects are territorially organized, it may be possible for families to earn payment credits by offering construction labor as a way of diminishing BHN expenses. If central allocations for BHN projects are primarily on capital account, it should be possible for local planners to work out a internal money exchange of capital support for recurrent support, with the money savings transferred to participating families as access credits (46).

Rural Resource Mobilization

Generally, the problem for rural communities is the lack of certain essential capital and managerial resources which would complement the resources currently at their disposal. At the same time, central governments have an obvius tendency to expand their capital budgets more quickly than their recurrent budgets. Thus, an expedient suggests itself: Wherever possible, it would seem appropriat to complement rural BHN projects with an ancillary capital resource which can be combined with local labor to produce revenues sufficient t cover a large proportion of recurrent labor and materials costs. Several variations on this same theme can be imagined, and the capital in question does not have to be a direct joint input to local productio

---

(46) Various ad hoc versions of this procedure have been employed by public agencies. In many Brazilian states, for example, the water authorities give preference to poor communities which volunteer labor time sufficient for system installation. In compensation, water services are subsequently provided at subsidized rates.

as long as it allows substantial labor resources to be freed up.

Consider, for example, the case of well provision in rural areas. The value of such a project to a particular village is primarily the opportunity cost of the time (generally that of village women) currently allocated to water-hauling and the additional net benefits (improved sanitation, health, nutrition) associated with the increased use of water at a much lower full price.

Once a well is installed, large quantities of female labor time are suddenly freed up. Under appropriate conditions, the result can be a sizable productivity increment captured by the village. The actual size of the increment would depend, of course, on local physical conditions and the availability of additional inputs to help boost agricultural or handicraft production. In any case, the village should exhibit substantial willingness to pay for well installation, and the existence of this willingness could be used to promote the simultaneous installation of basic welfare services. As always expectations would be important, and arrangements which would be readily agreed to in advance could be resented as burdensome if payment were claimed after the net benefits had begun flowing into private pockets. Thus, the best approach might be to allocate resources for water projects only to those villages which were willing to accept a package which included some technical assistance to the women concerned and a guarantee that some proportion of the resulting earned income (private or communal) would be earmarked for the support of recurrent costs associated with a service activity such as education or health.

The problem with all such proposals, of course, is that "technical assistance" in production and the assurance of marketing are easier to

describe than to provide. Rural areas are frequently characterized by chaotic transport, the lack of storage facilities, and the absence of the relevant technical expertise. Putting together the right kind of "package" can therefore be seen as a very tricky problem.

Pointing to difficulties in the assembly of appropriate production-support packages does not imply that the idea should be abandoned. Chaotic fiscal conditions have begun spontaneously generating exactly this kind of approach in many poor countries. The authors are familiar with the case of a rural Nigerian university whose current chancellor (an economist) has moved the university away from dependence on an erratic public fisc by attracting ancillary capital resources. The university currently operates a hydroelectric project, selling substantial excess supplies to a local city. At the same time, it owns a large farm which is simultaneously a site for agricultural experiments and a substantial source of revenue for the campus.

At the opposite extreme, we might cite the case of small-scale private aid organizations such as the Planned Parenthood Federation of America (PPFA), which has undertaken family planning projects in South Asia. In two villages in Bangladesh, PPFA money is being used to develop fish ponds whose revenues will be used to provide recurrent support for continuing family planning activities in the area. In other South Asian villages, PPFA resources have been used to develop a rice-husking mill and a mustard-oil press, with the same end in mind (47). At small scale, then, the capital resources which are sufficient to support BHN projects are not necessarily too complex for management

---

(47) See Mouat (1979).

by local personnel.

In summary, we might term this the "assets" approach to recurrent support as opposed to the "income" approach. What is really involved is simply a redefinition of the legal obligation of the central treasury to provide recurrent subsidies for local projects. When physical capital is placed in the hands of local personnel, an income stream is established which will be effectively impervious to fiscal fluctuations at the center. At the same time, it must be recognized that such drawing down of the capital budget is necessarily taking place at a high shadow cost. Environmental problems may force this approach, but it seems highly doubtful that the optimal distribution of physical investments will bear a close resemblance to the distribution of BHN projects, at least from the perspective of pure income earning. In addition, local managerial inadequacies are likely to erode the expected returns to such decentralized investment even further.

Thus, a further expedient may be called for in come cases. From the perspective of recurrent cost support for a local BHN project, the important thing is a claim on a steady income stream, not the possession of a physical asset. A natural alternative is therefore the use of interest-bearing general-obligation government bonds. The general obligation nature of such bonds would assure a payment flow (unless, of course, the government chose to move to a general state of default). At the same time, the use of legal claims on the income streams from physical assets rather than the assets themselves would allow scarce capital to be allocated to sites where it would yield the highest rate of return.

Both these approaches to strategic asset provision can serve the

function of recurrent cost support, although their implications for employment may differ substantially. There is probably nothing categorical that can be said about their relaive preferability. In many cases, the long-run benefits to local employment creation and skill formation may warrant the parceling out of ancillary physical capital. In other contexts, high rewards to the use of scarce capital at particular sites may justify the establishment of financial portfolios for the support of BHN projects.

Since the focus of the discussion here is on an altered form of capital budgeting, an appropriate role for international donor agencies seems relatively easy to define. Much of the physical capital could be imported, so that the traditional role of aid agencies as capital budget supporters could be maintained. One major change might have to be in the scale of projects which were considered eligible for the allocation of aid funds. The sort of assets approach under discussion here involves the provision of very small allotments of capital to widely-scattered communities. In the past, high overhead costs have frequently discouraged international donors from involvement in such small efforts. Where organizations such as Oxfam have already ventured, however, large-scale aid organizations might consider following.

We have one further thought on this subject which is not offered wholly with tongue in cheek. If the argument about scarcity value and management difficulties applies to locally-provided capital, it also applies to the international equivalent. If local management is inadequate, then a default will shut off the major source of recurrent support and strongly diminish the effectiveness of any BHN project. Thus, international donors might want to consider the "portfolio"

approach in some cases, as well. At first glance, the notion of a Bangladeshi village in possession of a few USAID-purchased AT&T bonds seems a little bizarre, but beauty is in the eye of the beholder.

Cost Minimization

While our research has underlined the importance of net benefit recapture as a source of recurrent support, it is important to consider the problem of cost minimization, as well. In the evaluation of alternative BHN facilities, the appropriate criterion, as always, is the present opportunity value of the resources used up in establishing and running the project. The ability of unassisted poor communities to earn money is often quite limited, and the relatively low shadow values of local labor and locally-available material inputs should pull the choice of technique for BHN delivery toward the facilities whose construction and operating characteristics are intensive in local resources.

Thus, cost minimization and the resource translation problem combine to enhance the appeal of BHN projects whose recurrent (and capital) support can be largely assured by local resources. Two implications follow from this. First, materials used for BHN projects must be locally-producible to the maximum extent possible. Secondly, service delivery personnel (teachers, health workers, water technicians, agronomists) must be supportable at levels which are not too far above those of local poor inhabitants. Given a small income differential and basically-similar tastes, much of the income of BHN project personnel can then be supported by locl residents through donations of labor (e.g. construction of residences) or products (food, clothing, household

implements, etc.).

This sort of reasoning is far from new, of course. It has been particularly highly developed in China, where support of "barefoot doctors" and simply-trained primary educators by urban and rural communes has facilitated a rapid expansion in the provision of basic health and education services to great numbers of people. Barefoot doctor programs particularly are attracting increasing attention as an appropriate approach to medical service provision for the poor. In some cases, rural communities have proven capable of offering adequate support to simply-trained primary educators (48).

From the perspective of cost, the conclusion seems unambiguous: The best BHN projects are those which can be constructed and maintained using local materials almost exclusively, with support personnel trained to levels such that their expected incomes (and associated demands for goods and services) are only slightly higher than those of their clients. Unfortunately, cost considerations are not sufficient for choosing projects wisely in this context any more than they generally are. Expected benefits count as well, and it is here that opinions begin differing sharply.

---

(48) S.I. Kiggundu, presently a Professor of Economics at Makerere University in Kampala, Uganda, has first-hand knowledge of the circumstances under which rural education was provided during the fiscally-catastrophic period at the end of the Amin regime. Local teacher incomes were often simply terminated during this period, but rural communities frequently mustered resources sufficient to persuade teachers to continue their work. Three reasons for this successful shouldering of the recurrent support burden by local communities can be cited. First, the perceived opportunity cost of rural educational employment undoubtedly declined during the latter Amin period. Secondly, relatively wealthy rural residents proved willing to accept a disproportionate financial burden. Finally, poor community members assured teachers a continuing flow of essential foods and some additional services.

Insofar as some items are concerned, of course, there is no controversy. Blackboards are generally necessary for teaching, drugs for healing, and appropriate materials for tube wells. Some possibilities for local input-substitution may exist in this regard, but it is obvious that at least some of the materials must be imported for almost any project. Such materials will not, however, be the main component of cost in many cases. Rather, personnel will carry the most financial weight. With respect to the provision of medical and educational services particularly, the controversy rages over the relationship between personnel costs and associated benefits. Unless medical and educational personnel are simply conscripted for service to the poor (in which case the quality of the services which they provide is apt to be extremely poor), the relevant notion is their own perception of opportunity cost. This, in turn, will vary with the degree of training which they have had.

Since client benefits obviously vary to some degree with service personnel training, the connection is complete. For BHN planners, the real question must be the relative shape of benefit curves for particular kinds of BHN service provision as training times increase. To advocates of "barefoot doctor-teacher" arrangements, the expected benefit curve is logistic, with the second inflection point at fairly low training levels. Critics of this approach may agree that the curve is logistic, but they would argue that the second inflection point and the maximum net benefit point are reached at much higher training levels. This question will be considered in substantially more detail in the next section. For present purposes, the main issue centers on the problem of necessary money support. The higher the training level

# APPENDIX A

## Maximum Likelihood Estimation of Modal Choice Models

of BHN personnel, the higher will be their perceived opportunity cost and the higher, therefore, their demand for money as a supplement to locally-produced goods and services.

## I. Alternative Approaches to Estimation

At present, two approaches to the mechanics of estimation are commonly employed in econometrics. These two estimation techniques might be called "classical" and "likelihood", in deference to the opposing philosophies which have produced them. The classical family is better known than its competitor. It includes all of the familiar least squares estimates, whose optimization criterion is the minimization of sums of squared differences between the actual and "fitted" values of dependent variables in econometric equations. The mechanics of least squares reflect the classical philosophy, in which any equation parameter has a fixed value whose location on the number line can be guessed more precisely as the sample size increases. Least squares estimation guarantees an asymptotically normal distribution of repeated sample estimates, centered around the true parameter value. Much of classical statistics uses the normality of sample distributions as the basis for inference.

While most economists are still (more or less) comfortably at home with the classical view, likelihoodism has arisen to challenge it. In the likelihoodist philosophy, the sample itself has a particular significance. Among all samples of identical size which could have been drawn, it is considered to have been "most likely" given the stochastic process generating the values observed in the underlying population. Philosophically, this approach amounts to a rejection of the classicist's "fixed" sampling distribution about a true parameter value. The likelihoodist regards the best guess about the parameter governing some stochastic process to be the value which makes the probability of the sample drawn greater than that of any other sample of equivalent size.

Inevitably, this approach is more probabilistic than its classical counterpart. In order to make any progress at all, the likelihoodist has to have a model of the stochastic process which is generating the observed values of a particular random variable. Unless this is done, no parameters can be estimated. The classicist, on the other hand, is under no such restriction. His approach to fitting yields parameter estimates quite independently of any underlying stochastic models.

To get a feel for the difference between these approaches, it may be useful to look at a simple estimation problem from both perspectives. Suppose that we have a sample (size N) of draws on a random variable, and we want to estimate the mean of the underlying population. In the classical mode, we could set up the problem as follows: Let $Y_i = \mu + \varepsilon_i$, where $Y_i$ is the ith random draw, $\mu$ is the true mean value of $Y_i$ and $\varepsilon_i$ is some random error (distribution unspecified) which causes $Y_i$ to depart from $\mu$ in each sample observation. To the classicist, a logical approach to estimation in this context is ordinary least squares. Given the sample of size N, an appropriate estimator of $\mu$ is the value $\hat{\mu}$ which minimizes the sum of squared residuals: $(\sum_i e_i^2 = \sum_i (Y_i - \hat{\mu})^2)$.

The general specification of the least squares estimator for the model

$$\underset{(N\times 1)}{Y} = \underset{(N\times K)}{X}\ \underset{(K\times 1)}{\beta} + \underset{(N\times 1)}{\varepsilon}$$

is $\hat{\beta} = (X'X)^{-1}X'Y$.

In our case, $\underset{(N\times 1)}{Y} = \underset{(N\times 1)}{i}\ \underset{(1\times 1)}{\mu} + \underset{(N\times 1)}{\varepsilon}$ (where i is an N-row vector of ones),

and the "degenerate" least squares estimator is:

$$(1) \quad \hat{\mu} = (i'i)^{-1}i'Y = (N)^{-1}\sum_i Y_i = \frac{\sum_i Y_i}{N} = \bar{Y}$$

Thus, the classical approach leads to the conventional estimate of the population mean.

In considering this estimation problem, the likelihoodist has to take a very different approach. At the outset, he faces a problem which does not immediately confront the classicist: He has to impose a model of the error-generating process. Suppose that he imposes the assumption of normality -- that is, $\varepsilon_i \sim N(0,\sigma^2)$ -- along with the condition that the draws represented in the sample be truly independent. Under these assumptions, the probability that a particular draw will take on the value $Y_i$ is given by the probability density function for a normally-distributed variable:

$$(2) \quad P(Y_i) = \frac{1}{\sqrt{2\Pi\sigma^2}} e^{\frac{-(Y_i-\mu)^2}{2\sigma^2}}$$

In this case, N draws have been taken from the population. The joint probability that two independent sample draws will take on values X and Y is given by $P(X,Y) = P(X)P(Y)$. Similarly, the joint probability for N independent draws is given by:

$$P(Y_1,\cdots,Y_N) = P(Y_1) \cdots P(Y_N) = \frac{1}{\sqrt{2\Pi\sigma^2}} e^{\frac{-(Y_1-\mu)^2}{2\sigma^2}} \cdots e^{\frac{-(Y_N-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{\sqrt{2\Pi\sigma^2}} e^{\frac{-\Sigma_i(Y_i-\mu)^2}{2\sigma^2}}$$

At this point, the sample and the stochastic model have been fitted together. There are only two unknowns in the problem, $\hat{\mu}$ and $\sigma^2$, and the likelihoodist approaches estimation in the familiar way. Since the sample $\Sigma(Y_1,Y_2,\cdots,Y_N)$ was actually drawn, the underlying true parameters $\mu$ and $\sigma^2$ are taken to be those which make the sample more likely than any other. The joint probability can

be relabeled the "likelihood" of the N-observation sample, and a simple problem in calculus results: Choose $\hat{\mu}$ and $\hat{\sigma}^2$ to maximize the value of the likelihood function:

$$(3) \quad L = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{\sum_i (Y_i-\mu)^2}{2\sigma^2}}$$

This is now a practical problem and there is a very important practical aid to solving it. Since there is a one-to-one correspondence between positive numbers and their logarithms, the estimates of $\hat{\mu}$ and $\hat{\sigma}^2$ which maximize the logarithm of L will also maximize L. The logarithic form of (3) is much more tractable mathematically.

$$(4) \quad \log L = \log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2} \sum_i (Y_i-\mu)^2$$

The appropriate values for $\hat{\sigma}$ and $\hat{\mu}$ can be read from the first-order condition for maximization. At present, only $\hat{\mu}$ is of interest.

$$\frac{d(\log L)}{d\mu} = 2 \left( \frac{1}{2\sigma^2} \right) \Sigma(Y_i-\hat{\mu}) = 0 = \sum_i (Y_i-\hat{\mu}) = 0 = \sum_i Y_i - N\hat{\mu} = 0$$

$$(5) \quad \hat{\mu} = \frac{\sum_i Y_i}{N} = \bar{Y}$$

Obviously, the classical and likelihoodist approaches do not necessarily lead to conflicting results. Maximum likelihood parameter estimates have all the usual attractive asymptotic properties (consistency, unbiasedness) in large samples. This is also true for least squares parameter estimates, of course. It is therefore natural to wonder why anyone would choose the relative

computational complexity of the maximum likelihood approach, along with the necessity for imposing a particular stochastic model. The answer is that in many applied cases least squares estimation is simply impractical, while maximum likelihood methods can be applied. The modeling of discrete choices provides an excellent case in point.

## II. Modeling Discrete Decision Problems

Microeconomic theory has focused on decision problems in which easily-divisible goods and services are consumed, produced, or marketed. In many interesting cases, however, quantities are indivisible and discrete choices must be made. Because discrete decision problems have frequently been encountered in public policy analysis, econometricians have recently devoted considerable effort to the specification and estimation of appropriate choice models.

These models follow their continuous counterparts in adopting the convenient mythology of the "representative individual". Insofar as the observable qualities of particular goods and services bear on a discrete decision, it is supposed that some average valuation of these qualities is common to all "similar" individuals in a population. "Similarity" is defined with reference to observable characteristics of the individuals themselves. At the same time, it is recognized that unobservable characteristics of individuals and commodities can have significant effects on valuation and choice. In discrete decision models it is these unobservable characteristics which can cause otherwise-identical people in seemingly-identical situations to choose different alternatives. Since both observable and unobservable characteristics are important to the modeling approach, it is advisable to develop this set of notions more formally.

For the representative individual i faced with j alternative consumption modes, the utility function can be defined as:

(6) $U_{ij}(X_{ij},S_i) + \varepsilon_{ij}(X_{ij},S_i)$

Here $U_{ij}(X_{ij},S_i)$ is the determinate component of utility. $X_{ij}$ is a vector of the levels of relevant and observable attributes of mode j from the perspective of individual i, while $S_i$ is a vector of the relevant and observable characteristics of individual i. Some specific model of utility has to be imposed, of course, but the associated utility of identical observable modal characteristics for identical individuals is taken to be exactly equivalent. The decision uncertainty comes from the stochastic incidence of unobservables, both in modal attributes and individual characteristics.

Given this way of looking at the world, the basic decision model for econometric analysis can be derived from the nearly-tautological statement of a fundamental inequality. Suppose that for mode 1, the associated utility for a particular individual i is:

(7) $U_{i1} = U_{i1}(X_{i1},S_i) + \varepsilon_{i1}(X_{i1},S_i)$

Suppose that for mode 2, the associated utility for the same individual is:

(8) $U_{i2} = U_{i2}(X_{i2},S_i) + \varepsilon_{i2}(X_{i2},S_i)$

Under the postulate of rationality, we know that the individual will choose mode 1 over mode 2 if $U_{i1} > U_{i2}$. By substitution, an equivalent expression is:

(9) $U_{i1}(X_{i1},S_i) + \varepsilon_{i1}(X_{i1},S_i) > U_{i2}(X_{i2},S_i) + \varepsilon_{i2}(X_{i2},S_i)$

By rearrangement, we have:

(10) $$U_{i1}(X_{i1},S_i) - U_{i2}(X_{i2},S_i) > \epsilon_{i2}(X_{i2},S_i) - \epsilon_{i1}(X_{i1},S_i)$$

The common sense underlying this re-expression is easy to interpret: The decision to choose mode 1 over mode 2 will follow from the observable utility superiority of mode 1, as long as the margin of superiority is greater than any countervailing effect of utility differences based on unobservable phenomena. This is, of course, an unexceptionable statement, but (10) has an important modeling advantage over (9): Since both random error terms are on the right-hand side of the inequality, it is in fact a stochastic specification which can be used for econometric modeling.

Two obvious problems must be resolved before any econometric model can actually be fitted. Some utility function must be imposed, and some plausible assumptions made about the structure of the errors which lead to deviations in the observed pattern of choice among observationally identical individuals.

Utility function specification is the easier of the two problems by a considerable margin. Although the econometric techniques which will subsequently be described depend on linearized models of utility, it is possible to specify such models in ways which are quite nonrestrictive. At the most general level, an appropriate utility function can be specified as:

(11) $$U_{ij} = \beta_0 + \beta_1 Z_{1i}(X_{ij},S_i) + \cdots + \beta_T Z_{Ti}(X_{ij},S_i)$$

This function is linear in its parameters ($\beta_0$, $\beta_1$, $\cdots$,$\beta_T$) but its arguments ($Z_1$, $\cdots$, $Z_T$) are themselves a series of transformations of modal and individual characteristics. Under such a specification, the Z-variables can be logarithms, reciprocals, multiples, exponentials, etc. Thus, a utility function of type (11) could be as simple as a purely linear (and quite unrealistic) specification:

165

$$(12)\quad U_{ij} = \beta_0 + \beta_1 X_{ij} + \cdots + \beta_G X_{Gj} + \beta_{G+1} S_{1j} + \cdots + \beta_T S_{Hj}$$

On the other hand, the same general specification encompasses the translog function:

$$(13)\quad \ln U_{ij} = \beta_0 + \sum_{j=1}^{T} \delta_j \ln W_{ij} + \sum_{k=1}^{T} \sum_{\ell=1}^{T} \beta_{k\ell} \ln W_{ik} \ln W_{i\ell}$$

where $W_{ij} = \{X_{ij}, S_i\}$.

The translog function is additive in the first powers, squares and cross-products of the logarithms of modal and personal characteristics. It is therefore a second-order approximation to any mathematically-defined utility function for any definable group in the population. This specification is undoubtedly general enough for most applications, so that the linearization of the utility function in a set of Z -transformations does not appear to be very restrictive. For convenience, we can adopt the following notation in the stochastic model:

$$(14)\quad U_i = Z_i \beta + \varepsilon_i \qquad \text{where } Z_i = [1\ Z_{1i} \cdots Z_{Ti}]\ \beta' = [\beta_0 \cdots \beta_T]$$

This imposition of a general utility function specification leads immediately to the problem of stochastic specification. A strong argument for normality in the error-generating process can be made in this case. In the choice model which has been proposed, the only reason for different discrete choices by observationally-identical individuals in the same choice environment must be the effect of unobservable modal attributes or personal attributes which do differ across individuals. In any real situation, of course, there will be a huge number of such characteristics. Thus, the error term in the utility function is the sum of the contribution to utility made by a very large number of characteristics which are distributed randomly across individuals.

As long as a fair number of these characteristics are independently distributed it makes no difference if each distribution of random occurrences has no regular pattern. The sum of the effects of the occurrences will approximate a normal distribution, and the Central Limit Theorem guarantees that the distribution will be closer to perfectly normal, the larger the number of independently-distributed unobservable characteristics.

Thus, the assumption that the error-generating process is normal seems reasonable for discrete choice modeling. The imposition of the assumption also has another advantage for the decision model which is being considered. If two random variables are normally distributed, then so is their sum or difference in repeated draws. Recall the fundamental inequality which governs the choice between two modes:

$$U_{i1}(W_{i1}) - U_{i2}(W_{i2}) > \varepsilon_{i2} - \varepsilon_{i1}$$

Now if $\varepsilon_{i1}$ and $\varepsilon_{i2}$ are normally-distributed, so must be their difference. After substitution from the linearized utility model (14), this inequality can be re-written as:

$$Z_{i1}\beta - Z_{i2}\beta > \eta_i ( = \varepsilon_{i2} - \varepsilon_{i1})$$

or

$$(15) \quad (Z_{i1} - Z_{i2})\ \beta > \eta_i$$

The recast decision model says that mode 1 will be chosen over mode 2 if their utility difference is greater than the value taken on by the normally-distributed variable $\eta_i$. Obviously, the greater the weighted sum of the distances between values, the greater the probability that it will be larger than the value taken on by the associated random draw. The normality of $\eta_i$ guarantees that this probability will decline very rapidly for possible values as they get farther away from the mean of the distribution. Nevertheless,

any value is possible, and the choice of mode 1 over mode 2 (or vice-versa) will therefore never be certain. The regularity of $\eta_1$ does allow for definite statements about probabilities. In fact, once the normality of $\eta_i$ has been assumed, the description of the choice problem in (15) above is simply a description of the cumulative normal probability function.

The probability that the choice of mode will dictate 1 rather than 2 is the same as:

$$\text{Prob}[(Z_{i1} - Z_{i2})\beta > \eta_i]$$

Since the distribution of $\eta_i$ is normal:

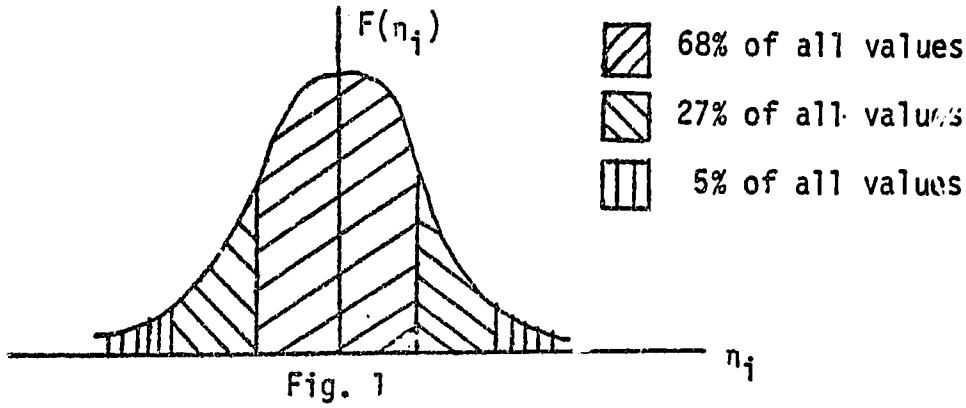



Fig. 1

This distribution can be re-expressed as a cumulative probability function which specifies the probability that $\eta_i$ is less than any particular value.

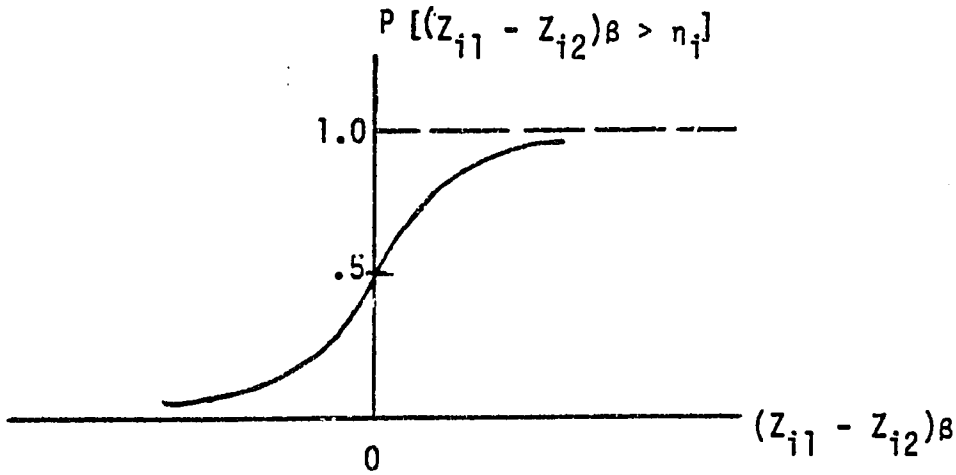



Fig. 2

Thus, the normal distribution of the error term in the original model of utility leads to a nice, regular function which relates differences in the modal attribute valuations to the probability that mode 1 will be chosen over mode 2. Unfortunately, the practical utility of the normal distribution of the error terms ends at this point. The cumulative normal probability function can only be expressed as a mathematical integral, and all the subsequent operations which lead to econometric estimation of the β's are vastly complicated by this fact. Thus, the original ease associated with normal error distributions evaporates when applications are considered.

The actual derivation of appropriate estimators of the β-values from a likelihood function based on repeated draws from cumulative normal distributions requires both rather complicated operations on mathematical integrals and the use of sophisticated computer algorithms for the numerical approximation of integral values. As this discussion progresses, it will become apparent that the use of the normal probability model does in fact have so many conceptual advantages that its application is indicated even in the face of great complexities. However, a continuation of the reasoning using mathematical integrals would certainly prove more confusing than enlightening at this stage.

We will therefore take another branch along the path of progressive simplification in order to preserve comprehensibility. Another characterization of the error-generating process will be used to illustrate the procedure for deriving maximum likelihood estimators for the β-values in this kind of problem. Although this revised view of the error-generating process will be an admitted source of some bias, it has such attractive practical qualities that they provide at least partial compensation for the loss.

## III. The Logit Model[1]

Our candidate for the revised view of the error-generating process is the Weibull distribution. We would have to reject any distribution whose shape was too far from normal on common-sense grounds, and the Weibull distribution satisfies this criterion. Its shape is fairly close to that of the normal distribution. At the same time, it is generated by a function which is much less complex than that of the normal distribution. The frequency distribution for the basic Weibull function is given by the following:

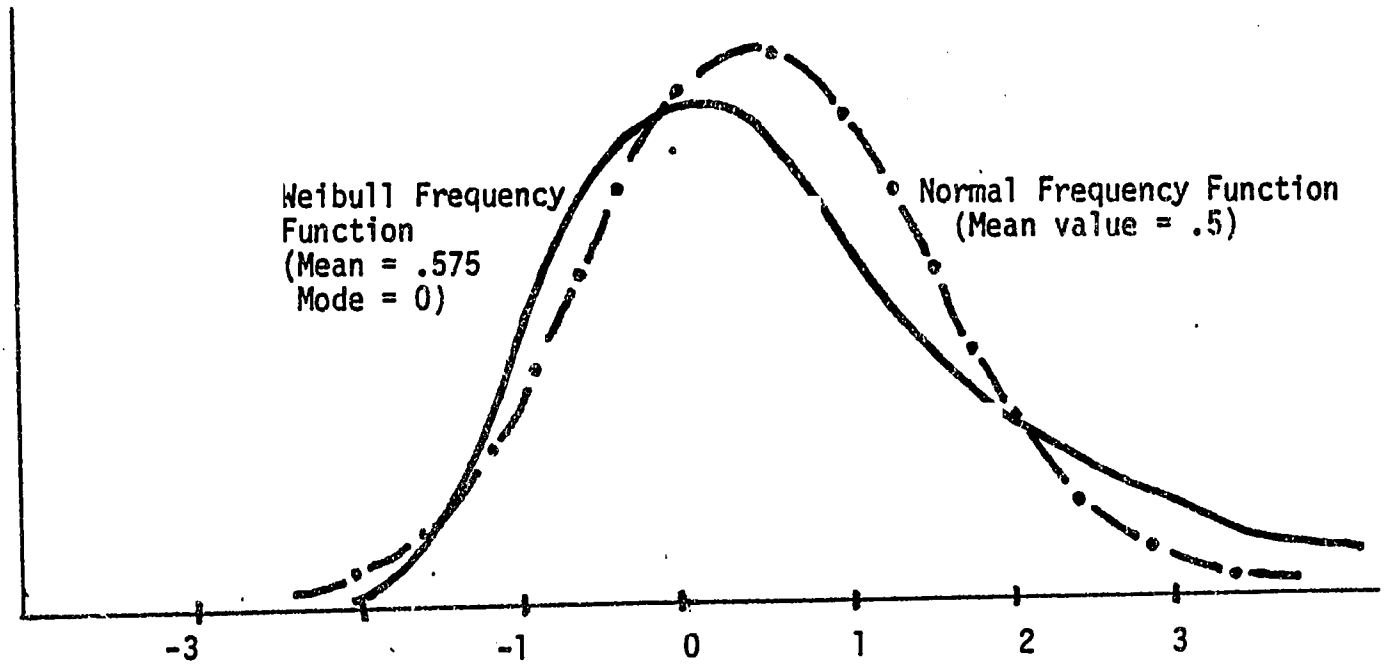
$$(16) \quad F(\eta) = e^{[-(\eta+\alpha)-e^{-(\eta+\alpha)}]}$$

where n is the value of the random variable which is Weibull-distributed and $\alpha$ is the parameter which sets the modal (most frequently-observed) value in the distribution. For purposes of comparison Fig. 3 shows a comparative plot of a Weibull frequency function with mode $\alpha = 0$ and a normal distribution with mean = .5 and standard deviation = 1 (=variance, in this case). The mean value for the Weibull-distribution variable is .575, so that the illustration is designed to show the true degree of divergence in the two distributions.

---

1. Much of the material in Section III represents a detailed re-working of the presentation in McFadden (1975).

Figure 3



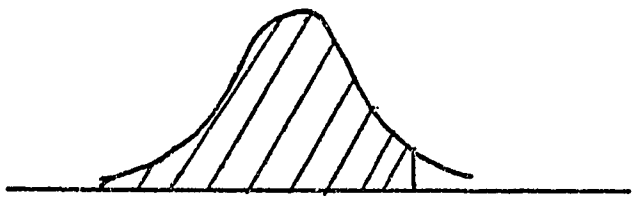

By comparison with the normal frequency function, the Weibull function is somewhat skewed to the left, so that negative values of $\eta_i$ descend more rapidly in frequency than positive values. Nevertheless, if we want to consider the Weibull distribution as an approximation of the normal distribution, we can see that it is not badly biased. The major advantage of the Weibull distribution as a characterization of the error term in the utility model emerges from one simple proposition: The cumulative probability function associated with the difference between two Weibull-distributed random variables is an extremely simple mathematical form called the binary logit. Since the difference between two random error terms is the stochastic foundation of the choice model, this simple result greatly facilitates maximum likelihood estimation. Before this can be clearly seen, however, it will be helpful to show how the above result has been established.

Suppose that the cumulative probability function for Weibull-distributed variable $\eta_2$ is being considered.

Recall that for any continuous random variable, the cumulative probability function (also called the cumulative distribution function, or simply the distribution function) is defined as

$$F(\eta) = \int_{-\infty}^{\eta} F(\eta)d\eta$$

This is a mathematical representation of the proportion of total area under the probability density function (p.d.f.) from $-\infty$ to $\eta$. Thus:



$F(\eta) =$

It is also clear that the p.d.f. can be defined as the first derivative of the c.d.f., evaluated at $\eta$.

$$F'(\eta) = (\eta) \quad [\text{This follows from } F(\eta) = \int_{-\infty}^{\eta} F(\eta)d\eta]$$

Thus, we can see that for a Weibull-distributed variable, the c.d.f. is:

$$F(\eta) = e^{-e^{-(\eta+\alpha)}}$$

$$F(\eta) = F'(\eta) = d\left[e^{-e^{-(\eta+\alpha)}}\right] / d\eta = (-1)\left[-e^{-(\eta+\alpha)}\right] e^{-e^{-(\eta+\alpha)}}$$

$$= e^{-(\eta+\alpha)}\, e^{-e^{-(\eta+\alpha)}}$$

We know already that this latter formula is the frequency function for a Weibull-distributed variable. Thus, the c.d.f. for Weibull-distributed variable $\eta_2$ is given by:

$$(16)\quad F(\eta_2) = \text{Prob}[\eta_2 < \eta] = e^{-e^{-(\eta+\alpha)}}$$

This c.d.f. also has a name -- it is called the Gnedenko distribution. Thus, for the Weibull variable $\eta_2$, the probability that any draw from its distribution will be less than a particular value is given by $F(\eta_2)$ above. As $\eta$ (the comparison value) rises, so (logically enough) does the probability that a random draw from the $\eta_2$-distribution will be less than $\eta$. The Gnedenko formula simply yields numerical values for this rising probability in the case of a random variable whose frequency function has the Weibull shape.

Now, let us suppose that the comparison value is itself a draw from a Weibull distribution, so that its relative frequency is given by:

$$(17)\quad F(\eta_1) = e^{-(\eta+\alpha)}\, e^{-e^{-(\eta+\alpha)}}$$

Consider the comparison problem in two-step sequence. Each value of $\eta_2$ can vary from $-\infty$ to $+\infty$, although not all comparison values are equally probable. The relative frequencies can be read out of the Weibull distribution for $\eta_1$. For each of these comparison values in turn, there is a definite probability that $\eta_2$ is smaller. This probability can be read out of the Gnedenko distribution for $\eta_2$. Thus, for any given pair of values, we know that the probability

that the $\eta_1$ draw will be greater than the $\eta_2$ draw is given by:

(18) $\text{Prob}\,[\eta_{1i}, \eta_{1i} > \eta_2] = \text{Prob}(\eta_{1i}) \cdot \text{Prob}(\eta_{1i} > \eta_2)$

The joint probability is the product of the individual probabilities here because the draws on $\eta_1$ and $\eta_2$ are independent of one another. Now, (18) gives us the joint probability for one particular value of $\eta_1$. There are, however, an infinite number of possible values of $\eta_1$ (from $-\infty$ to $+\infty$). We know from simple probability theory that Prob(A or B) = Prob(A) + Prob(B) when A and B are draws on a random variable. Thus, if there were a finite number (M) of possible values in the distribution $\eta_1$, we would have

$$\text{Prob}[\eta_{11} \text{ or } \eta_{12} \text{ or } \cdots \eta_{1m}] = \text{Prob}(\eta_{11}) + \cdots + \text{Prob}(\eta_{1m}) = \sum_{j=1}^{m} \text{Prob}(\eta_{1j})$$

Similarly, we would have

(19) $\text{Prob}[(\eta_{11}, \eta_{11} > \eta_2) \text{ or } (\eta_{12}, \eta_{12} > \eta_2) \text{ or } \cdots (\eta_{1m}, \eta_{1m} > \eta_2)]$

$= \text{Prob}(\eta_{11})\,\text{Prob}(\eta_{11} > \eta_2) + \cdots + \text{Prob}(\eta_{1m})\,\text{Prob}(\eta_{1m} > \eta_2)$

$= \sum_{j=1}^{m} \text{Prob}(\eta_{1j})\,\text{Prob}(\eta_{1j} > \eta_2)$

In our case, of course, we have an infinite number of such possible draws on $\eta_1$. Recall that $f_1(\eta)$ is another way of representing the p.d.f. for $\eta_1$, and $F_2(\eta)$ is another way of representing the Gnedenko distribution for $\eta_2$. We can therefore replace the discrete probability expression above with an analogous integral to represent summation along values of the continuous variable $\eta_1$.

(20) $\text{Prob}[\eta_1, \eta_1 > \eta_2] = \int_{-\infty}^{+\infty} f_1(\eta) F_2(\eta)\, d\eta$

We now have the probability function which we actually need to get a determinate solution to the Weibull case. Substituting, we obtain

$$\int_{-\infty}^{+\infty} f_1(\eta)F_2(\eta)d\eta = \int_{-\infty}^{+\infty} e^{-(\eta+\alpha_1)}\, e^{-e^{-(\eta+\alpha_1)}}\, e^{-e^{(-\eta+\alpha_2)}}\, d\eta$$

$$= \int_{-\infty}^{+\infty} e^{-(\eta+\alpha_1)}\, e^{-e^{-\eta}[e^{-\alpha_1}+e^{-\alpha_2}]}\, d\eta = \int_{-\infty}^{+\infty} e^{-\alpha_1}\, e^{-\eta}\, e^{-\theta e^{-\eta}}$$

$$= e^{-\alpha_1} \int_{-\infty}^{+\infty} e^{-\eta}\, e^{-\theta e^{-\eta}} \quad (\theta = [e^{-\alpha_1} + e^{-\alpha_2}]$$

Although the integral looks complex, the magical properties of e, the natural logarithm base, make it quite simple. It can easily be seen to be

$$\frac{1}{\theta} e^{-\theta e^{-\eta}}$$

[Thus, $\frac{d}{d\eta}[\frac{1}{\theta} e^{-\theta e^{-\eta}}] = \frac{1}{\theta}(-\theta e^{-\eta})(-1)\, e^{-\theta e^{-\eta}} = e^{-\eta}\, e^{-\theta e^{-\eta}}$]

The joint probability turns out to be:

$$\int_{-\infty}^{+\infty} f_1(\eta)\, F_2(\eta)d\eta = e^{-\alpha_1} \int_{-\infty}^{+\infty} e^{-\eta}\, e^{-\theta e^{-\eta}}\, d\eta = e^{-\alpha_1} \frac{1}{\theta} e^{-\theta e^{-\eta}}\Big]_{-\infty}^{+\infty}$$

$$\theta = [e^{-\alpha_1} + e^{-\alpha_2}]$$

Thus, we have

$$\frac{e^{-\alpha_1}}{e^{-\alpha_1}+e^{-\alpha_2}} \cdot e^{-\theta e^{-\eta}}\Bigg]_{-\infty}^{+\infty} = \frac{e^{-\alpha_1}}{e^{-\alpha_1}+e^{-\alpha_2}} \cdot \frac{1}{e^{\theta \frac{1}{e^{\eta}}}}\Bigg]_{-\infty}^{+\infty}$$

$$= \frac{e^{-\alpha_1}}{e^{-\alpha_1}+e^{-\alpha_2}} \quad \frac{1}{e^{\theta(1/\infty)}} - \frac{1}{e^{\theta(\infty)}} = \frac{e^{-\alpha_1}}{e^{-\alpha_1}+e^{-\alpha_2}}[1-0] = \frac{e^{-\alpha_1}}{e^{-\alpha_1}+e^{-\alpha_2}}$$

This is a truly remarkable result. For two independent, Weibull-distributed variables with modal values $\alpha_1$ and $\alpha_2$, we have

$$\text{Prob}\ (\eta_1, \eta_1 > \eta_2) = \frac{e^{-\alpha_1}}{e^{-\alpha_1}+e^{-\alpha_2}} = \frac{1}{1+e^{\alpha_1-\alpha_2}}$$

This is a binomial result because we have been considering the problem of a simple choice between modes. However, we can generalize this result greatly by establishing a few relatively simple propositions.

(1) If $\eta_i$ has a Weibull distribution with mode $\alpha_i$, then for any scalar V, $\eta_i + V$ has a Weibull distribution with parameter $\alpha_i - V$.

We can establish this most easily by using the c.d.f. of the Weibull distribution (the Gnedenko distribution) for $\eta_i$.

Since $\eta_i + V < \eta \Rightarrow \eta_i < \eta - V$, we have

$$\text{Prob}\ [\eta_i + V < \eta] = \text{Prob}[\eta_i < \eta - V]$$

But we know that $\eta_i$ has the Gnedenko distribution:

$$\text{Prob}\ [\eta_i < \eta] = e^{-e^{(-\eta+\alpha_i)}}$$

Here, we simply replace $\eta$ by $\eta - V$:

$$(21)\quad \text{Prob}[\eta_i < \eta - V] = e^{-e^{-(\eta-V+\alpha_i)}} = e^{-e^{-(\eta+\alpha_i-V)}}$$

We can immediately see that $\eta_i + V$ is distributed as Weibull with mode $\alpha_i - V$.

(2) If $\eta_1, \cdots, \eta_M$ are all Weibull-distributed random variables, then in repeated sets of independent draws on all the random variables, the distribution of the maximum values for each set will also be Weibull.

The easiest way to think about this problem is to imagine a particular comparison value, $\eta$. In any simultaneous draw on our M independent Weibull variables, we know that one variable value will be the maximum in the set, but we cannot know which one. Therefore, we can express the probability that the largest draw in the set will be less than $\eta$ by:

$$(22)\quad \text{Prob}\ [\max_{(i=1,\cdots M)} \eta_i < \eta] = \text{Prob}[(\eta_i < \eta),\cdots, \eta_M < \eta] = \text{Prob}(\eta_1 < \eta) \cdots \text{Prob}(\eta_M < \eta)$$

This formulation holds because the draws are independent. But the latter product equals:

$$e^{-e^{-(\eta+\alpha_1)}} \cdot e^{-e^{-(\eta+\alpha_2)}} \cdot \ldots \cdot e^{-e^{-(\eta+\alpha_M)}} = e^{-\left[e^{-\eta+\alpha} + \cdots + e^{-\eta+\alpha_M}\right]}$$

$$= e^{-e^{\eta}(e^{-\alpha_1} + \cdots + e^{-\alpha_M})}$$

$$(23) \quad = e^{-e^{-\eta} \sum_{i=1}^{M} e^{-\alpha_i}}$$

But since $e^{-\alpha_1}, \cdots, e^{-\alpha_M}$ are all fixed values, we can find an $\alpha$ such that

$$e^{-\alpha} = \sum_{i=1}^{M} e^{-\alpha_i}$$

We can therefore substitute (21) into (22):

$$(24) \quad \text{Prob}[\max_{i=1,\cdots,M} \eta_i < \eta] = e^{-e^{-\eta} \sum_i e^{-\alpha_i}} = e^{-e^{-\eta} e^{-\alpha}} = e^{-e^{-(\eta+\alpha)}}$$

which is distributed as Weibull with modal value $\alpha$.

Now we can use propositions (1), and (2)., just demonstrated, to generalize our binomial comparison probability to a multinomial probability. Recall that

$$\text{Prob}(\eta_1 > \eta_2) = \frac{e^{-\alpha_1}}{e^{-\alpha_1} + e^{-\alpha_2}}$$

where $\alpha_1$ and $\alpha_2$ are the modal values for the two distributions.

By (1) above, we know that

$$(25) \quad \text{Prob}(\eta_1 + V_1 > \eta_2 + V_2) = \frac{e^{-(\alpha_1 - V_1)}}{e^{-(\alpha_1 - V_1)} + e^{-(\alpha_2 - V_2)}} = \frac{e^{V_1 - \alpha_1}}{e^{V_1 - \alpha_1} + e^{V_2 - \alpha_2}}$$

178

since $(\eta_1 + V_1)$ and $(\eta_2 + V_2)$ are both distributed as Weibull with modal values $(\alpha_1 - V_1)$ and $(\alpha_2 - V_2)$, respectively. From (25), we have

$$(26) \quad \text{Prob}\,[\eta_1 + V_1 > \eta_2 + V_2] = \text{Prob}\,[\eta_2 - \eta_1 < V_1 - V_2] = \frac{e^{V_1-\alpha_1}}{e^{V_1-\alpha_1} + e^{V_2-\alpha_2}}$$

$$= \frac{1}{1 + e^{(\alpha_1-\alpha_2) - (V_2-V_1)}}$$

When written this way, (26) tells us what we need to know in order to do econometric work in the context of binomial choice. Recall that this all began with a model of discrete choice in the context of binary comparison.

If $U_{i1} = Z_{i1}\beta + \varepsilon_{i1}$ and $U_{i2} = Z_{i2} + \varepsilon_{i2}$, then the probability that 1 will be chosen over 2 is given by

$$\text{Prob}\,[U_{i1} > U_{i2}] = \text{Prob}[Z_{i1}\beta + \varepsilon_{i2} > Z_{i2}\beta + \varepsilon_{i1}] = \text{Prob}[(Z_{i1} - Z_{i2})\beta > \varepsilon_{i2} - \varepsilon_{i1}]$$

Now, if we suppose that both errors are distributed as Weibull rather than normally, we can see where our results have brought us. Since $Z_{i1}$ and $Z_{i2}$ are known attribute vectors and the $\beta$'s are to be thought of as fixed parameter values, we can think of $\beta' Z_{i1}$ and $\beta' Z_{i2}$ as our scalars $V_1$ and $V_2$ and substitute into (26)

$$(27) \quad \text{Prob}[1 \text{ chosen over } 2] = \text{Prob}[\varepsilon_{i2} - \varepsilon_{i1} < (Z_{i1} - Z_{i2})\beta =$$

$$= \frac{1}{1 + e^{(\alpha_1-\alpha_2) + (Z_{i2}-Z)\beta}}$$

As we will see shortly, our mathematical groundlaying has now brought us quite close to the point where we can actually begin thinking about ways of estimating the $\beta$-values. First, however, let's generalize (27) using (24) and (21).

$$\text{Prob}\,[\eta_1 + V_1 > \eta_1^* + V_1^*] = \frac{e^{V_1-\alpha_1}}{e^{V_1-\alpha_1} + e^{-\alpha}}$$

where $\eta_1^* + V_1^* = \underset{i=1,\ldots,M}{\text{Max}}(\eta_i + V_i)$

Translating this into a form which is convenient for our work, we have

$$(28)\quad \text{Prob}[\eta_1^* - \eta_1 < V_1 - V_1^*] = \frac{e^{V_1-\alpha_1}}{e^{V_1-\alpha_1} + e^{-\alpha}} = \frac{e^{V_1-\alpha_1}}{\sum_{i=1}^{M} e^{V_i-\alpha_i}}$$

This allows us to reformulate our binary choice model as a multiple choice model. Suppose that M choices are available to the consumer instead of two. Then the consumer will choose 1 over the other (M-1) alternatives only if

$$(29)\quad U_{ij}(Z_{i1}) + \varepsilon_{i1} > \underset{j=1,\ldots,M-1}{\text{Max}}[U_{ij}(Z_{ij}) + \varepsilon_{ij}] \quad \text{or} \quad U_{i1}(Z_{i1}) + \varepsilon_{i1} > U_{ij}^*(Z_{ij}^*) + \varepsilon_{ij}^*$$

Thus, the probability that mode 1 will be chosen by the ith consumer, given M-1 other choices, is given by

$$(30)\quad \text{Prob}[1 \text{ is chosen}] = \text{Prob}[U_{ij}(Z_{i1}) + \varepsilon_{ij} > U_{ij}^*(Z_{ij}^*) + \varepsilon_{ij}^*]$$

$$= \text{Prob}[\varepsilon_{ij}^* - \varepsilon_{i1} < U_{i1}(Z_{i1}) - U_{ij}^*(Z_{ij}^*)] = \text{Prob}[\varepsilon_{ij}^* - \varepsilon_{i1} < (Z_{i1} - Z_{ij}^*)\beta$$

$$= \frac{e^{V_1-\alpha_1}}{\sum_{j=1}^{M} e^{\beta' Z_{ij}-\alpha_j}} = \frac{e^{V_1-\alpha_1}}{\sum_{j=1}^{M} e^{Z_{ij}\beta-\alpha_j}}$$

Expressions (30) and (27) will serve as the foundations for our discussion of maximum-likelihood estimation of the β-values in the binary and multiple choice cases, respectively.

## IV. Parametric Estimation with the Logit Model

At this point, it may be worthwhile to turn back to pp. 3-5 to review the essentials of the likelihoodist approach. In brief, we can say that the likelihoodist regards the sample of observations which has been drawn as the most likely of all possible samples of the same size. Thus, the appropriate estimates for the underlying parameters in the system model are regarded as being those which maximize the likelihood of the sample. Maximum likelihood estimates are always determined from the application of the calculus to a likelihood function specified with reference to a particular model of the error distribution in a particular context.

In returning to the problem of parameter estimation, then, we will begin with a consideration of appropriate forms for the likelihood function. As a prelude, we will introduce the most common general specification of the likelihood function:

$$(31) \quad L(\theta) = \prod_{i=1}^{M} \prod_{k=1}^{N} P_{ik}^{y_{iK}}$$

where the likelihood function $L(\theta)$ is defined over all households ($i = 1,\ldots,M$) and alternatives ($K = 1,\ldots,N$); $y_{iK} = 1$ if household i chooses alternative K and 0 otherwise. $\sum_{K=1}^{N} P_{iK} = 1$ for any individual i.

We can begin our explanation of this function by considering a simple unconditional case. Suppose that there are two households in our sample ($M = 2$) and two modal alternatives under consideration ($N = 2$). Since the sample points are drawn independently of one another (this has to be assured by the sampling method), we know that the joint probability of a particular set of modal choices for the two households is just the product of their separate probabilities. Thus, suppose that we have an unconditional probability problem of the following

sort.

| Household | Mode 1 | Mode 2 |
|---|---|---|
| 1 | P = 1/3 | P = 2/3 |
| 2 | P = 1/4 | P = 3/4 |

The following joint probabilities hold (here P(i,j) means the probability that household 1 will choose mode i and household 2 mode j simultaneously):

$P(1,1) = 1/3 \cdot 1/4 = 1/12$

$P(1,2) = 1/3 \cdot 3/4 = 3/12$

$P(2,1) = 2/3 \cdot 1/4 = 2/12$

$P(2,2) = 2/3 \cdot 3/4 = 6/12$

Note the common-sense sets of totals in this problem. For each family,

$\sum_{K=1}^{2} P_{iK} = 1$ [Family 1: 1/3 + 2/3 = 1; Family 2: 1/4 + 3/4 = 1]

In addition, the sum of all the joint probabilities is equal to 1:

$P(1,1) + P(1,2) + P(2,1) + P(2,2) = 1/12 + 3/12 + 2/12 + 6/12 = 1$

Again, of course, this is no accident. Under the criterion that some choice be made, the four joint probabilities provide an exhaustive description of the possible outcomes. Since one of those outcomes must occur, the sum of the joint probabilities must be 1.

Now, there is very little difference between the simple specification of joint probability outcomes which we have just presented and the general specification of the likelihood function in (30). Recall that the likelihood function itself is nothing more than a statement of the joint probability of a set of sample draws. The challenge lies in specifying a general likelihood function which characterizes the full set of possible draws for a set of households. For our 2-household, 2-mode case, 4 such joint draws are possible.

When we examine (31), we can see that it is actually designed to include all of these possibilities simultaneously.

In order to see how, we will write out a full version of (31) for our 2 x 2 case.

$$(32)\ L(\theta) = \prod_{i=1}^{2} \prod_{K=1}^{2} P_{iK}^{y_{iK}} = P_{11}^{y_{11}} \cdot P_{12}^{y_{12}} \cdot P_{21}^{y_{21}} \cdot P_{22}^{y_{22}} = (1/3)^{y_{11}} (2/3)^{y_{12}}$$

$$(1/4)^{y_{21}} (3/4)^{y_{22}}$$

The actual value of the likelihood function will depend on the values taken by the $y_{iK}$. Recall the original definition: $y_{iK} = 1$ for the alternative which a household actually chooses and 0 for the others. Thus, for our case we have the four possibilities which we have already observed:

| Choice | $y_{11}$ | $y_{12}$ | $y_{21}$ | $y_{22}$ | $L(\theta)$ |
|---|---|---|---|---|---|
| (1,1) | 1 | 0 | 1 | 0 | $(1/3)^1 (2/3)^0 (1/4)^1 (3/4)^0 = 1/12$ |
| (1,2) | 1 | 0 | 0 | 1 | $(1/3)^1 (2/3)^0 (1/4)^0 (3/4)^1 = 3/12$ |
| (2,1) | 0 | 1 | 1 | 0 | $(1/3)^0 (2/3)^1 (1/4)^1 (3/4)^0 = 2/12$ |
| (2,2) | 0 | 1 | 0 | 1 | $(1/3)^0 (2/3)^1 (1/4)^0 (3/4)^1 = 6/12$ |

It is clear that (32) is simply a general statement of joint probability which allows for all the choices made by households. Now we can make use of (32) to go one step further and consider the kind of problem which is of direct interest. Note that in the example which has just been presented, the probabilities under consideration were unconditional. That is, we were regarding them as fixed, rather than as varying with the characteristics of the households or modes themselves. As a first introduction to maximum likelihood estimation in the current context, let us continue the assumption of unconditional probability and observe the way in which we would set maximum-likelihood estimates of modal choice probabilities for a set of households whose choices have been

observed.

We will suppose that we are observing the choice between two modes made by five households. The general likelihood function is

$$L(\theta) = \prod_{i=1}^{5} \prod_{K=1}^{2} P_{iK}^{y_{iK}} = P^{y_{11}}(1-P)^{y_{12}} \cdot P^{y_{21}}(1-P)^{y_{22}} \ldots P^{y_{52}}(1-P)^{y_{52}}$$

(Note: Since P is unconditional, it doesn't vary across households and modes.)

In our sample, we observe the following choices:

| | | | | | |
|---|---|---|---|---|---|
| Household: | 1 | 2 | 3 | 4 | 5 |
| Mode: | 1 | 2 | 1 | 1 | 2 |

Now we have a likelihood function which is tied to the choices already made:

$$(33)\ L(\theta) = P^1(1-P)^0P^0(1-P)^1P^1(1-P)^0P^1(1-P)^0P^0(1-P)^1 = P^3(1-P)^2$$

We are interested in estimating P. In the likelihoodist view, P should be the probability which makes the likelihood (joint probability) of the sample drawn the highest possible. One further trick is useful in solving this problem (we saw this trick before on p. 4). Since there is a one-to-one correspondence between positive numbers and their logarithms, we can replace the likelihood function with its logarithm without changing anything. Thus, we can rewrite (33) as

$$\log(L) = 3\log P + 2\log(1-P)$$

An application of simple calculus yields the result which we are seeking:

$$\frac{d(\log L)}{dP} = \frac{3}{\hat{P}} - \frac{2}{1-\hat{P}} = \frac{3-3\hat{P}-2\hat{P}}{\hat{P}(1-\hat{P})} = 0 \Longrightarrow 3-5\hat{P} \text{ and } \hat{P} = 3/5$$

Thus, 3/5 is the "maximum likelihood estimator" of P, the unconditional probability that a household in our sample will choose mode 1. Referring back to our original sample, we can see that we have always used maximum likelihood estimation in our calculation of probabilities without knowing it. Since 3 of

5 houses choose mode 1, our reflexive estimate of the unconditional probability would also be 3/5.

Hopefully, this simple set of examples has communicated the flavor of maximum likelihood estimation (MLE) in modal choice situations. Now we are ready to move on to a conditional probability case. Suppose that we have drawn a sample of M families, each confronting the same set of N alternative modes. Our likelihood function is:

$$L = \prod_{i=1}^{M} \prod_{j=1}^{N} P_{ij}^{y_{ij}}$$

or (in logarithmic form):

$$(34)\quad L = \prod_{i=1}^{M} \prod_{j=1}^{N} y_{ij} \log P_{ij}$$

Before going further, we have to specify the conditional probability function which is to be put into the likelihood function. Recall from our earlier discussion of multimodal choice that the probability that a mode (K) will be chosen over its (N-1) competitors by a consumer (i) is given by:

$$(35)\quad \text{Prob}[U_{iK}(W_{iK}) + \varepsilon_{iK} > U^*_{ij}(W_{ij}) + \varepsilon^*_{ij}]$$

where $U^*_{ij}(W_{ij}) + \varepsilon^*_{ij} = \underset{j \neq K}{\text{Max}} [U_{ij}(W_{ij}) + \varepsilon_{ij}]$

As in (30) we obtain

$$(36)\quad \text{Prob}[K \text{ is chosen}] = \frac{e^{Z_{iK}\beta - \alpha_K}}{\sum_{j=1}^{N} e^{Z_{ij}\beta - \alpha_j}}$$

At this point, we can introduce one simplification. Note that we have been incorporating a constant term into our utility model, so that any expression such as $(Z_{ij}\beta - \alpha_j)$ has two constant terms. Since it is more convenient expositionally

(and since we cannot in any case estimate two constant terms in the same function), we will simply absorb the $\alpha$'s into the constant components of the utility function definition, so that we can rewrite (36) as:

$$(37)\quad \text{Prob[K is chosen]} = \frac{e^{Z_{iK}\beta}}{\sum_{j=1}^{N} e^{Z_{ij}\beta}}$$

With this general specification of the conditional probability that one mode will be chosen among several, we can fill in the likelihood function:

$$(38)\quad L = \prod_{i=1}^{M} \prod_{j=1}^{N} y_{ij} \log P_{ij} = \sum_{i=1}^{M} \sum_{j=1}^{N} y_{ij} \left[ \log \frac{e^{Z_{iK}\beta}}{\sum_{j=1}^{N} e^{Z_{ij}\beta}} \right]$$

which can also be written as:

$$L = - \sum_{i=1}^{M} \sum_{j=1}^{N} y_{ij} \log \frac{\sum_{j=1}^{N} e^{Z_{ij}\beta}}{e^{Z_{iK}\beta}}$$

Through additional manipulation of the probability term we obtain:

$$(39)\quad L = - \sum_{i=1}^{M} \sum_{j=1}^{N} y_{ij} \log \left[ \left( \sum_{j=1}^{N} e^{Z_{iK}\beta} \right) \left( e^{Z_{ij}\beta} \right)^{-1} \right]$$

$$= - \sum_{i=1}^{M} \sum_{j=1}^{N} y_{ij} \log \left[ g_1(\beta)\, g_2(\beta) \right]$$

In the unconditional problem which we used to introduce MLE methods, the probability was the parameter which we sought and we solved for it directly. In the present case, however, the probability that any mode will be chosen de-

pends on various modal and personal characteristics. The relationship between these characteristics and consumer utility is assumed to be defined by a set of utility function parameters ($\beta$). The unknown $\beta$-values can be determined by the set of first-order conditions for a maximum value of (39) with respect to $\beta$:

$$\frac{\partial L}{\partial \beta} = - \sum_{i=1}^{M} \sum_{j=1}^{N} y_{ij} \frac{1}{g_1(\beta) g_2(\beta)} [g_1'(\beta) g_2(\beta) + g_1(\beta) g_2'(\beta)]$$

$$= - \sum_{i=1}^{M} \sum_{j=1}^{N} y_{ij} \frac{g_1'(\beta)}{g_1(\beta)} + \frac{g_2'(\beta)}{g_2(\beta)}$$

Here $g_1'(\beta) = \dfrac{\partial \left[ \sum_{j=1}^{N} e^{Z_{ij}\beta} \right]}{\partial \beta} = \sum_{j=1}^{N} Z_{ij} e^{Z_{ij}\beta}$

and $g_2'(\beta) = \dfrac{\partial \left[ (e^{Z_{iK}\beta})^{-1} \right]}{\partial \beta} = -Z_{iK} e^{-Z_{iK}\beta}$

Thus, $\dfrac{g_1'(\beta)}{g_1(\beta)} = \dfrac{\sum_{j=1}^{N} Z_{ij} e^{Z_{ij}\beta}}{\sum_{j=1}^{N} e^{Z_{ij}\beta}} = \dfrac{Z_{i1} e^{Z_{i1}\beta} + \cdots + Z_{iN} e^{Z_{iN}\beta}}{\sum_{j=1}^{N} e^{Z_{iN}\beta}}$

A glance at (37) makes it clear that the expression above is equal to

$$Z_{i1} P_{i1} + \cdots + Z_{iN} P_{iN}$$

In addition,

$$\frac{g_2'(\beta)}{g_2(\beta)} = \frac{-Z_{ij} e^{-Z_{ij}\beta}}{e^{-Z_{ij}\beta}} = -Z_{ij}$$

Thus, we have

$$\frac{\partial L}{\partial \beta} = - \sum_{i=1}^{M} \sum_{j=1}^{N} y_{ij} \frac{g_1'(\beta)}{g_1(\beta)} + \frac{g_2'(\beta)}{g_2(\beta)} = - \sum_{i=1}^{M} \sum_{j=1}^{N} y_{ij} [\sum_{j=1}^{N} Z_{ij}P_{ij} - Z_{ij}]$$

$$= - \sum_{i=1}^{M} [(\sum_{j=1}^{N} Z_{ij}P_{ij}) \sum_{j=1}^{N} y_{ij} - \sum_{j=1}^{N} y_{ij}Z_{ij}]$$

Recall that $\sum_{j=1}^{N} y_{ij} = 1$, since each family can choose only one of the available modes.

Thus, $\frac{\partial L}{\partial \beta} = - \sum_{i=1}^{M} [\sum_{j=1}^{N} Z_{ij}P_{ij} - \sum_{j=1}^{N} y_{ij}Z_{ij}]$

or equivalently,

$$(40) \quad \frac{\partial L}{\partial \beta} = - \sum_{i=1}^{M} \sum_{j=1}^{N} (Z_{ij}P_{ij} - y_{ij}Z_{ij}) = - \sum_{i=1}^{M} \sum_{j=1}^{N} (P_{ij} - y_{ij})Z_{ij}$$

$$= \sum_{i=1}^{M} \sum_{j=1}^{N} (y_{ij} - P_{ij})Z_{ij}$$

Since the MLE estimates of the β's are obtained from an evaluation of the first partial derivatives at the point where the likelihood function is stationary, we have:

$$(41) \quad \frac{\partial L}{\partial \beta} = \sum_{i=1}^{M} \sum_{j=1}^{N} (y_{ij} - P_{ij})Z_{ij} = 0$$

188

This general expression is appropriate for the estimation of parameter values in the case of multinominal (multimodal) choice. It is easy to simplify this expression in the case of binomial (bimodal) choice:

$$(42)\quad \frac{\partial L}{\partial \beta} = \sum_{i=1}^{M} \sum_{j=1}^{2} (y_{ij} - P_{ij})Z_{ij} = \sum_{i=1}^{M} [y_{i1} - P_{i1})Z_{i1} - (y_{i2} - P_{i2})Z_{i2}]$$

$$= \sum_{i=1}^{M} [(y_{i1} - P_{i1})Z_{i1} + [(1 - y_{i1}) - (1 - P_{i1})]\ Z_{i2}]$$

$$= \sum_{i=1}^{M} [(y_{i1} - P_{i1})Z_{i1} - (y_{i1} - P_{i1})Z_{i2}]$$

$$= \sum_{i=1}^{M} (y_{i1} - P_{i1})(Z_{i1} - Z_{i2}) = 0$$

Vector equations (41) and (42) define the simultaneous sets of nonlinear equations which can be solved to yield maximum-likelihood parameter values in the binomial and multinomial choice cases. In order to write out the appropriate simultaneous systems of equations, we have only to remember that $\frac{\partial L}{\partial \beta}$ is a vector of first partial derivatives and that the Z's are corresponding transformations of personal and modal characteristic variables which enter into the determination of utility in each case.

As a preliminary illustration, consider the bimodal case. Let C be the number of households who have chosen mode 1 and reorder the observations so that these households appear first in the sample. We have a set of T equations:

$$(43)\quad \frac{\partial L}{\partial \beta} = \begin{bmatrix} \frac{\partial L}{\partial \beta} \\ \vdots \\ \frac{\partial L}{\partial \beta_T} \end{bmatrix} = \sum_{i=1}^{M} (y_{i1} - P_{i1}) \begin{bmatrix} Z_{i1}^{1} - Z_{i2}^{1} \\ \vdots \\ Z_{i1}^{T} - Z_{i2}^{T} \end{bmatrix}$$

or,

$$(2)\quad \sum_{i=1}^{C} (Z_{i1}^{1} - Z_{i2}^{1}) - \sum_{i=1}^{C} (Z_{i1}^{1} - Z_{i2}^{1}) \left( \frac{e^{Z_{i1}\hat{\beta}}}{e^{Z_{i1}\hat{\beta}} + e^{Z_{i2}\beta}} \right) = 0$$

$$\vdots$$

$$(T)\quad \sum_{i=1}^{C} (Z_{i1}^{T} - Z_{i2}^{T}) - \sum_{i=1}^{C} (Z_{i1}^{T} - Z_{i2}^{T}) \left( \frac{e^{Z_{i1}\hat{\beta}}}{e^{Z_{i1}\hat{\beta}} + e^{Z_{i2}\hat{\beta}}} \right) = 0$$

We can now extend this same logic to the more complex, N-mode case. Here, we will use $C_1, \ldots, C_N$ to indicate the number of households who have chosen modes $1, \ldots, N$. In addition, we will assume the households to be ordered by choice $\{1 \rightarrow N\}$ for expositional convenience.

$$(44)\quad \frac{\partial L}{\partial \beta} = \begin{bmatrix} \frac{\partial L}{\partial \beta_1} \\ \vdots \\ \frac{\partial L}{\partial \beta_T} \end{bmatrix} = \sum_{i=1}^{M} \left[ (y_{i1} - P_{i1}) \begin{bmatrix} Z_{i1}^{1} \\ \vdots \\ Z_{i1}^{T} \end{bmatrix} + \ldots + (y_{iN} - P_{iN}) \begin{bmatrix} Z_{iN}^{1} \\ \vdots \\ Z_{iN}^{T} \end{bmatrix} \right]$$

or,

$$(1)\quad \sum^{C_1} Z^1_{i1} + \sum^{C_2} Z^1_{i2} + \dots + \sum^{C_N} Z^1_{iN} - \sum_{i=1}^{M} \left[ \frac{\sum_{j=1}^{N} Z^1_{ij}\, e^{Z^1_{ij}\hat{\beta}}}{\sum_{j=1}^{N} e^{Z^1_{ij}\hat{\beta}}} \right] = 0$$

$$\vdots$$

$$(T)\quad \sum^{C_1} Z^T_{i1} + \sum^{C_2} Z^T_{i2} + \dots + \sum^{C_N} Z^T_{iN} - \sum_{i=1}^{M} \left[ \frac{\sum_{j=1}^{N} Z^T_{ij}\, e^{Z^T_{ij}\hat{\beta}}}{\sum_{j=1}^{N} e^{Z^T_{ij}\hat{\beta}}} \right] = 0$$

In each of the preceding cases, we have T non-linear equations in T unknowns. Since the equations are nonlinear, the solutions must be generated from iterative procedures. D. McFadden (1975) has shown that the existence of unique solution values is virtually guaranteed in samples of modest size. Several non linear solution methods are available, including the Newton-Raphson, Fletcher-Powell, and Davidon routines.

## V. Prediction Bias and the Logit Model

Generally, discrete choice models have two significant uses. To the extent that the relative attractiveness of a particular mode is dependent upon the value of some policy variable, a fully estimated choice model reveals the sensitivity of modal choice to policy changes. At the same time, the fact that choice models relate the attractiveness of alternatives modes to measurable differences in common underlying characteristics allows for the prediction of demand for new modes which become available. This is because these new modes also have (at least provisionally) measurable values of the same underlying characteristics. Thus, after a full choice model has been estimated for the existing alternatives, the new mode can simply be inserted as an additional equation in the model and the demand shares for each mode recalculated.

In order to see how this process works for the multinomial logit model (and as a prelude to a critique), this section will begin with a simple example. Suppose that we currently have three modes available to consumers in a particular area and that the evaluation of these modes by consumers has been modeled as a linear function of 4 variables. The fitted model has given us the following common parameters of the utility function:

$$U_j = .3X_{1j} + .8X_{2j} - .4X_{3j} + .1X_{4j} + e_j$$

Suppose additionally that the three existing modes have the following attribute values.

| Mode | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| 1 | 1 | 3 | 2 | 6 |
| 2 | 4 | 2 | 1 | 2 |
| 3 | 5 | 2 | 4 | 2 |

In the multinomial choice case, we have been specifying the probability that a particular mode (K) will by chosen by a particular individual (i) as

$$P_{ik} = \frac{e^{Z_{ik}\beta}}{\sum_{j=1}^{N} e^{Z_{ij}\beta}}$$

Note, then, that the relative likelihood that mode K will be chosen over another mode (h) is given by:

$$(45) \quad \frac{P_{iK}}{P_{ih}} = \frac{e^{Z_{iK}\beta} / \sum_{j=1}^{N} e^{Z_{ij}\beta}}{e^{Z_{ih}\beta} / \sum_{j=1}^{N} e^{Z_{ij}\beta}} = \frac{e^{Z_{iK}\beta}}{e^{Z_{ih}\beta}} = e^{(Z_{iK} - Z_{ih})\beta}$$

We can immediately see that this relative likelihood calculation is the key to the estimation of shares, since the share of each mode should be in direct proportion to the probability of its use by a representative consumer. Only attribute characteristics count here. We can therefore calculate the share quite easily, using any of the three modes as the basis of probability comparisons. For our example, mode 1 will be used as the base for shares calculations:*

$$\frac{P_2}{P_1} = e^{.3(4-1) + .8(2-3) - .4(1-2) + .1(2-6)} = e^{.9-.8+.4-.4} = e^{.1}$$

$$\frac{P_3}{P_1} = e^{.3(5-1) + .8(2-3) - .4(4-2) + .1(2-6)} = e^{1.2-.8-.8-.4} = e^{-.8}$$

---

* Here the i's are suppressed because consumer characteristics do not interact with model characteristics to produce a direct effect on modal choice.

ıus,

$\frac{S_2}{S_1} = \frac{P_2}{P_1} = 1.105$ and $\frac{S_3}{S_1} = \frac{P_3}{P_1} = .449$ (Here $S_i$ is the share of mode i)

$S_1 + S_2 + S_3 = S_1 + 1.105S_1 + .449S_1 = 2.554S_1 = 1$

$S_1 = .39 \quad S_2 = .43 \quad S_3 = .18$

Suppose that we introduce another mode, with the following attributes: $X_1 = 2$, $X_2 = 4$, $X_3 = 1$, $X_4 = 3$

Now we see the ease with which multimodal logit models can be employed. Since all the comparisons are binary, the first three ratios are unchanged. Only one additional calculation is necessary.

$$\frac{S_4}{S_1} = \frac{P_4}{P_1} = e^{.3(2-1) + .8(4-3) - .4(1-2) + .1(3-6)} = e^{1.2} = 3.32$$

or $S_4 = 3.332S_1$

Now the shares can be re-estimated:

$S_1 + 1.105S_1 + .449S_1 + 3.32S_1 = 5.875S_1 = 1$

$S_1 = .17 \quad S_2 = .19 \quad S_3 = .08 \quad S_4 = .56$

Two things are immediately apparent about this solution. First, of course, the relative attractiveness of the new mode has resulted in a tremendous diversion of demand away from the original three. Of more fundamental importance is a second point: While mode 4 has diverted much of the existing demand, the shares going to the remaining three modes retain their former proportions (.17/.19/.08 vs..39/.43/.19). This result follows, of course, from our modeling approach: Only binary comparisons matter, so that the relative valuation of the first three modes by consumers is not supposed to

change as a result of the introduction of another mode. Since the form of the logit model embodies this assumption, we should examine it very closely. Just how bad is it likely to be in practice? To test its implications, suppose that the new fourth mode has exactly the same characteristics as the original mode 1:

$$\frac{S_4}{S_1} = \frac{P_4}{P_1} = e^{.3(1-1) + .8(3-3) - .4(2-2) + .1(6-6)} = e^0 = 1$$

and $S_1 + 1.105S_1 + .449S_1 + S_1 = 3.554S_1 = 1$

$$S_1 = .28 \quad S_2 = .31 \quad S_3 = .13 \quad S_4 = .28$$

Suddenly, things look very wrong. Mode 1 and the hypothetical new mode 4 are absolutely identical. Effectively, then, modes 1 and 4 together are simply mode 1, and we would expect them to split the former share of mode 1 -- .39. Our common-sense prediction would therefore be:

$$S_1 = .195 \quad S_2 = .43 \quad S_3 = .18 \quad S_4 = .195$$

When we compare the two predictions above, we see how the use of strict binary comparison in shares calculations can violate common sense. The closer a new mode is to an existing mode, in fact, the larger this kind of bias will be. It is a major limitation of the standard multinomial logit model in cases where a new mode is in many ways comparable to one which already exists.

We have already seen that this problem stems from the binary comparisons which are inherent in the modeling approach. Thus, the model ignores the common-sense restriction that shifts in the shares of existing modes when a new one is introduced must be related to their similarity to the new mode. The greater the similarity, the larger the loss in existing share.

Because the multinomial logit is implicitly based on a model in which modes are totally distinguishable from one another, the underlying principle

is called "the independence of irrelevant alternatives". Since it is clear that alternatives may in fact by highly relevant for modal switching decisions, J. Hausman and D. Wise (1978) have critically re-stated this principle as "the independence of relevant alternatives". It is a tremendous weakness of the standard logit model when the introduction of new modes is being considered.

It might be supposed that the multinomial logit is on firmer ground when interest focuses on the manipulation of policy variables which bear on consumer choice among currently-available modes. Unfortunately, the standard logit model suffers from a serious specification problem even in this more restricted domain. This problem also stems from the imposition of the "independence of relevant alternatives". In order to understand its nature, let us again turn to a simpler (this time four-mode) example:

| Mode | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| 1 | 1 | 3 | 2 | 6 |
| 2 | 4 | 2 | 1 | 2 |
| 3 | 5 | 2 | 4 | 2 |
| 4 | 5 | 2 | 1 | 4 |

Recall that our estimated valuation function here is

$$U_j = .3X_{1j} + .8X_{2j} - .4X_{3j} + .1X_{4j} + e_j$$

The standard calculations yield:

$$\frac{P_4}{P_1} = e^{.3(5-1) + .8(2-3) - .4(1-2) + .1(4-6)} = e^{.6} = 1.822$$

$$S_1 + 1.105S_1 + .499S_1 + 1.822S_1 = 4.426S_1 = 1$$

$$S_1 = .23 \quad S_2 = .25 \quad S_3 = .11 \quad S_4 = .41$$

Now, suppose that the government were to shift the value of attribute $X_2$ in mode 2 in order to make that mode more attractive. Obviously, the ratio $S_2/S_1$ would be affected by this shift. However, the independence of

196

"irrelevant" alternatives assumes that neither $S_3/S_1$ nor $S_4/S_1$ would be affected. Thus, the ratio $S_3/S_4$ $(=S_3/S_1)/(S_4/S_1)$ would not change.

Now, although the ratio $S_3/S_4$ cannot change as a result of a change in the attractiveness of mode 2, it is clear that $S_3$ and $S_4$ will both diminish. In order for their ratio to remain constant, in fact, they must diminish in exactly the same proportion. Since elasticities are defined as ratios of percent changes, we can see immediately that the logit specification implies the constancy of most cross-elasticities in the model. A shift in the value of an attribute of mode 2 will always have the same proportional effects on the shares of modes 3 and 4 when mode 1 attributes are used as the basis for binary comparison. More generally, we could say that for N modes, with mode i used for binary comparison, the cross elasticities of response to an attribute shift in any mode j are constant for all other modes in the model. Obviously, this is highly restrictive and it represents a form of specification bias.

It is clear that the standard multinomial logit model has serious weaknesses, both for the analysis of modal introduction and for policy analysis within an existing system. How can these problems be overcome? At the present time, two approaches are available. One, developed by J. Hausman, C. Manski and others, goes back to first principles and builds a complete model of discrete choice based on normally-distributed rather than Weibull-distributed errors. This approach turns out to have the potential for extreme generality. Besides the obvious virtue of building upon a plausible model of random errors in the choice process, it also allows for the automatic inclusion of the asymmetric reallocation choices which have just been discussed. This modeling technique is known as the "covariance probit" approach, and it will be fully discussed in Sections VII-X. The

state-of-the-art in existing applications software allows for the inclusion of no more than 10 discrete alternatives.

This computational limitation does not apply to the second approach. which has recently been developed by R.S. Hartman. Hartman has generalized the multinomial logit model in a way which overcomes some of the undesirable results of the imposed "independence of irrelevant alternatives". As we will see in the following section, the generalized multinomial logit model offers considerable promise for policy analysis in existing systems, although it does not adapt well to the introduction of new modes.

## VI. The Generalized Multinomial Logit[2]

A brief reconsideration of the logit approach reveals a fortunate coincidence. The computational ease associated with the logit results from the assumption that individual departures from "representative" valuations of alternative modes are random draws from Weibull distributions. Although the Weibull distribution is somewhat skewed by the comparison with the normal distribution (which is certainly a more plausible description of the random error pattern in this case), it is hard to believe that this small degree of bias can have fundamental consequences for prediction capability. This is hopeful, since there are clear benefits to retaining the Weibull distribution.

The more fundamental reason for biased predictions by logit models lies in the way in which the consumer utility function has commonly been specified. For the general class of linearized utility functions, we have:

$$\text{Prob[K is chosen]} = \text{Prob}[\beta_0 + \beta_1 Z^1_{iK} + \ldots + \beta_T Z^T_{iK} + \varepsilon_K > \beta_0 + \beta_1 Z^{*1}_i + \ldots + \beta_T Z^{*T}_i + \varepsilon^*_i]$$

(Here * indicates attribute values for the best alternative mode.)

or,

$$\text{(46)} \quad \text{Prob[K is chosen]} = \text{Prob}[\varepsilon^* - \varepsilon_K < \beta_1(Z^1_{iK} - Z^{*1}_i) + \ldots + \beta_T(Z^T_{iK} - Z^{*T}_i)]$$

From our previous development of the simple linear model, we know that relative probabilities will be basically determined by the differences $[Z_{iK} - Z^*_i]$ on the right-hand side of the inequality above and all comparisons will be effectively binary. With this in mind, we can ask a fundamental question: What happens when, for any individual (i), the utility function elements $Z_{ij}$ are the same for two modes which are being compared? Obviously, the answer

---

[2] This section draws heavily on Hartman (1976b)

is that they will disappear from (46), since subtraction will yield a value of zero. It is this way of specifying the utility function which forces the principle of "independence of relevant alternatives". The model imposes a function which evaluates a mode only according to its own characteristics and the characteristics of the individual who is making a choice. The hedonic specification does not simultaneously incorporate the value taken on by the same attribute in another mode. Once this way of looking at the world is imposed, prediction bias becomes an inescapable companion of share calculations based on binary comparisons.

Hartman's generalization of the multinomial logit attempts to circumvent the problem by fundamentally respecifying the underlying choice model. The Hartman model incorporates two methodological innovations. First, the valuation function is respecified so that the valuation of a particular mode is sensitive both to its own measured value of a particular attribute and to the values of that same attribute which characterize other modes. Thus, suppose that three modes are being considered. We could generalize the valuation function as:

$$(47) \quad U_{ij} = \beta_0 + \beta_1 Z_1 [W^1_{i1}, W^1_{i2}, W^1_{i3}] + \ldots + \beta_T Z_T [W^T_{i1}, W^T_{i2}, W^T_{i3}] + \varepsilon_{ij}$$

At first glance, this formulation of the utility model appears to contain a paradox. If the Z-functions are all the same, and all modal attributes are in every utility function, then all the modal valuations are identical and the whole logit approach collapses. This obviously doesn't occur, and the apparent paradox has been introduced only to demonstrate the inevitability of another change in specification once all system variables have been put into each valuation equation. This is a change in the β's. In Hartman's generalization, the valuation weights are assumed to vary across modes. Thus, the appropriate specification is:

$$(48)\quad V_{ij} = \beta_0^j + \beta_1^j Z_1(W_i^1) + \ldots + \beta_T^j Z_T(W_i^T) + \varepsilon_{ij} = Z\beta^j + \varepsilon_{ij}$$

This re-casting makes it clear that we can no longer think of (48) as a utility function in the same sense that this idea applies to a formulation like (47). Rather, we should think of (48) as a prediction model which is in some ways analagous to a consumer demand function for a continuous (non-discrete-choice) case. However, demand functions in the usual case are derived from a stable, underlying utility function in which continuous trade-offs are possible. In the Hartman model, on the other hand, the implicit utility function has parameters which are themselves random variables within the relevant population. Thus, particular modes are chosen because different subgroups of the population apply different value-weights to observable modal attributes and are themselves characterized by different kinds of choice behavior, even when identical personal attributes are observed.

Once the inherently predictive character of the generalized multinomial logit is accepted, it can be quite useful for policy analysis. To see how, we have only to substitute the valuation function (48) into the Weibull-based modal comparison system which has already been developed.

$$(49)\quad \text{Prob[K is chosen]} = \frac{e^{V_K}}{\sum_{i=1}^{N} e^{V_j}} = \frac{e^{Z\beta^K}}{\sum_{j=1}^{N} e^{Z\beta j}} = \frac{1}{\sum_{j=1}^{N} e^{Z(\beta^j - \beta^K)}} = \frac{1}{\sum_{j=1}^{N} e^{Z\gamma_{jK}}}$$

Since we know Z, the matrix of multiple attribute values, we can estimate the values of the parameters $\gamma_{jK}$ as long as we have sufficient degrees of freedom. Since each $\gamma_{jK}$ is equal to a difference in underlying valuation parameters $(\beta^j - \beta^K)$, we can not deduce the parameter values themselves from our results. The model is primarily useful for prediction, however, so this indeterminancy presents no real problems. Recall that all attribute values

for all modes are in each modal choice equation, so that predicting the consequences of an attribute shift for shares allocation is straightforward.

The result of maximum likelihood estimation in this case will be a set of estimated coefficients which characterize differences in responsiveness across modes with respect to each common attribute. There is no arbitrary constraint on cross-elasticities, so that the generalized multinomial logit can come closer than the standard logit to an unbiased prediction of share adjustments to price changes.

While the generalized multinomial logit approach represents a clear improvement for policy analysis within existing choice environments, it does not adapt well to the introduction of new modes. To see why, we only have to glance at (48) and (49). The $\gamma_{jk}$'s which are estimated are specific to particular existing modes. Obviously, no such coefficients exist for a new mode, and we therefore have no way of predicting its share. In order to handle this difficulty, we have to turn to the second and more complex approach -- the covariance probit.

## VII. Covariance Probit Estimation[3]

As previously noted, the generalized multinomial logit (GML) approach provides a useful alternative to conventional logit models when existing systems are being analyzed because it allows all modal alternatives to have an impact on binary comparisons. The parameters of GML equations are more specific, however, so that this approach cannot be used for forecasting demand diversion to new modes. The logit specification appears to be inherently incapable of yielding satisfactory predictions when new modes are being considered.

Thus, we find ourselves confronted with a difficult choice. If we opt for Weibull-distributed errors, we become capable of modeling choice when many modal alternatives are available but incapable of accurate demand forecasting when a new mode is proposed. Abandonment of the Weibull distribution and the accompanying logistic probability function however, threatens to undermine our ability to estimate modal choice models at all. In the search for an unbiased prediction methodology, we must now return to our original specification of utility and examine our options more closely. The analysis will be developed using three existing modes as the basis for valuation. When a better approach to forecasting has been developed, consideration will be given to the demand for one additional mode.

Suppose that three modes are currently available to individuals in a particular context, and that the basis for valuation is an additive utility function whose two arguments are $X_1$ and $X_2$. For any "representative" individual in the population, the familiar approach to utility modeling leads to the

---

[3]The following sections represent a very detailed re-working of Hausman (1978)

following three valuation functions:

$$U_1 = \beta_1 X_{11} + \beta_2 X_{12} + \varepsilon_1$$

$$(50)\quad U_2 = \beta_1 X_{21} + \beta_2 X_{22} + \varepsilon_2$$

$$U_3 = \beta_1 X_{31} + \beta_2 X_{32} + \varepsilon_3$$

Notice that all three alternatives are being evaluated with the same utility function, so that the parameters ($\beta_1$, $\beta_2$) are assumed to be fixed across modes. As previously mentioned, this approach is based on the notion of a "representative individual" who typifies the decision-makers in a particular population. At the same time, we have recognized the existence of a random error which characterizes the degree to which individual valuations differ from the typical valuation. There are at least two different ways to rationalize the phenonemon which this error term describes. In the preceding sections, the error term was taken to represent the impact of unobservable individual and modal characteristics. With equal plausibility, however, we could re-specify the error by revising the utility model.

Some reflection about the nature of personal valuations leads to the conclusion that the "representative individual" in a population does not really exist. Rather, the parameters of any general utility function can best be thought of as central tendencies in distribution of individual valuations. Thus, we could plausibly re-specify (50) as:

$$U_{i1} = (\bar{\beta}_1 + \beta_{1i})X_{11} + (\bar{\beta}_2 + \beta_{2i})X_{12} + v_{i1}$$

$$(51)\quad U_{i2} = (\bar{\beta}_1 + \beta_{1i})X_{21} + (\bar{\beta}_2 + \beta_{2i})X_{22} + v_{i2}$$

$$U_{i3} = (\bar{\beta}_1 + \beta_{1i})X_{31} + (\bar{\beta}_2 + \beta_{2i})X_{32} + v_{i3}$$

Here, $\bar{\beta}_1$ and $\bar{\beta}_2$ could be thought of as the mean values of the marginal valuations present in the population. Similarly, $\beta_{1i}$ and $\beta_{2i}$ could be thought

of as the deviations form the norm which characterize any individual i in the population. If we adopt the assumption that these deviations are normally-distributed, then we know that

$$\beta_1 \sim N(0, \sigma^2_{\beta_1}) \text{ and } \beta_2 \sim N(0, \sigma^2_{\beta_2})$$

Since it is clear that individual valuations do differ considerably in a population, this way of approaching the problem seems to be quite plausible. Undoubtedly, the problem of unobservable modal and personal attributes will continue to exist, however, so that it also makes sense to continue to add an error term $v_j$, which will be assumed to be normally-distributed. Thus,

$$v_j \sim N(0, \sigma^2_{v_j})$$

We can re-write (51) generally as

$$U_1 = \bar{\beta}_1 X_{11} + \bar{\beta}_2 X_{12} + (\beta_1 X_{11} + \beta_2 X_{12} + v_1)$$

$$(52) \quad U_2 = \bar{\beta}_1 X_{21} + \bar{\beta}_2 X_{22} + (\beta_1 X_{21} + \beta_2 X_{22} + v_2)$$

$$U_3 = \bar{\beta}_1 X_{31} + \bar{\beta}_2 X_{32} + (\beta_1 X_{31} + \beta_2 X_{32} + v_3)$$

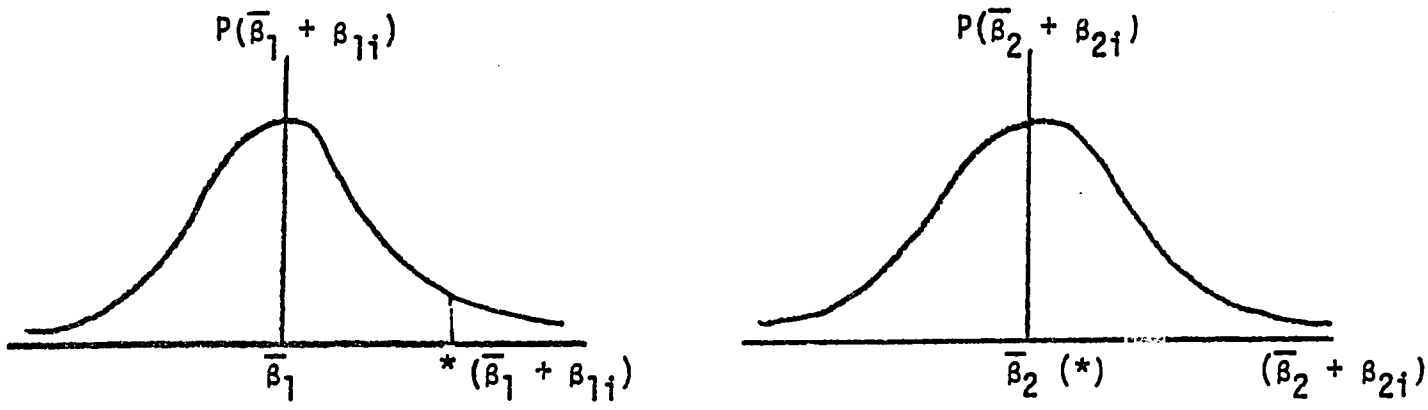
Notice that if $\beta_1$, $\beta_2$, and $v_j$ are all normally-distributed, then so is their weighted sum $(\beta_1 X_{ji} + \beta_2 X_{j2} + v_j)$. This weighted sum is equivalent to $\varepsilon_j$, the error term in the first model of valuation. In our treatment of the errors $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ to this point, we have always assumed that they are independently distributed. It is not time to see whether some alternative specification is not more appropriate.

It seems desirable to begin with a brief discussion of the concept of independent variation in this context. We can characterize all the patterns of individual and joint variation for the errors in our 3-equation model using the following covariance matrix:

$$(53)\quad V(\varepsilon) = \begin{bmatrix} Var(\varepsilon_1) & Cov(\varepsilon_1\varepsilon_2) & Cov(\varepsilon_1\varepsilon_3) \\ Cov(\varepsilon_2\varepsilon_1) & Var(\varepsilon_2) & Cov(\varepsilon_2\varepsilon_3) \\ Cov(\varepsilon_3\varepsilon_1) & Cov(\varepsilon_3\varepsilon_2) & Var(\varepsilon_3) \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

We know that the variance of a normally distributed random variable is a measure of the degree to which its values are dispersed around the mean value. The covariance of two normally-distributed random variables, on the other hand, is a measure of their tendency to vary together. In the utility model which we employed in the preceding sections, the assumption of error independence was maintained. For (53), this is equivalent to assuming that all error covariances are zero.

In the "variable coefficients" model which has just been specified, however, the error covariances will not be zero. The structure of the model makes it relatively easy to understand why this should be so. Suppose some individual places a much higher marginal valuation on attribute $X_1$ than is characteristic of the population, while his marginal valuation of $X_2$ is at the norm. Thus, for the individual in question, we have:




If we know the true values for the typical valuation parameters $\beta_1$ and $\beta_2$ then the systematic valuation of each choice by the "representative individual" is:

$$\hat{U}_1 = \hat{\bar{\beta}}_1 X_{11} + \hat{\bar{\beta}}_2 X_{12} + \varepsilon_1$$

$$(54) \quad \hat{U}_2 = \hat{\bar{\beta}}_1 X_{21} + \hat{\bar{\beta}}_2 X_{22} + \hat{\varepsilon}_2$$

$$\hat{U}_3 = \hat{\bar{\beta}}_1 X_{31} + \hat{\bar{\beta}}_2 X_{32} + \hat{\varepsilon}_3$$

As an illustration, we will suppose that for modes 1 and 2 the value of $X_1$ is very large and that of $X_2$ negligible, while the converse is true for mode 3. Since our individual has a $\beta_1$ value which is much higher than the norm, his valuations of modes 1 and 2 will both be above the norm (and for mode 3, his valuation will be below the norm). Thus, both $\varepsilon_1$ and $\varepsilon_2$ for our individual will be large and positive. For some other individual with a substantially below-normal marginal valuation of $X_1$, the converse would be true -- $\hat{\varepsilon}_1$ and $\hat{\varepsilon}_2$ would be large and negative.

This outcome may be understood more fully with the aid of a few numbers. Let the typical valuation equation be as follows:

$$\hat{U}_1 = (3 + \beta_1)X_{11} + (-2 + \beta_2)X_{12} + \hat{v}_1$$

$$(55) \quad \hat{U}_2 = (3 + \beta_1)X_{21} + (-2 + \beta_2)X_{22} + \hat{v}_2$$

$$\hat{U}_3 = (3 + \beta_1)X_{31} + (-2 + \beta_2)X_{32} + \hat{v}_3$$

For our individual, the marginal valuation of $X_1$ is substantially above the norm ($\beta_1 = 2$) while the valuation of $X_2$ is at the norm ($\beta_2 = 0$). For the "representative individual" and our individual, the valuation equations are as follows:

Typical Individual

$$\hat{U}_1 = 3X_{11} - 2X_{12}$$

(56) $$\hat{U}_2 = 3X_{21} - 2X_{22}$$

$$\hat{U}_3 = 3X_{31} - 2X_{32}$$

Our Individual

$$\hat{U}_{1(1)} = 5X_{11} - 2X_{12} + v_{1(1)}$$

(57) $$\hat{U}_{2(1)} = 5X_{21} - 2X_{22} + v_{2(1)}$$

$$\hat{U}_{3(1)} = 5X_{31} - 2X_{32} + v_{3(1)}$$

Now, let us suppose that modes 1 and 2 have the same measured level for attribute 1($X_{11} = X_{21} = 4$) and that its level is much lower for mode 3 ($X_{31} = 0$). We now have sufficient information to calculate the values of $\hat{\varepsilon}_1$, $\hat{\varepsilon}_2$, and $\hat{\varepsilon}$ for our individual.

$$\hat{\varepsilon}_{1(1)} = \hat{U}_{1(1)} - \hat{U}_1 = 5X_{11} - 2X_{12} + v_{1(1)} - 3X_{11} + 2X_{12} = 2X_{11} + v_{1(1)} = 8 + v_{1(1)}$$

(58) $$\hat{\varepsilon}_{2(1)} = \hat{U}_{2(1)} - \hat{U}_2 = 5X_{21} - 2X_{22} + v_{2(1)} - 3X_{21} + 2X_{22} = 8 + v_{2(1)}$$

$$\hat{\varepsilon}_{3(1)} = U_{3(1)} - U_3 = 5X_{31} - 2X_{32} + v_{3(1)} - 3X_{31} + 2X_{32} = 0 + v_{3(1)}$$

Thus, we can see that modes 1 and 2 have the same large, positive error component (8). The component is there for two reasons. First, the individual in question has a preference for attribute 1 which is substantially above the norm. At the same time, modes 1 and 2 have an important common characteristic ($X_{11} = X_{31} = 4$) which makes them similar from the perspective of the individual. Hopefully, the general implication of this illustration is clear. Because individual deviations from normal valuations will be the same across equations, the error terms for modes which are similar will have substantial positive covariances.

Large negative covariances are obviously also possible when modal attributes are negatively rather than positively correlated. Only when modal attributes are effectively orthogonal (uncorrelated) would we expect error covariances to approach zero and the assumption of independence to be approximately sustainable.

Now that error dependence in the variable coefficients model (or random coefficients model, as it is commonly called) has been established intuitively, it should be relatively easy to follow the mathematical development of the same result. We have three statistical assumptions to go on at this point:

$$\beta_1 \sim N(0,\sigma^2_{\beta_1}) \qquad \beta_2 \sim N(0,\sigma^2_{\beta_2}) \qquad v_i \sim N(0,\sigma^2_{v_i})$$

For the three modes and two attributes which we have been employing in our example, these assumptions generate the following covariance matrices:

$$v(\beta) = \begin{bmatrix} \sigma^2_{\beta_1} & 0 \\ 0 & \sigma^2_{\beta_1} \end{bmatrix} \qquad V(v) = \begin{bmatrix} \sigma^2_{v_1} & 0 & 0 \\ 0 & \sigma^2_{v_2} & 0 \\ 0 & 0 & \sigma^2_{v_3} \end{bmatrix}$$

In addition, we impose the assumption that $\beta_1$, $\beta_2$, and $v_i$'s are independently distributed, so that the mathematical expectation of any of their cross-products is zero (e.g., $E(\beta_1 V_1) = E(\beta_2 V_1) = 0$). With the aid of a little matrix notation we can now establish the covariance matrix $V(\varepsilon)$ which is of interest.

$$\text{Let} \quad X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ X_{31} & X_{32} \end{bmatrix} \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} \qquad v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

From (52) we know that

$$\varepsilon_1 = \beta_1 X_{11} + \beta_2 X_{12} + v_1$$

$$(59) \quad \varepsilon_2 = \beta_1 X_{21} + \beta_2 X_{22} + v_2$$

$$\varepsilon_3 = \beta_1 X_{31} + \beta_2 X_{32} + v_3$$

In matrix form, this same three-equation set can be written as

$$(60) \quad \varepsilon = X\beta + v$$

It is easy to establish that the mathematical expectation of $\varepsilon$ is zero:

$$E(\varepsilon) = E[X\beta + v] = E(X\beta) + E(v) = XE(\beta) + E(v) = X(0) + 0 = 0$$

This in turn makes it relatively easy to determine $V(\varepsilon)$, the full error covariance matrix:

$$(61) \quad V(\varepsilon) = E[\varepsilon - E(\varepsilon)][\varepsilon - E(\varepsilon)]' = E(\varepsilon - 0)(\varepsilon - 0)' = E(\varepsilon\varepsilon')$$

$$= E[X\beta + v][X\beta + v]' = E[X\beta + v][\beta'X' + v']$$

$$= E[X\beta\beta'X' + X\beta v' + v\beta'X' + vv']$$

$$= E(X\beta\beta'X') + E(X\beta v') + E(v\beta'X') + E(vv')$$

$$= XE(\beta\beta')X' + 0 + 0 + E(vv')$$

$$= XV(\beta)X' + V(v)$$

The interior expectation items in the fourth line of the deviation are zero because we have imposed the assumption (which seems quite reasonable) that the $\beta$'s and V's are distributed independently of one another.

We can write out the expression for $V(\varepsilon)$ in this case by substituting from the original matrix defininitions:

$$(62) \quad V(\varepsilon) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ X_{31} & X_{32} \end{bmatrix} \begin{bmatrix} \sigma^2_{\beta_1} & 0 \\ 0 & \sigma^2_{\beta_1} \end{bmatrix} \begin{bmatrix} X_{11} & X_{21} & X_{31} \\ X_{12} & X_{22} & X_{32} \end{bmatrix} + \begin{bmatrix} \sigma^2_{v_1} & 0 & 0 \\ 0 & \sigma^2_{v_2} & 0 \\ 0 & 0 & \sigma^2_{v_3} \end{bmatrix}$$

Thus, given a fixed parameter variance ($\sigma^2_{\beta_1}$, $\sigma^2_{\beta_2}$) in the population, it is the relative sizes of particular modal attributes which determine the covariance matrix. Here we have been working with two modal attributes. However it is easy to expand (61) to the case of K attributes and N modes:

$$V(\varepsilon) = \begin{bmatrix} X_{11} & \cdots & X_{1K} \\ \vdots & & \vdots \\ X_{N1} & \cdots & X_{NK} \end{bmatrix} \begin{bmatrix} \sigma^2_{\beta_1} & & 0 \\ & \ddots & \\ 0 & & \sigma^2_{\beta_K} \end{bmatrix} \begin{bmatrix} X_{11} & \cdots & X_{N1} \\ \vdots & & \vdots \\ X_{1K} & \cdots & X_{NK} \end{bmatrix} + \begin{bmatrix} \sigma^2_{v_1} & & 0 \\ & \ddots & \\ 0 & & \sigma^2_{v} \end{bmatrix}$$

$$= X\Sigma_\beta X' + \Sigma_N$$

We now have a completely general formulation of the error covariance matrix, under assumptions about the random coefficients model which have already been specified. This completes part of the story which must be told in order to solve our problem. Obviously, however, the tale is far from completed. Now that we have arrived at a characterization of the full error covariance matrix for our valuation model, how can we use it to produce unbiased predictions of demand diversion when a new mode is introduced?

## VIII. Modeling with Multivariate Normal Functions

It is at this juncture that we need to step back and draw some parallels between the current problem and the one which we have already solved in the case of the multinomial logit. Recall that in the case of the logit, a tautological statement of utility-superiority and modal choice led to a fundamental stochastic inequality in which the relevant error term was the difference between error terms in the valuation function. Thus, $[Z_1 - Z_2)\beta > \eta_1(\varepsilon_2 - \varepsilon_1)]$ is the fundamental inequality for modes 1 and 2. Under the assumption that $\varepsilon_1$ and $\varepsilon_2$ are Weibull-distributed we are led to the logistic distribution as an appropriate measurement of cumulative probability.

In this case, we have abandoned the Weibull distribution for the normal distribution, but much of the story must remain the same. We must therefore present a brief development of the normal-distribution-based choice model.

In the present case we have (in matrix form):

(63) $u = X\bar{\beta} + \varepsilon$

where $\varepsilon = X\beta + v$

$E(\varepsilon) = 0$

$V(\varepsilon) = X\Sigma_\beta X' + \Sigma_v$

As we have seen previously, the choice of a particular mode will depend upon whether its utility-superiority over other modes outweighs any countervailing random error effects for a particular individual. Thus, for example, we know that mode 1 will be chosen in preference to mode 2 if $U_1 > U_2$, which implies that $U_1 - U_2 > \varepsilon_2 - \varepsilon_1$. Since we have our characterization of the valuation function and the error term, we can employ them to express the inequality above as:

(64) $\bar{\beta}_1 X_{11} + \bar{\beta}_2 X_{12} - \bar{\beta}_1 X_{21} - \bar{\beta}_2 X_{22} > \varepsilon_2 - \varepsilon_1 (=\eta_1)$

$\bar{\beta}_1(X_{11} - X_{21}) + \bar{\beta}_2(X_{12} - X_{22}) > \eta_1$

212

Since both $\varepsilon_1$ and $\varepsilon_2$ are normally-distributed random variables with fixed means and variances, so is $\eta_1$. In fact, it is relatively easy to calculate the two moments for $\eta_1$:

If $w_1' = [-1\ 1\ 0]$, then $\eta_1 = w_1'\varepsilon$

$$E(\eta_1) = E(w_1'\varepsilon) = w_1'E(\varepsilon) = 0$$

$$V(\eta_1) = V(w_1'\varepsilon) = E(w_1'\varepsilon)(\varepsilon' w_1) = w_1'V(\varepsilon)w_1$$

Similarly, if $w_2' = [-1\ 0\ 1]$, then $\eta_2 = \varepsilon_3 - \varepsilon_1 = w_2'\varepsilon$

$$E(\eta_2) = w_2'E(\varepsilon) = 0$$

$$V(\eta_2) = w_2'V(\varepsilon)w_2$$

and $\mathrm{Cov}(\eta_1\eta_2) = E(w_1'\varepsilon)(\varepsilon' w_2) = w_1'V(\varepsilon)w_2$

Thus, rather than giving direct consideration to the covariance matrix $V(\varepsilon)$, we can consider the covariance matrix of the $\eta$'s, which are normally-distributed variables found from differences in the $\varepsilon$'s:

$$(65)\quad V(\eta) = \begin{bmatrix} w_1'V(\varepsilon)w_1 & w_1'V(\varepsilon)w_2 \\ w_2'V(\varepsilon)w_1 & w_2'V(\varepsilon)w_2 \end{bmatrix} = w'V(\varepsilon)w$$

where $w = [w_1\ w_2]$

For the sake of brevity, we will now use the symbol "$\gamma$" to refer to variances and covariances in this context.

$$(66)\quad V(\eta) = w'V(\varepsilon)w = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}$$

Having reduced a three-error case ($\varepsilon_1$, $\varepsilon_2$, $\varepsilon_3$) to a two-error case ($\eta_1$, $\eta_2$) we can pass to a consideration of the multivariate normal distribution. Since we are considering a case with two normally-distributed variables $\eta_1$ and $\eta_2$, we will give specific attention to the bivariate normal distribution. We will

continue to use V to represent the associated covariance matrix. Then the technical definition of the probability density function (p.d.f.) characterizing the joint normal distribution of several random variables is:

$$(67)\quad f(X) = \frac{1}{(2\pi)^{n/2}\sqrt{|V|}}\, e^{\frac{[-(X-\mu)'V^{-1}(X-\mu)]}{2}}$$

where X = an $n$-row vector of random (normally-distributed) variables.

$\mu$ = an $n$-row vector of the means (expected values) of those variables.

This rather complex-looking formula becomes more comprehensible if we examine it using a simple example. We will use our two normal variables $n_1$ and $n_2$, and assume that they are independent of one another. Thus, under the assumption that

$$(68)\quad V(n) = \begin{bmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \end{bmatrix}$$

we have

$$(69)\quad F\binom{n_1}{n_2} = \frac{1}{(2\pi)^{2/2}\sqrt{\begin{vmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \end{vmatrix}}}\, e^{-1/2\left([(n_1-0)'(n_2-0)']\begin{bmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \end{bmatrix}^{-1}\begin{bmatrix} n_1-0 \\ n_2-0 \end{bmatrix}\right)}$$

$$= \frac{1}{2\pi\sqrt{\gamma_{11}\,\gamma_{22}}}\, e^{-1/2[n_1'\ n_2']\begin{bmatrix} 1/\gamma_{11} & 0 \\ 0 & 1/\gamma_{22} \end{bmatrix}\begin{bmatrix} n_1 \\ n_2 \end{bmatrix}}$$

$$= \frac{1}{2\pi\sqrt{\gamma_{11}\,\gamma_{22}}}\, e^{-1/2[1/\gamma_{11}n_1'n_1 + 1/\gamma_{22}n_2'\ n_2]}$$

$$= \frac{1}{\sqrt{2\pi\,\gamma_{11}}}\, e^{-1/2\eta_1'\eta_1/\gamma_{11}} \cdot \frac{1}{\sqrt{2\pi\,\gamma_{22}}}\, e^{-1/2\eta_2'\eta_2/\gamma_{22}} = f(\eta_1)\cdot f(\eta_2)$$

Here, independence gives us a diagonal covariance matrix, and after a few mathematical manipulations we arrive at what we know should be the result. The probability of joint occurrence of any two values of $\eta_1$ and $\eta_2$ is the product of their individual probabilities, where both of these individual probabilities are generated by the standard equation for a single normally-distributed variable. In other words, we have the familiar case:

(70) $P(AB) = P(A)\ P(B)$ where A and B are independent draws

In the case of our choice model, of course, the problem is that there are very good reasons for supposing that the random error terms are <u>not</u> independent. The general statement of joint probability which is analogous to (70) is:

(71) $P(AB) = P(A)\ P(B|A) = P(B)\ P(A|B)$

In this more general statement of probability, room is left for the existence of conditional dependence between A and B(i.e., knowing that A has taken on a particular value will be helpful in predicting the value of B and conversely). The covariance matrix in the p.d.f. for the multivariate normal distribution specifies this dependency exactly. Separately, the random variables in the vector X are normally-distributed (e.g., for particular fixed values of $X_2,\ldots,X_N$, the values taken on by $X_1$ are normally distributed). Some dependence exists among them, however, so that the probability of occurrence for any particular value of one variable will be affected by the values taken on by the other variables in the joint distribution.

Having said all this, we can now go back and examine the bivariate normal distribution for the case where $\eta_1$ and $\eta_2$ are not independent of one another.

215

Any multivariate distribution is sufficiently complex to cause problems for maximum likelihood estimation because multiple integrals are involved. We are therefore going to examine the model intensively by moving through a series of handy simplifications. The notation will be changed slightly for expositional convenience:

$$(72) \quad V(\eta) = \begin{bmatrix} \gamma_1^2 & \gamma_{12} \\ \gamma_{21} & \gamma_2^2 \end{bmatrix}$$

We can now introduce the first substantial simplification which is possible in this two-variable case. For any two variables, $X_1$ and $X_2$, we know that the coefficient of correlation between $X_1$ and $X_2$ is defined as:

$$\rho = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)Var(X_2)}}$$

In our case, the correlation coefficient for $\eta_1$ and $\eta_2$ can be written as:

$$\rho = \frac{\gamma_{12}}{\sqrt{\gamma_1^2 \gamma_2^2}} = \frac{\gamma_{12}}{\gamma_1 \gamma_2} = \frac{\gamma_{21}}{\gamma_1 \gamma_2}$$

or $\gamma_{21} = \gamma_{12} = \rho\gamma_1\gamma_2 = \rho\gamma_2\gamma_1$

Thus, we can re-express the covariance matrix as:

$$(73) \quad V(\eta) = \begin{bmatrix} \gamma_1^2 & \gamma_{12} \\ \gamma_{21} & \gamma_2^2 \end{bmatrix} = \begin{bmatrix} \gamma_1^2 & \rho\gamma_1\gamma_2 \\ \rho\gamma_2\gamma_1 & \gamma_2^2 \end{bmatrix}$$

Recall that in the bivariate normal distribution the covariance matrix V makes two appearances. We can therefore simplify the full probability density function expression by solving as follows:

$$|V| = \begin{vmatrix} \gamma_1^2 & \rho\gamma_1\gamma_2 \\ \rho\gamma_2\gamma_1 & \gamma_2^2 \end{vmatrix} = \gamma_1^2\gamma_2^2 - \rho^2\gamma_1^2\gamma_2^2 = \gamma_1^2\gamma_2^2(1-\rho^2)$$

The inverse of V is also easy to calculate in this simple two-by-two case:

$$\left( \text{(Recall: For } A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad A^{-1} = \frac{1}{|A|}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \right)$$

$$V^{-1} = \frac{1}{|V|}\begin{bmatrix} V_{22} & -V_{12} \\ -V_{21} & V_{11} \end{bmatrix} = \frac{1}{\gamma_1^2\gamma_2^2(1-\rho^2)} \cdot \begin{bmatrix} \gamma_2^2 & -\rho\gamma_1\gamma_2 \\ -\rho\gamma_2\gamma_1 & \gamma_1^2 \end{bmatrix}$$

$$= \frac{1}{1-\rho^2}\begin{bmatrix} \frac{1}{\gamma_1^2} & \frac{-\rho}{\gamma_1\gamma_2} \\ \frac{-\rho}{\gamma_2\gamma_1} & \frac{1}{\gamma_2^2} \end{bmatrix}$$

This leads us to:

$$(74)\quad f\binom{\eta_1}{\eta_2} = \frac{1}{\sqrt{\gamma_1^2\gamma_2^2(1-\rho^2)}}\, e^{-\frac{1}{2(1-\rho^2)}\left[[\eta_1'\eta_2']\begin{bmatrix} 1/\gamma_1^2 & \frac{-\rho}{\gamma_1\gamma_2} \\ \frac{-\rho}{\gamma_2\gamma_1} & 1/\gamma_2^2 \end{bmatrix}\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}\right]}$$

217

At first glance this doesn't seem like much help. After a small digression, however, we will see that this way of formulating the p.d.f. is in fact extremely useful.

We can begin to solve our problem by converting $\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}$ to its "standard normal" form:

$$\text{If } \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{bmatrix} \gamma_1^2 & \rho\gamma_1\gamma_2 \\ \rho\gamma_2\gamma_1 & \gamma_1^2 \end{bmatrix}\right]$$

We can establish easily that

$$(75) \quad \begin{bmatrix} \eta_1/\gamma_1 \\ \eta_2/\gamma_2 \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right]$$

[This is because $\eta_1/\gamma_1 \sim N(0,1)$ and $\eta_2/\gamma_2 \sim N(0,1)$, so that

$$V = \begin{bmatrix} (1)^2 & \rho(1)(1) \\ \rho(1\eta 1) & (1)^2 \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

]

$$\text{Thus, } V^{-1} = \frac{1}{|V|}\begin{bmatrix} V_{22} & -V_{12} \\ -V_{21} & V_{11} \end{bmatrix} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}$$

Given this result, it is easy to set up the joint p.d.f. for the "standard normal" vector $\begin{pmatrix} \frac{\eta_1}{\gamma_1} \\ \frac{\eta_2}{\gamma_2} \end{pmatrix}$ :

2-18

$$(76)\quad F\begin{pmatrix}\frac{\eta_1}{\gamma_1}\\ \frac{\eta_2}{\gamma_2}\end{pmatrix} = \frac{1}{2\pi\sqrt{|V|}}\, e^{-\frac{1}{2}\left[\frac{\eta_1'}{\gamma_1}\ \frac{\eta_2'}{\gamma_2}\right] V^{-1} \begin{pmatrix}\frac{\eta_1}{\gamma_1}\\ \frac{\eta_2}{\gamma_2}\end{pmatrix}}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}\left[\frac{\eta_1'}{\gamma_1}\ \frac{\eta_2'}{\gamma_2}\right] \frac{1}{1-\rho^2}\begin{bmatrix}1 & -\rho\\ -\rho & 1\end{bmatrix}\begin{pmatrix}\frac{\eta_2}{\gamma_2}\\ \frac{\eta_2}{\gamma_2}\end{pmatrix}}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{\frac{1}{2(1-\rho^2)}\left[\frac{\eta_1'\eta_1}{\gamma_1^2} - \frac{2\rho\eta_1'\eta_2}{\gamma_1\gamma_2} + \frac{\eta_2'\eta_2}{\gamma_2^2}\right]}$$

Thus, we have established the joint p.d.f. for two standard normal variables $\frac{\eta_1}{\gamma_1}$, $\frac{\eta_2}{\gamma_2}$ whose covariance is given by $\rho$. An additional mathematical trick, an appropriate "transformation of variables," must be introduced at this point. Suppose that we define a new variable, $r_2$, such that

$$r_2 = \frac{\eta_2}{\gamma_2} - \rho\frac{\eta_1}{\gamma_1}$$

It is relatively easy to establish that $F(r_2) = f\left[\frac{\eta_2}{\gamma_2} \,\middle|\, \frac{\eta_1}{\gamma_1}\right]$

Recall that $f\left(\frac{\eta_1}{\gamma_1}\frac{\eta_2}{\gamma_2}\right) = f\left(\frac{\eta_1}{\gamma_1}\right) f\left(\frac{\eta_2}{\gamma_2}\middle|\frac{\eta_1}{\gamma_1}\right)$ must always be true. Thus, demonstrating that $f\left(\frac{\eta_1}{\gamma_1}\frac{\eta_2}{\gamma_2}\right) = f\left(\frac{\eta_1}{\gamma_1}\right) f(r_2)$ is equivalent to demonstrating that

$$f(r_2) = f\left(\frac{\eta_2}{\gamma_2}\middle|\frac{\eta_1}{\gamma_1}\right).$$

The demonstration is as follows:

$$\frac{\eta_1}{\gamma_1} \sim N(0,1) \Longrightarrow F\left(\frac{\eta_1}{\gamma_1}\right) = \frac{1}{\sqrt{2\pi}} e^{\frac{\eta_1'\eta_1}{\gamma_1^2}}$$

For $r_2$, we have

$$E(r_2) = E\left(\frac{\eta_2}{\gamma_2} - \rho\frac{\eta_1}{\gamma_1}\right) = E\left(\frac{\eta_2}{\gamma_2}\right) - \rho E\left(\frac{\eta_1}{\gamma_1}\right) = 0$$

$$V(r_2) = E\left[\left(\frac{\eta_2}{\gamma_2} - \rho\frac{\eta_1}{\gamma_1}\right) - E\left(\frac{\eta_2}{\gamma_2} - \rho\frac{\eta_1}{\gamma_1}\right)\right]' \left[\left(\frac{\eta_2}{\gamma_2} - \rho\frac{\eta_1}{\gamma_1}\right) - E\left(\frac{\eta_2}{\gamma_2} - \rho\frac{\eta_1}{\gamma_1}\right)\right]$$

$$= E\left[\frac{\eta_2}{\gamma_2} - \rho\frac{\eta_1}{\gamma_1}\right]' \left[\frac{\eta_2}{\gamma_2} - \rho\frac{\eta_1}{\gamma_1}\right]$$

$$= E\left[\frac{\eta_2'\eta_2}{\gamma_2^2} - 2\rho\frac{\eta_1\eta_2}{\gamma_1\gamma_2} + \rho^2\frac{\eta_1'\eta_1}{\gamma_1^2}\right]$$

$$= E\left(\frac{\eta_2'\eta_2}{\gamma_2^2}\right) - 2\rho E\left(\frac{\eta_1\eta_2}{\gamma_1\gamma_2}\right) + \rho^2 E\left(\frac{\eta_1'\eta_1}{\gamma_1^2}\right)$$

Since $V(\eta_2) = E(\eta_2'\eta_2) = \gamma_2^2$, and $E(\eta_1\eta_2) = \mathrm{Cov}\ (\eta_1,\eta_2) = \rho\gamma_1\gamma_2$,

the above expression is equivalent to:

$$V(r_2) = 1 - 2\rho^2 + \rho^2 = 1-\rho^2$$

Therefore $r_2 \sim N(0,1-\rho^2)$

and $$f(r_2) = \frac{1}{\sqrt{2\Pi(1-\rho^2)}}\, e^{-\frac{1}{2}\frac{r_2'r_2}{1-\rho^2}}$$

It follows that

$$F\ \frac{\eta_1}{\gamma_1}\ f(r_2) = \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}\left[\frac{\eta_1'\eta_1}{\gamma_1^2} + \frac{r_2'r_2}{1-\rho^2}\right]}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}\left[\frac{\frac{\eta_1'\eta_1}{\gamma_1^2}(1-\rho^2) + r_2'r_2}{1-\rho^2}\right]}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}\left[\frac{\frac{\eta_1'\eta_1}{\gamma_1^2} - \rho^2\frac{\eta_1'\eta_1}{\gamma_1^2} + \frac{\eta_2'\eta_2}{\gamma_2^2} - 2\rho\frac{\eta_1\eta_2}{\gamma_1\gamma_2} + \rho^2\frac{\eta_1'\eta_1}{\gamma_1^2}}{1-\rho^2}\right]}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left[\frac{\eta_1'\eta_1}{\gamma_1^2} - 2\rho\frac{\eta_1\eta_2}{\gamma_1\gamma_2} + \frac{\eta_2'\eta_2}{\gamma_2^2}\right]}$$

$$= F\left(\frac{\eta_1}{\gamma_1}\ \frac{\eta_2}{\gamma_2}\right)$$

22-1

Thus, we can write our bivariate normal function either as $F\left[\frac{\eta_1}{\gamma_1}\ \frac{\eta_2}{\gamma_2}\right]$ or as $F\left[\frac{\eta_1}{\gamma_1}\right] \cdot F\left[\frac{\eta_2}{\gamma_2} \cdot \rho\frac{\eta_1}{\gamma_1}\right]$. However, it quickly becomes apparent that the latter form has one very strong advantage. It greatly facilitates the combination of integration with maximum likelihood procedures.

222

## IX. Maximum Likelihood Estimation of Covariance Probit Models

Recall that we began this entire exercise with an attempt to characterize modal choice probabilities in a meaningful way. The fundamental underlying notion employed was that of cumulative probability: Given a mathematical characterization of modal valuation up to some determinate error factor, the cumulative probability that a particular mode will be chosen over some alternative mode can be systematically related to the difference in valued characteristics between the two modes. As we saw in the logistic case, much of the analysis rests on the specification of the stochastic properties of the errors in the valuation model. Once specific probability density functions (p.d.f.'s) have been imposed, the link between cumulative choice probability and differences in valued characteristics is established through the application of integral calculus. The development of binary and multinomial logistic models of choice provided one example of this process.

Now we are ready to have a second look, this time in the context of normally-distributed errors. A major portion of the preceding Section has been devoted to the analysis and simplification of the multivariate normal density function which results when errors in the valuation model are assumed to be characterized by normal distributions and some dependence upon one another. We have introduced this dependency in a formal way through our characterization of a non-diagonal covariance matrix for the multivariate normal distribution. With this as our model, we have addressed ourselves specifically to a three-mode case. We have now passed through two simplifying steps in considering our characterization of the three-mode problem.

First, recognizing that the problem is essentially comparative, we

have reduced a three-variable problem to a two-variable problem by examining the relationship between differences in valued modal attributes and the associated errors. Thus, we began with a 3 x 3 error covariance matrix:

$$V(\varepsilon) = X\Sigma_\beta X' + \Sigma_V$$

We re-expressed this as a 2 x 2 covariance matrix of the error differences, $\eta_1$ and $\eta_2$:

$$V(\eta) = w'V(\varepsilon)\, w = \begin{bmatrix} \gamma_1^2 & \rho\gamma_1\gamma_2 \\ \rho\gamma_2\gamma_1 & \gamma_2^2 \end{bmatrix}$$

Now this re-expressed error covariance matrix can be joined to the problem at hand. Recall that:

$$\text{Prob [1 is chosen]} = \Pr \begin{bmatrix} u_1 > u_2 \Rightarrow \hat{u}_1 - \hat{u}_2 > \varepsilon_2 - \ _1(=\eta_1) \\ \text{and } u_1 > u_3 \quad \hat{u}_1 - \hat{u}_3 > \varepsilon_3 - \varepsilon_1(=\eta_2) \end{bmatrix}$$

We can also express the probability that mode 1 will be chosen in the following way:

$$\Pr(1) = \Pr\,[(u_1 - u_2 > \eta_1)(u_1 - u_3 > \eta_2)]$$

$$\int_{\eta_1 = -\infty}^{u_1 - u_2} \int_{\eta_2 = -\infty}^{u_1 - u_3} f(\eta)\,d\eta$$

$$= \int_{\eta_1 = -\infty}^{u_1 - u_2} \int_{\eta_2 = -\infty}^{u_1 - u_3} \frac{1}{2\pi\sqrt{|V\eta\ |}}\, e^{-\frac{1}{2}[\eta' V_\eta^{-1} \eta]}\, d\eta_2 d\eta_1$$

Because we are working with the cumulative probability of a joint distribution of two variables here, we have to employ a double integral. But since we know that all such probability statements are destined for entry into a likelihood function whose maximization will determine parameter values for our utility specification, it is clear that simplification of the double integral is a necessary prelude to econometrics. In the pages immediately preceding this Section, two simplifications were undertaken with this end in view.

First, the double integral was transformed into an equivalent value by converting the problem to standardized form:

$$(77) \quad \int_{n_1=-\infty}^{u_1-u_2} \int_{n_2=-\infty}^{u_1-u_3} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} m'm} \, dm_2 m_1 \quad \left(\text{where } m_i = \frac{n_i}{\gamma_i}\right)$$

We saw in turn that the joint density function in the above expression was equal to a product of two separate density functions:

$$f(m_1 m_2) = f(m_1)f(r_2) = f(m_1)f(m_2|m_1) \qquad \text{where } r_2 = m_2 - \rho m_1$$

Thus, we could also write the double integral in (77) as:

$$(78) \quad \int_{m_1=-\infty}^{\frac{u_1-u_2}{\gamma_1}} \int_{r_2=-\infty}^{\frac{u_1-u_3}{\gamma_2} - \rho m_1} f(m_1)f(m_2|m_1)dm_2 dm_1$$

Now we achieve an opportunity for major simplification. Notice that the double integral above must be evaluated from the inside out, so that the first integral to be evaluated is:

225

$$(79) \qquad \int_{r_2 = -\infty}^{\frac{u_1-u_3}{\gamma_2} - \rho m_1} \frac{1}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{1}{2}\frac{r_2' r_2}{1-\rho^2}} \, dr_2$$

Again, the principle of standardization is useful in performing one further simplification. We have already established that $V(r_2) = 1-\rho^2$.

Thus, if we define the standardized variable $q_2 = \left(\frac{r_2}{\gamma_{r_2}} = \frac{r_2}{\sqrt{1-\rho^2}}\right)$

we can translate (79) to an equivalent integral value:

$$(80) \qquad \int_{q_2 = -\infty}^{\frac{\left[\frac{u_1-u_3}{\gamma_2} - \rho m_1\right]}{\sqrt{1-\rho^2}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} q_2' q_2} \, dq_2$$

Notice that (80) is simply an expression of the cumulative density function for the standardized variable $q_2$, which is distributed normally with mean zero and standard deviation one. For expositional convenience, we can write the cumulative distribution function evaluated at $q_2$ as $F(q_2)$. Thus, (78) can be re-written as

$$\int_{m_1 = -\infty}^{\frac{u_1-u_2}{\gamma_1}} f(m_1) \, F\left[\frac{u_1-u_3}{\gamma_2\sqrt{1-\rho^2}} - \frac{\rho m_1}{\sqrt{1-\rho^2}}\right] dm_1$$

At this point, we have reached a stage of simplification from which econometric estimation is possible.

Before going any further, we can take into account the mathematical specification of the utility function which we have actually been employing:

$$u = X\beta + \varepsilon$$

So that in our case

$$u_1 - u_2 = (X_1 - X_2)\beta$$

and $u_1 - u_3 = (X_1 - X_3)\beta$

$$\text{Thus,}\quad Pr(1) = \int_{m_1 = -\infty}^{\frac{u_1-u_2}{\gamma_1}} f(m_1)F\left[\frac{u_1-u_3}{\gamma_2\sqrt{1-\rho^2}} - \frac{\rho m_1}{\sqrt{1-\rho^2}}\right] dm_1$$

$$= \int_{m_1 = -\infty}^{\frac{(X_1-X_2)\beta}{\gamma_1}} f(m_1)F\left[\frac{(X_1-X_3)\beta}{\gamma_2\sqrt{1-\rho^2}} - \frac{\rho(X_1-X_2)\beta}{\gamma_1\sqrt{1-\rho^2}}\right] = \int_{-\infty}^{m_1(\beta)} f[m_1(\beta)]F[q_2(\beta)]dm_1$$

Notice that since we have been working with a double integral here, we could obtain an exactly-equivalent expression by reversing the order of integration:

$$(81)\quad f(m_1) = \int_{-\infty}^{m_1(\beta)} \int_{-\infty}^{\frac{m_2(\beta)-\rho m_1}{\sqrt{1-\rho^2}}} f(m_2|m_1)f(m_1)dm_2dm_1$$

$$= \int_{-\infty}^{m_1(\beta)} \int_{-\infty}^{\frac{m_2(\beta)-\rho m_1}{\sqrt{1-\rho^2}}} f(m_1|m_2)f(m_2)dm_1dm_2 = f(m_2)$$

227

Both of these expressions are equal to the probability that mode 1 will be chosen in a particular context. Now, we have $P_1 = F(\beta)$, so that we know that when maximum likelihood is employed we are really going to be taking the following derivative:

$$\frac{dP_1}{d\beta} = \frac{\partial f}{\partial m_1} \cdot \frac{\partial m_1}{\partial \beta} + \frac{\partial f}{\partial m_2} \cdot \frac{\partial m_2}{\partial \beta}$$

For $f(m_1)$ in (81) we have

$$\frac{df}{dm_1} = d\left[\int_{-\infty}^{\frac{(X_1-X_2)\beta}{\gamma_1}} f(m_1)F\left[\frac{(X_1-X_3)\beta}{\gamma_2\sqrt{1-\rho^2}} - \frac{\rho(X_1-X_2)\beta}{\gamma_1\sqrt{1-\rho^2}}\right] dm_1\right] \qquad dm_1$$

$$= f\left(\frac{(X_1-X_2)\beta}{\gamma_1}\right) F\left[\frac{(X_1-X_3)\beta}{\gamma_2\sqrt{1-\rho^2}} - \frac{\rho(X_1-X_2)\beta}{\gamma_1\sqrt{1-\rho^2}}\right]$$

Conversely,

$$\frac{df}{dm_2} = f\left(\frac{(X_1-X_3)\beta}{\gamma_2}\right)F\left[\frac{(X_1-X_2)\beta}{\gamma_1\sqrt{1-\rho^2}} - \rho\frac{(X_1-X_3)\beta}{\gamma_2\sqrt{1-\rho^2}}\right]$$

At the same time,

$$\frac{\partial m_1}{\partial \beta} = \frac{\partial\left[\frac{(X_1-X_2)\beta}{\gamma_2}\right]}{\partial \beta} = \frac{X_1-X_2}{\gamma_1}$$

and $$\frac{\partial m_2}{\partial \beta} = \frac{\partial\left[\frac{(X_1-X_3)\beta}{\gamma_2}\right]}{\partial \beta} = \frac{X_1-X_3}{\gamma_2}$$

228

Therefore, we know that one whole set of equations used as first-order conditions for maximum-likelihood estimation will have the following form:

$$(82)\quad \frac{dP_1}{d\beta} = f\left[\frac{(X_1-X_2)\beta}{\gamma_1}\right] F\left[\frac{(X_1-X_3)\beta}{\gamma_2\sqrt{1-\rho^2}} - \rho\frac{(X_1-X_2)\beta}{\gamma_1\sqrt{1-\rho^2}}\right]\frac{(X_1-X_2)}{\gamma_1}$$

$$+ f\left[\frac{(X_1-X_3)\beta}{\gamma_3}\right] F\left[\frac{(X_1-X_2)\beta}{\gamma_1\sqrt{1-\rho^2}} - \rho\frac{(X_1-X_3)\beta}{\gamma_2\sqrt{1-\rho^2}}\right]\frac{(X_1-X_3)}{\gamma_2}$$

Note that this expression has replaced indefinite integrals with sums of products of standard normal probability density functions and cumulative density functions. Since the values of these functions are easy to approximate numerically, it is quite possible to evaluate the expressions across alternative values for elements of the β-vector.

It is important to notice, of course, that the β's are not the only unknowns in this system. One of the main strengths of the approach lies in its inclusion of the error covariances in the system, as well as the valuation parameters themselves. For each choice-probability, then, we also have to determine the gradient of first partial derivatives with respect to the (unknown) covariance parameters. The calculation of this gradient depends on recalling that although ρ is not a function of the β's in this problem, it is a function of the variances of the β's.

Recall that in our case

$$V(n) = \begin{bmatrix} \gamma_1^2 & \rho\gamma_1\gamma_2 \\ \rho\gamma_2\gamma_1 & \gamma_2^2 \end{bmatrix} = W'X\Sigma_\beta X'W + W'\Sigma_V W \qquad \text{where } \Sigma_\beta = \begin{bmatrix} \gamma_{\beta_1}^2 & 0 \\ 0 & \gamma_{\beta_2}^2 \end{bmatrix}$$

229

It is clear that the value of $\rho\left(= \frac{\gamma_{12}}{\gamma_1\gamma_2}\right)$ can be regarded as a function of the $\gamma^2_{\beta_i}$. Therefore, the full gradient must be:

$$\frac{dF}{d\sigma^2_{\beta_i}} = \frac{\partial F}{\partial m_1} \cdot \frac{\partial m_1}{\partial \sigma^2_{\beta_i}} + \frac{\partial F}{\partial m_2} \cdot \frac{\partial m_2}{\partial \sigma^2_{\beta_i}} + \frac{\partial F}{\partial \rho} \cdot \frac{\partial \rho}{\partial \sigma^2_{\beta_i}}$$

$\frac{\partial F}{\partial m_1}$ and $\frac{\partial F}{\partial m_2}$ are as before.

For $\frac{\partial m_1}{\partial \sigma^2_{\beta_i}}$ we have

$$\frac{\partial \left[\frac{(X_1-X_2)\beta}{\sqrt{\gamma_1^2}}\right]}{\partial \sigma^2_{\beta_i}} = \frac{\partial \; (X_1-X_2)\beta(\gamma_1^2)^{-1/2}}{\partial \sigma^2_{\beta_i}}$$

$$= -\frac{1}{2\gamma_1^2} \frac{(X_1-X_2)\beta}{\sqrt{\gamma_1^2}} \cdot \frac{\partial \gamma_1^2}{\partial \gamma^2_{\beta_i}}$$

Similarly, $\frac{\partial m_2}{\partial \gamma^2_{\beta_i}} = -\frac{1}{2\gamma_2^2} \frac{(X_1-X_3)\beta}{\sqrt{\gamma_2^2}} \cdot \frac{\partial \gamma_2^2}{\partial \gamma^2_{\beta_i}}$

Finally, for the third term, we make use of a theorem in mathematical statistics which assures us that the partial derivative of a bivariate normal cumulative density function with respect to the associated correlation co-

230

efficient is simply the bivariate normal probability density function.

Thus,

$$(83)\quad \frac{dF}{d\rho} = \frac{d\left[\int_{-\infty}^{M_1}\int_{-\infty}^{Q_2} f(m_1)f(q_2)\,dm_2\,dm_1\right]}{d\rho} = f(M_1)f(Q_2)$$

So in our case,

$$\frac{dF}{d\rho} = f\left[\frac{(X_1-X_3)\beta}{\gamma_2}\right] F\left[\frac{(X_1-X_2)\beta}{\gamma_1\sqrt{1-\rho^2}} - \rho\,\frac{(X_1-X_3)\beta}{\gamma_2\sqrt{1-\rho^2}}\right]$$

Recall that

$$\rho = \frac{\gamma_{12}}{\sqrt{\gamma_1^2\,\gamma_2^2}} = \gamma_{12}(\gamma_1^2)^{-1/2}\,(\gamma_2^2)^{-1/2}$$

Thus,

$$\frac{d\rho}{d\gamma^2_{\beta_i}} = \frac{\partial\rho}{\partial\gamma_{12}}\cdot\frac{\partial\gamma_{12}}{\partial\gamma^2_{\beta_i}} + \frac{\partial\rho}{\partial\gamma_1^2}\cdot\frac{\partial\gamma_1^2}{\partial\gamma^2_{\beta_i}} + \frac{\partial\rho}{\partial\gamma_2^2}\cdot\frac{\partial\gamma_2^2}{\partial\gamma^2_{\beta_i}}$$

$$= \frac{1}{\sqrt{\gamma_1^2\,\gamma_2^2}}\,\frac{\partial\gamma_{12}}{\partial\gamma^2_{\beta_i}} - \frac{1}{2}\,\frac{\gamma_{12}}{\gamma_1^2\sqrt{\gamma_1^2\gamma_2^2}}\,\frac{\partial\gamma_1^2}{\partial\gamma^2_{\beta_i}} - \frac{1}{2}\,\frac{\gamma_{12}}{\gamma_2^2\sqrt{\gamma_1^2\gamma_2^2}}\cdot\frac{\partial\gamma_2^2}{\partial\gamma^2_{\beta_i}}$$

$$= \frac{1}{\sqrt{\gamma_1^2\gamma_2^2}}\,\frac{\partial\gamma_{12}}{\partial\gamma^2_{\beta_i}} - \frac{\rho}{2\gamma_1^2}\,\frac{\partial\gamma_1^2}{\partial\gamma^2_{\beta_i}} - \frac{\rho}{2\gamma_2^2}\,\frac{\partial\gamma_2^2}{\partial\gamma^2_{\beta_i}}$$

Our gradient of first partials for can be written in the following way:

$$(84)\quad \frac{dP_1}{d\gamma^2_{\beta_i}} = -f\left[\frac{(X_1-X_2)\beta}{\sqrt{\gamma_1^2}}\right] F\left[\frac{(X_1-X_3)\beta}{\sqrt{\gamma_2^2}\sqrt{1-\rho^2}} - \frac{\rho(X_1-X_2)\beta}{\sqrt{\gamma_1^2}\sqrt{1-\rho^2}}\right]\left(\frac{1}{2\gamma_1^2}\frac{(X_1-X_2)\beta}{\sqrt{\gamma_1^2}}\right)\frac{d\gamma_1^2}{d\gamma^2_{\beta_i}}$$

$$-f\left[\frac{(X_1-X_3)\beta}{\sqrt{\gamma_2^2}}\right] F\left[\frac{(X_1-X_2)\beta}{\sqrt{\gamma_1^2}\sqrt{1-\rho^2}} - \frac{\rho(X_1-X_3)\beta}{\sqrt{\gamma_2^2}\sqrt{1-\rho^2}}\right]\left(\frac{1}{2\gamma_2^2}\frac{(X_1-X_3)\beta}{\sqrt{\gamma_2^2}}\right)\frac{d\gamma_1^2}{d\gamma^2_{\beta_i}}$$

$$+ f\left[\frac{(X_1-X_3)\beta}{\sqrt{\gamma_2^2}}\right] F\left[\frac{(X_1-X_2)\beta}{\sqrt{\gamma_1^2}\sqrt{1-\rho^2}} - \frac{\rho(X_1-X_3)\beta}{\sqrt{\gamma_2^2}\sqrt{1-\rho^2}}\right]\left[\frac{1}{\sqrt{\gamma_1^2\gamma_2^2}}\frac{\partial\gamma_{12}}{\partial\gamma^2_{\beta i}} - \frac{\rho}{2\gamma_1^2}\frac{\partial\gamma_1^2}{\partial\gamma^2_{\beta_i}} - \frac{\rho}{2\gamma_2^2}\frac{\partial\gamma_2^2}{\partial\gamma^2_{\beta_i}}\right]$$

Having moved through an admittedly staggering set of transformations, we have now arrived at a series of gradient equations which define the first-order conditions necessary for maximizing the likelihood function. Recall that the general form of this function is:

$$L(\theta) = \prod_{i=1}^{M}\prod_{j=1}^{N} P_{ij}^{y_{ij}}$$

or (in log form)

$$L(\theta) = \sum_{i=1}^{M}\sum_{j=1}^{N} y_{ij}\log P_{ij}$$

Thus, the first-order conditions for maximization are given by

$$(85)\quad \frac{dL}{d\theta} = \sum_{i=1}^{M}\sum_{j=1}^{N}\frac{y_{ik}}{P_{ij}}\frac{dP_{ij}}{d\theta} = 0$$

232

We have jsut developed two exampoes of the kinds of gradients which must be calculated for this exercise:

$$\left[\frac{dP_1}{d\beta_k}\right] \quad \left[\frac{dP_1}{d\sigma^2_{\beta k}}\right]$$

A full specification of the first order conditions simply requires appropriate substitution into the same formulas for the other probabilities in the model.

The advantage of the derivations which we have just performed is that even the rather complex-looking equations which have been developed incorporate known values for most of the included variables. The only parameters which have to be estimated in the model are the valuation-function coefficients (the $\beta$'s) and the variances of those coefficients (the $\gamma^2_\beta$'s). Therefore, the available degrees of freedom will always be N-2T, where N is the number of observations and T is the number of parameters in the valuation function.

## X. Prediction with Covariance Probit Models.

Now that we have explained the basis for estimating the parameters of covariance probit models, we must turn to the question of model use. For this, we must turn our attention again to the known and unknown elements. Two prediction problems are of interest here: (1) The use of parameter estimates to predict the impact of a shift in the value of some attribute of the choice probabilities associated with each mode; and (2) The use of the parameter estimates in forecasting the demand for a new mode.

### (1) Shifting Attribute Values in the Existing System

Recall that our probabilities are specified as double integrals in the three-choice case, and the upper limits to integration are generated by systematic differences in modal utility generated by differences in modal attribute values and their interactions (if any) with personal characteristics. In order to see how the prediction system works, let us review the state of our knowledge after maximum likelihood estimation. The known elements in the system are as follows:

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ X_{31} & X_{32} \end{bmatrix} \qquad \hat{\Sigma}_\beta = \begin{bmatrix} \hat{\gamma}^2_{\beta_1} & 0 \\ 0 & \hat{\gamma}^2_{\beta_2} \end{bmatrix} \qquad \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \qquad \hat{\Sigma}_v = \begin{bmatrix} \hat{\gamma}^2_{v_1} & 0 & 0 \\ 0 & \hat{\gamma}^2_{v_2} & 0 \\ 0 & 0 & \hat{\gamma}^2_{v_3} \end{bmatrix}$$

Recall that

$$\hat{V}_\eta = w'X\hat{\Sigma}_\beta X'w + w'\Sigma_V w$$

and (using mode (1) as an example):

$$(86) \quad Pr(1) = \int_{-\infty}^{(X_1-X_2)\hat{\beta}} \int_{-\infty}^{(X_1-X_3)\hat{\beta}} \frac{1}{2\pi\sqrt{|\hat{V}|}} e^{-\frac{1}{2}\eta'\hat{V}^{-1}\eta} d\eta$$

Here we have all the numbers we need to calculate the value of the double integral which measures the probability of mode 1. The same thing is true for the probabilities of modes 2 and 3.

Thus, given a matrix X of observations in particular modal and personal characteristics, we can construct an appropriate covariance matrix, set the limits of integration, and calculate modal choice probabilities. By the same logic, we can examine the effects of a shift in one variable value by repeating the probability calculations with the altered X-matrix (note that the altered X has an impact both on the covariance matrix estimates and on the upper limits for integration in the new calculation.)

(2) The Demand for a New Mode

Covariance probit analysis is obviously helpful in the analysis of demand shifts in response to attribute value changes in the current system. The use of a non-diagonal covariance matrix in the calculation of modal choice probabilities does away with the "independence of 'irrelevant' alternatives". The problem of constant cross-elasticities which exists in the standard multinomial logit model is therefore done away with. In its ability to model shifts within the current system, however, covariance probit is largely duplicated by the generalized multinomial logit. It is

the ability to predict the demand for new modes which really distinguishes covariance probit. In order to see why the technique handles this problem so neatly, we will consider a case in which a hypothetical fourth mode is added to the three existing modes. We will again use the choice probability of mode 1 as the basis for comparison.

Two questions are relevant in deciding whether we can predict the new choice probability. (1) Can we establish the numerical values of the upper bounds for the relevant integrals? (2) Can we set up the relevant covariance matrix? Fortunately, the answer to both questions is yes. Once we have set the two attribute values for our hypothetical new mode, as well as any interactions with personal characteristics of the relevant decision-makers, we have a new X-matrix:

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ X_{31} & X_{32} \\ X_{41} & X_{42} \end{bmatrix}$$

Here $X_{41}$ and $X_{42}$ are the new attribute values, with the rest of the X-matrix as before.

Since we have a new mode which is being valued, we also have to extend our error vector:

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

For integration, we have an additional error difference: $\varepsilon_4 - \varepsilon\varepsilon_1$

Thus, we have to establish a new w-matrix (such that $w'\varepsilon = \eta$)

In this case,

$$w' = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \quad \text{and } w' \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = \begin{bmatrix} \varepsilon_2 - \varepsilon_1 \\ \varepsilon_3 - \varepsilon_1 \\ \varepsilon_4 - \varepsilon_1 \end{bmatrix} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix}$$

With a new X-matrix and w-matrix, it is clear that we have a new covariance matrix:

$$\hat{V} = w'X\hat{\Sigma}_\beta X'w + w'\hat{\Sigma}_V w$$

Notice, however, that $\hat{\Sigma}_\beta$ and $\hat{\Sigma}_V$ are just as always. We need no new parameter estimates in order to build the new covariance matrices, given the model of random utility with which we began. Notice, in addition, that we can still set all the necessary upper limits to integration since the utility function parameters are just as before.

Thus, armed only with attribute values $X_{41}$ and $X_{42}$ and suitable re-adjustments of w, $\varepsilon$, and $\eta$ we can set up our new cumulative probability calculation as a triple integral of a trivariate normal p.d.f.:

$$(87) \quad Pr(1) = \int_{-\infty}^{(X_1-X_4)\hat{\beta}} \int_{-\infty}^{(X_1-X_3)\hat{\beta}} \int_{-\infty}^{(X_1-X_2)\hat{\beta}} \frac{1}{2\pi\sqrt{|\hat{V}|}} e^{-\frac{1}{2}\eta' V^{-1}\eta} d\eta$$

Although a triple integral may be a little vexatious, all the essential numbers are in hand and the calculation of modal choice probabilities is a straightforward matter. Thus, covariance probit appears to provide an attractive way of modeling the demand for new modes which avoids the predic-

tion bias inherent in the multinomial logit approach. At present, the main weakness of the method lies in the limited ability of existing algorithms to handle more than a few modes. In this chapter, we have worked out a 3-mode case, and even here a rather involved transformation of variables was necessary for the generation of a probability function appropriate for maximum likelihood estimation.

The same general approach has been extended by Hausman and Wise to the case of 4 modes, with the claim that further extension to five should create no major difficulties other than those associated with computation cost. Beyond five modes, however, this approach seems (at present) to lose tractability. Albright, Lerman, and Manski (1977) have developed an alternative estimation routine which offers the promise of extension to at least ten modes, but the method for calculating system parameters appears to be substantially less efficient (and therefore more costly) than the Hausman-Wise procedure. As processing costs continue to fall, however, the cost-benefit calculus may well shift favorably.

Appendix B

## Basic Needs Fulfillment and Economic Growth: A Simultaneous Model

Recent discussions of basic needs satisfaction in the process of development emphasize the importance of changes in the physical quality of life for the poor (1). Although this approach introduces at least one novel element by supporting a broader definition of social justice, it does not explicitly reject the conventional distinction between growth and equity as development objectives. Thus some basic needs proponents advocate the provision of essential goods and services to the poor regardless of productivity effects, while partisans of rapid growth predictably reply with opportunity cost arguments (2). Common to both positions is the implicit assumption that changes in national output an the basic welfare of the poor are separable in the short run, so that the choice of development strategy must ultimately depend on politically-determined value weightings.

In the particular case of basic needs, this intellectual distinction between "welfare" and "productivity" may be very misleading Streeten and others have argued that improvements in health, nutrition,

---

(1) The literature on basic needs is already quite extensive. Good introductions can be found in Ghai (1977) and Lisk (1977).
(2) The strongest critics of the basic needs approach dismiss it as misguided welfarism. In this view, basic needs policies resemble the "consumption" scenario provided by Ahluwalia and Chenery (1974): Durin an extended period, substantial consumption transfers to the poor at th expense of capital formation leave the poor worse off than they would have been under a policy of non-intervention.

and education have labor-augmenting productivity effects, so that basic needs expenditures strongly complement capital accumulation in the development process (3). Conversely, it can be argued that additional resources from output growth will generate widespread improvements in the physical quality of life, even when money incomes remain unequally distributed.

No one would deny, of course, that some sympathetic interaction exists between productivity and the general state of health, nutrition, and education in poor countries. It is the significance of this interaction which is at issue, and quantitative research has been hampered by the relative scarcity of appropriate time-series data. Recently, however, the publication of a large and reasonably consistent set of social indicators by the World Bank has made it possible to begin econometric work (4).

This section attempts to contribute to a more integrated view of the development process by specifying and estimating a growth model which is explicitly simultaneous in national output and generalized measures of health, nutrition, and education. Consistent and relatively efficient parameter estimates are obtained through the application of three-stage least squares, using data from a large sample of poor countries in Africa, Asia, and Southern America. The results give support to the hypothesis that changes in basic welfare are strong contributors to labor productivity change, and conversely. In addition,

---

(3) See Streeten (1979). A strong case for basic human resource development as a pre-condition for equitable growth is made in Adelman and Morris (1973).
(4) The two publications which are used extensively for the empirical work reported in this Appendix are IBRD (1976) and IBRD (1978).

they suggest substantial differences in the pattern of simultaneous interaction at different income and welfare levels. In two illustrativ experiments, total response elasticities are used to demonstrate the apparent importance of initial conditions in determining the impacts of alternative policy approaches.

## (1). A Simultaneous Model of Welfare and Productivity Change

If education is defined as a basic need, then the existence of at least one link between needs satisfaction and economic growth is well known in industrial societies. Work by Denison (1967,1974), Intriligator (1965), and Barger (1969) has established the importance o embodied education as a labor-augmenting component of technical change in the U.S. and Western Europe. Any direct inference from these result to the relationship between truly basic education and productivity growth in poor countries would, of course, be unwarranted. Of more obvious relevance is a recent paper by G.R. Saxonhouse (1977) which clearly demonstrates the impact of primary schooling on labor productivity in early twentieth century Japan. Other studies of educational effects in poor countries have been plentiful, but they ha yielded mixed results (4).

---

(4) Work on agricultural productivity by Hayami and Ruttan (1970) suggests an important role for technical and general education in explaining the differential between rich and poor countries. On the other hand, Nadiri (1972) concludes from a survey of the literature th education has generally been unimportant as a determinant of differentials across poor countries, although within-country contributions often appear significant. An implication of this somewha paradixical result is that education interacts strongly with other

Comparable efforts to examine the productivity impacts of health and nutrition at the aggregate level have been rare. In part this scarcity is due to difficulties in measurement and data acquisition, and in part it undoubtedly reflects a neglect of the problem in the mainstream growth literature. One notable exception to the general trend is a study of Latin American countries by Correa, who finds a strong correlation between nutrition and health mesures and output growth (as well as an apparently insignificant role for education) (5).

Until recently, the same scarcity has characterized the literature on linkages among the basic welfare indices themselves. As the publication of comparative social indicators by various international agencies has increased, some cross-section studies have begun filling the gap. Existing regression results demonstrate, for example, that differences in indices of public health across countries are significantly correlated with differences in per capita income, nutrition, and education (6).

Unfortunately, these studies suffer from two deficiencies which limit their usefulness for policy discussion. First, they compare levels across countries at one point in time and are therefore particularly susceptible to the argument that observed correlations mask the effects of unobserved variables. Cross-section regressions which suggest strong correlations often suffer an embarrassing loss of

---

(unobserved) variables in augmenting productivity.
(5) See Correa (1970). Some micro-level studies of nutritional effects on productivity are also available. See, for example, Basta and Churchill (1974).
(6) Recent pioneering work by Morawetz provides much useful information. In Morawetz (1978), for example, his regression results indicate a strong correlation between literacy rate and life expectancy across poor countries.

"explanatory" power when levels are replaced by intertemporal changes (7). The second weakness of linkage studies (shared by most of the labor productivity research) is their assumption of uni-directional causation when regression equations are estimated, although it is obvious that welfare and productivity changes are simultaneously determined. The existence of simultaneity guarantees some degree of bias in coefficients estimated by ordinary least squares, and it is therefore difficult to interpret many of the results which have been reported in the literature.

Since the requisite data have recently become much more plentiful, it is now possible to do empirical work which confronts both of the problems mentioned above. If cross-section estimates are to be truly useful for policy discussion, they should be able to survive the translation from levels to changes. At the same time, they should be obtained from models of welfare-productivity interaction which are explicitly simultaneous. Here both criteria are applied. Attention is focused on a simultaneous four-equation model of productivity and

---

(7) Again, recent work by Morawetz illustrates this phenomenon. In Morawetz (1977), he presents results for a large number of regressions relating indices of basic needs fulfillment to levels of per capita GNP. For 30 of 32 regressions run using levels for 1960 and 1970, a statistically significant relationship is evident. When changes (1960-1970) in basic needs indicators are regressed on changes in per capita GNP, however, only 5 of 16 regressions exhibit a significant relationship.

welfare determination whose general form (for country i during period t) is given by (8):

(1a) $Q\{it\} = Q\{it\}[\ K\{it\},\ L'\{it\}(L\{it\}, H\{it\}, N\{it\}, E\{it\}),\ A\{t\}\ ]$

(1b) $N\{it\} = N\{it\}[\ Q\{it\}/P\{it\}\ ]$

(1c) $E\{it\} = E\{it\}[\ Q\{it\}/P\{it\},\ G\{Eit\}\ ]$

(1d) $H\{it\} = H\{it\}[\ Q\{it\}/P\{it\},\ N\{it\},\ E\{it\},\ G\{Hit\}\ ]$

where $Q\{it\}$ = Output

$P\{it\}$ = Population

$K\{it\}$, $L\{it\}$ = The use levels of capital and labor services

$L'\{it\}$ = The use level of "effective labor" (i.e. the measure of labor which accounts for productivity improvements).

$H\{it\}$, $E\{it\}$, $N\{it\}$ = Measures of general levels of health, education, and nutrition, respectively

$G\{Eit\}$, $G\{Hit\}$ = Measures of capacity levels for the provision of public education and health services, respectively

$A\{it\}$ = Some measure of the general state of production technology during period t

As is always the case with simultaneous models, the specification incorporates an explicit notion of causality. National output, health, education, and nutrition are endogenously determined, while capital and labor inputs, population, and public capacity levels in health and education are taken to be predetermined. The predetermined variables

---

(8) In this report, brackets of the form "[ ]" will be used to denote subscripts.

are viewed as reflecting decisions by individuals and governments which are not systematically related to contemporaneous values of the endogenous variables.

The first equation is a generalized form of the production function which incorporates the labor-augmenting effects of health, nutrition, and education. The three welfare equations reflect the income-elasticity of consumption patterns which promote nutrition, education, and health. They are themselves quasi-reduced-form equations, since per capita income enters as the primary determinant of demand and the associated supply conditions are taken to be either perfectly elastic (1b) or exogenously determined (1c and 1d).

In (1b), where the link is obviously through the income-elastic demand for food, the potential availability of imports is taken to impl a flat nutrient supply curve. Equation (1c) presumes that the demand for education (G{Eit}) is considered to be a major determinant of the supply of services. In (1d), the prevailing nutrition level is assumed to have an automatic impact on public health. Increased education also enters, through its effect on dietary and sanitary practices. In addition, the demand for other goods and services which are health-promoting (e.g. safe water, sanitation systems, fly screens) is taken to be income elastic, so that a direct impact of per capita incon on health should be observable. On the supply side, the level of publi provision of health personnel and facilities is regarded as the major constraint.

Specification

Since the proposed growth model is explicitly simultaneous, the four equations must be specified compatibly. Existing work provides little guidance for the imposition of particular functional forms on the welfare equations. The properties of alternative production functions, on the other hand, are quite familiar. Attention has therefore been focused on the production equation, with the welfare equations specified in the form which is appropriate for system estimation.

In modeling the production process, it is necessary to give simultaneous attention to the labor-augmenting role of welfare inputs and the role ot capital and effective labor in determining output. The labor augmenting contributions of health, nutrition, and education are assumed to be characterized by unitary elasticity of substitution (9):

(4) L'{t} = L{t} * (H{t}**B1) * (N{t}**B2) * (E{t}**B3)

Two alternative specifications of the production function itself have been considered:

(5a) Augmented CES:

Q{t} = A{t} * [a*(K{t}**p) + (1-a)*(L'{t}**p)]**(1/p)

---

(9) The notation "**" will be used to denote exponentiation in this report and "*" will denote multiplication.

(5b) Augmented Cobb-Douglas:

$$Q\{t\} = A\{t\} * (K\{t\}**g1) * (L'\{t\}**g2)$$

Empirical testing has yielded the conclusion that (5a) is not significantly better than (5b) as a representation of production in this case. Thus, production is assumed to exhibit unitary elasticity of substitution throughout:

$$(6)\quad A\{t\} = A\{t\} * (K\{t\}**g1) * [\ L\{t\}*(H\{t\}**b1)*(N\{t\}**b2)*(E\{t\}**b3)\ ]**g2$$

Here g1 and g2 can be interpreted as the output elasticities of capital and effective labor, while the b's are the labor-augmenting elasticitie of health, nutrition, and education.

Equation (6) would be appropriate for estimation if time series fo a single country were being considered. However, only two sets of observations (for 1960 and 1970) are available for each country in the sample, and the model must be specified to reflect the consequences of one-period changes across countries. The time derivative of (6) is therefore used as the production equation (10).

$$(7)\quad \langle dq\rangle = \langle da\rangle + g1*\langle dk\rangle + g2*\langle dl\rangle + g2*b1*\langle dh\rangle + g2*b2*\langle dn\rangle + g2*b3*\langle de\rangle$$

---

(10) The notation "< >" will be used to denote time derivatives in th report.

where $\langle dx \rangle = [dX\{t\}/dt]/X\{t\}$

This equation is linear in percent changes, and the same structure has been imposed on the three welfare equations. The resulting model is presented in Figure 1 in stochastic form, along with the definitions of the variables actually employed.

Figure 1

Interactive Growth Model

Equations

(8a) $\langle dq \rangle = [\langle da1 \rangle + \langle da2 \rangle * c] + g11 * \langle dk \rangle + g12 * \langle dl \rangle + g12 * b1 * \langle dh \rangle$
$+ g12 * b2 * \langle dn \rangle + g12 * b3 * \langle de \rangle + v1$

(8b) $\langle dn \rangle = g20 + [g21 + g22 * \ln N\{io\}] * [\langle dq \rangle - \langle dp \rangle] + v2$

(8c) $\langle de \rangle = g30 + [g31 + g32 * \ln E\{io\}] * [\langle dq \rangle - \langle dp \rangle]$
$+ [g33 + g34 * \ln E\{io\}] * \langle ds \rangle + v3$

(9d) $\langle dh \rangle = g40 + [g41 + g42 * \ln H\{io\}] * [\langle dq \rangle - \langle dp \rangle]$
$+ [g43 + g44 * \ln H\{io\}] * \langle dd \rangle + [g45 + g46 * \ln H\{io\}] * \langle dr \rangle$
$+ [g47 + g48 * \ln H\{io\}] * \langle dn \rangle + [g49 + g410 * \ln H\{io\}] * \langle de \rangle$
$+ v4$

Variable Definitions (11).

<dq> = Change in gross domestic product

<dk> = Total investment (1960-1970) divided by national output in 1960 (See explanation in text)

<dl> = Change in total labor force

<dd> = Change in population per doctor

<dr> = Change in population per nursing person

<dn> = Change in per capita calories available

<de> = Change in adult literacy rate

<ds> = Change in primary school enrollment ratio

<dp> = Change in population

<dh> = Change in life expectancy at birth

c = A dummy variable accounting for the presence (c = 1) of substantial manufacturing investment by multinational firms

N{o},E{o},H{o} = 1960 levels of calorie availability, adult literacy rate, and life expectancy at birth, respectively

v1,...,v4 = Stochastic error terms (assumed to be additive and normally distributed)

---

(11) All variables defined as percent changes from 1960-1970, except where otherwise indicated. The data are taken from IBRD (1976) and IBR (1978).

The production equation is precisely as specified in (8), with the exception of a shift dummy which has been introduced to allow for possible productivity increments in countries where multinational firms have played a dominant role in manufacturing investment and technology transfer (12). The welfare equations include several measures of public resource availability. In (8c), the change in the primary school enrollment ratio is used as a proxy (admittedly a poor one) for the change in government promotion of basic education. Since rising literacy converges to an asymptote at 100 percent and primary enrollments are likely to have the greatest impact at low literacy levels, (8c) is specified to allow for a declining response elasticity. The same kind of reasoning applies to the introduction of variable response elasticities for medical personnel availability in (8d), as well as for the effects of nutrition and education on health.

In all three welfare equations, the change in per capita income is assumed to play an important role. In the nutrition equation, the income elasticity of demand for nutrients is expected to decline as the nutritional status of the population rises. Similarly, the income responsiveness of demand for health-promoting goods and services should be highest where initial health levels are low (13). Finally, it is assumed that the income elasticity of demand for basic educational services will drop as basic educational needs are satisfied.

---

(12) Among the 54 countries for which complete data are available, the following 12 have been dominant as regional centers for multinational investment: South Korea, Taiwan, Malaysia, Indonesia, Thailand, Pakistan, Iran, Kenya, Ivory Coast, Mexico, Panama, Brazil.
(13) One source of ambiguity here is the probability that part of the measured effect of per capita income increases will represent the greater availability of public resources for health programs, although such resources are not necessarily forthcoming at higher income levels.

Data

The data for this study have been drawn from observations from 77 poor countries in Africa, Asia, and Southern America which are included in two recent IBRD publications (1976 and 1978). By combining data from these tables, it has been possible to construct complete variable sets for 54 countries, which are listed in the appendix. In most cases, the variables employed for estimation (summarized in Figure 1) can only be defended as the best available.

The input and output measures in the three basic welfare equations are all national averages, so that substantial differences in distribution are masked by the data. This problem is apt to be more severe for the measurement of availability of medical personnel than for the other indices. Aggregative measures of population per doctor or population per "nursing person" say nothing about differences in the length or efficacy of training programs, the consistency with which the term "nursing person" is defined across countries, or the geographical distribution of medical personnel within countries. Thus, the two measures are not very satisfactory as indices of generalized access to medical care.

The measures of health, nutrition, and education, on the other hand, may not be too bad. Life expectancy at birth seems to have been accepted as the best available indicator of the general health status of the population, and the variances around national means are certainly much lower for this measure than for income per capita (14). As a

---

(14) See Morawetz (1977).

measure of nutritional status per capita calorie availability is a reasonable approximation. It is now generally accepted that calorie sufficiency is much more likely to indicate nutritional adequacy than any other index. At the same time, available microeconomic studies always show a rapid drop in the income elasticity of calorie consumption across income classes within countries, so that it is reasonable to link expansion in calorie consumption with increases in the nutritional status of the poor as per capita income rises (15). Among measures of the change in embodied basic education, the change in the adult literacy rate is undoubtedly the best available index. Primary school enrollment ratios, on the other hand, are not very reliable as input measures since they mask considerable variance in actual student attendance rates and student-teacher ratios across contries.

The conventional economic indices in the study are subject to many of the usual strengths and weaknesses. The measured change in gross domestic product seems acceptable, aside from the usual index number and exchange rate conversion problems. Similarly, the change in total population presents no particular problem. The way in which <dk> and <dl> are employed in the simple production model is unfortunately much less satisfactory. The use of percent changes in capital and labor in a production model incorporates the assumption of full utilization of the available services. This is obviously wrong (and particularly so for labor in poor countries), but nothing can be done about it because reliable information on capacity utilization rates and unemployment rates for these countries simply does not exist.

---

(15) Complete discussions of these and related issues can be found in Beutlinger and Selowsky (1975).

An additional complication is introduced by the complete non-availability of reliable and comparable capital stock figures for these countries, so that it is impossible to obtain direct measures of percent changes for the period 1960-1970. Available data on yearly investment and output do allow for estimation under the assumption (admittedly heroic) that the general capital-output ratio for the countries in question did not change significantly during the period 1960-1970. Some manipulation of the labor-augmenting production model yields:

$$[dQ/dt]/Q\{o\} = [dA/dt]/A\{o\} + g11*[Q\{o\}/K\{o\}]*[(dK/dt)/Q\{o\}] + g12*[dL'/dt]/L'\{o\}$$

or

$$(7) \quad \langle dq \rangle = \langle da \rangle + g11*[(Q/K)*\langle dk \rangle] + g12*\langle dl' \rangle$$

Since we can observe $[(dK/dt)/Q\{o\}]$ for the period 1960-1970, estimation with some unknown degree of bias is possible. It is comforting to note that the mean capital output ratio must be somewhere in the range (2-5) and the combined output elasticities for capital and effective labor can be somewhat greater than one at most, so the econometric results must conform to certain obvious restrictions.

(2) The Interactive Growth Model: 3SLS Estimates

In the growth model as specified in (8), all of the equations are over-identified. Second-stage parameter estimates were obtained using the predetermined variables in the model and the 1960 levels of

population per doctor and nursing person. A series of dummy variables for major sub-regions were also introduced in the second stage to aid in absorbing unexplained variance, since it was clear that local culture and natural conditions could have an important influence on the processes being observed (16). The third stage estimation applied the conventional procedure, constructing the estimated error variance matrix from the residuals of the second-stage estimates and using the multi-equation equivalent of generalized least squares to generate more efficient estimates for the full set of model parameters (17). Three-stage estimates for the full model are presented in Figure 2. A brief sketch of the results will be presented first, followed by a more detailed discussion for each equation.

---

(16) The sub-regions defined for this work were the Caribbean, Central America, the Andean countries, the rest of South America, North Africa, the Sahel, West Africa, East Africa, West Asia, South Asia, Southeast Asia, East Asia.

(17) The classic discussions of this approach can be found in Zellner and Theil (1962), and Madansky (1964).

Figure 2

3SLS Estimates: Interactive Growth Model(a)

(54 observations)

| | SE | R**2 |
|---|---|---|
| **Productivity** | | |
| <dq> = [-.203 + .355e] + .184<dk> + .238<dl> | .252 | .52 |
| (.091) (.063) (.035) (.179) | | |
| + 1.697<dh> + 2.652<dn> +.0078<de> | | |
| (.751) (.391) (.042) | | |
| **Nutrition** | | |
| <dn> = .019 + [2.932 - .626lnN{o}]*[<dq>-<dp>] | .086 | .15 |
| (.018) (1.402) (.310) | | |
| **Literacy** | | |
| <de> = .568 + [1.630 - .420lnE{o}]*[<dq>-<dp>] | .685 | .19 |
| (.180) (.952) (.293) | | |
| + [2.470 - 1.088lnE{o}]*<ds> | | |
| (.706) (.283) | | |
| **Health** | | |
| <dh> = .099 + [.171 - .028lnH{o}]*[<dq>-<dp>] | .049 | .02 |
| (.015) (.436) (.115) | | |
| + [-.320 + .085lnH{o}]*<dd> + [-.756 + .205lnH{o}]*<dr> | | |
| (.802) (.208) (.614) (.161) | | |
| + [3.140 - .803lnH{o}]*<dn> + [.707 - .200lnH{o}]*<dn> | | |
| (2.864) (.724) (.487) (.134) | | |

---

(a) In the reported results, standard errors are in parentheses under the estimated coefficients.

The productivity equation suggests a very strong labor-augmenting role for health and nutrition in the determination of output (income) change for poor countries. Conversely, when income change outpaces population growth it has a significant impact on general levels of nutrition and literacy. Some impact on health may also be present, but multicollinearity from the large number of interactive terms in the health equation has generated standard errors which are too large for much confidence to be placed in the results. In the first two welfare equations, the indicated effect of change in per capita income is substantial at low welfare levels and declines to values near zero at relatively high levels (100 percent calorie adequacy; 50 percent literacy). A similar pattern of declining income responsiveness appears in the health equation, although the impact of income does not seem to be large even at very low health levels.

Among the policy-determined variables which have been introduced (<ds>,<dd>,<dr>), the results are generally as expected, with an apparent pattern of declining elasticity as initial welfare levels rise (18). In the case of education, the results suggest a significant but rapidly declining responsiveness to enrollment changes. The standard errors in the health equation are again quite large, but the same sort of pattern seems to be present. The indicated impact of change in nursing persons is particularly high at low health levels, which supports the notion that such personnel make their most valuable contribution in raising basic standards of hygiene and child care.

---

(18) It should be borne in mind that the indices for medical personnel have been entered as population/doctor and population/nursing person, so that the expected value of the associated coefficients is less than zero.

In the health equation, the simultaneous relationship with income is complemented by those of nutrition and education. In both cases, the suggested responsiveness is quite high at low levels of health and exhibits the familiar pattern of marginal decline. Although the cofficients represent the best estimates in the usual sense, multicollinearity has again generated standard errors which are uncomfortably large.

Generally, then, the hypothesis suggested by Streeten and others appears to be supported by the data, even over relatively short periods. Changes in the physical quality of life appear to have a strong effect on productivity in poor countries, and conversely. The one unexpected result is the seemingly negligible impact of literacy change on productivity. This problem will be considered again after some attention has been given to technical detail in the results.

In the productivity equation, imposition of constant returns to scale (g11+g12 = 1) yields an estimated output elasticity of .76 for capital and an implied capital-output ratio of 4.14. While the latter measure is unsurprising, the indicated magnitude of the output elasticities precisely reverses the pattern generally observed in industrial societies (19). It seems that the growth of available labor at constant quality levels has had a relatively minor impact on output change. The effects of changes in basic labor quality, on the other hand, appear to have been dramatic. The labor-augmenting elasticities of health and nutrition implied by the parameter estimates (7.13 and

---

(19) For wealthier countries, the estimated elasticities for labor and capital are generally around 0.70 and 0.30, respectively. See for example Barger (1969).

11.14, respectively) are the most important impacts observable in the results (20). When the measured contributions of capital and augmented labor to output are calculated using the sample mean values, the impact of basic welfare changes leads to a labor contribution which is approximately equal to that of capital.

The nutrition equation results imply no systematic unexplained component of improvement, although the proportion of unexplained variation is obviously large. The implied response elasticity with respect to per capita income is .27 at 70 percent of average adequacy (the minimum world figure in 1960) and almost zero at the 100 percent level. In the literacy equation, the results suggest a large increase in the general rate of literacy during the 1960's which can be attributed neither to rising per capita income levels nor to changes in the available measure of basic schooling (21). At the same time, a

---

(20) The 3SLS parameter estimates for <dm> and <dn> are relatively efficient from a systems perspective, but a look at the two-stage estimates leads to the suspicion that the impact of nutrition is somewhat exaggerated in the final stage by estimation problems in the literacy equation. It has been necessary to specify this equation as linear in percent changes for compatibility with the productivity equation. Unfortunately, five African countries exhibit grossly disproportionate percent changes for the sampling period because of low initial levels (5 percent or less). When these five observations are deleted from the sample and the entire system is re-estimated, the error of the third equation, the constant term, and the estimated impact of enrollment changes at low literacy levels are all substantially reduced. All coefficients remain significant at the 95 percent level, however. In the productivity equation, the resulting estimates for the coefficients of <dm>, <dn>, and <de> are changed from 1.697 to 1.729, 2.652 to 2.235, and .0078 to .141, respectively, and the last estimate becomes significant at the 90 percent level. While selective elimination of outliers is generally not a very good idea, their removal in this case does suggest some over-estimation of the effect of health on productivity and an under-estimation of the effect of literacy when the full sample is employed. Although the full sample results will be used here, this caveat should be kept in mind.

(21) This is true whether or not the five literacy outliers are in the sample although the estimate is smaller when they are deleted.

substantial marginal impact for changes in income and schooling is indicated at very low levels of literacy. At a rate of 3 percent, for example, the implied elasticities are 1.17 for per capita income and 1.27 for primary enrollment, while the marginal contributions decline to zero at literacy rates of 48 percent and 10 percent, respectively. Obviously, the systematic part of this equation is relevant only for countries whose literacy rates were low initially.

In the health equation, a significant part of the systematic rise in world health during the 1960's remains unexplained. In the sample under consideration the mean increase for the period 1960-1970 was about 10 percent and the maximum somewhat greater than 20 percent, so an estimated constant of .099 is substantial. The fit for this equation is generally the worst of the lot, although multicollinearity has greatly affected the standard error estimates. All signs in the equation are as expected, but only nutrition responsiveness appears to be very sizable at low health levels. All elasticities seem to decline to the vanishing point at a life expectancy in the early 40's, which was about the international median in 1960.

## Total Impact Evaluation

The size of the unexplained variance in the three welfare equations makes it clear that specific national factors have had a lot to do with changes in basic welfare. From the perspective of public policy, it is of interest to examine the patterns of total response to alternative approaches which are suggested by the data. Although the estimates for the structural model cannot yield information about

relative costs, they can indicate the marginal benefits associated with investment and basic welfare changes. Using some notional cost figures, it is possible to obtain extremely crude estimates of the relative net returns to alternative approaches. As an aid to investigation in this context, the structural model has been re-estimated to allow for sensitivity to initial average income levels in the productivity equation parameters. The results are presented in Figure 3.

Figure 3

3SLS Estimates: Modified Growth Model

(54 observations)

| | SE | R**2 |
|---|---|---|
| **Productivity** | | |
| $\langle dq\rangle = [-.143 + .382c] + .228\langle dk\rangle + .124\langle dl\rangle$ (.118) (.092) (.058) (.222) | .225 | .64 |
| $+ [8.027 - 1.392\ln(Q/P)\{o\}]*\langle dm\rangle + [-5.970 + 1.539\ln(Q/P)\{o\}]*\langle d$ (4.256) (1.027) (4.352) (.833) | | |
| $+ [-.510 + .102\ln(Q/P)\{o\}]*\langle de\rangle$ (.662) (.146) | | |
| **Nutrition** | | |
| $\langle dn\rangle = .025 + [3.154 - .679\ln N\{o\}]*[\langle dq\rangle-\langle dp\rangle]$ (.018) (1.554) (.343) | .086 | .16 |
| **Literacy** | | |
| $\langle de\rangle = .602 + [1.841 - .518\ln E\{o\}]*[\langle dq\rangle-\langle dp\rangle]$ (.178) (.906) (.275) | .680 | .21 |
| $+ [2.331 - 1.026\ln E\{o\}]*\langle ds\rangle$ (.670) (.272) | | |
| **Health** | | |
| $\langle dh\rangle = .096 + [.257 - .053\ln H\{o\}]*[\langle dq\rangle-\langle dp\rangle]$ (.016) (.463) (.122) | .050 | .22 |
| $+ [-.298 + .078\ln H\{o\}]*\langle dd\rangle + [-.643 + .174\ln H\{o\}]*\langle dr\rangle$ (.841) (.218) (.648) (.170) | | |
| $+ [2.938 - .733\ln H\{o\}]*\langle dm\rangle + [.783 - .221\ln H\{o\}]*\langle de\rangle$ (2.989) (.758) (.496) (.137) | | |

No significant change in the welfare equation estimates is generated by this re-specification of the first equation. In the productivity equation itself, some interesting patterns seem to emerge. As per capita income increases, the impact of health changes declines sharply, while the response elasticity for nutrition evidently

increases. A similar pattern of increase is evident for literacy responsiveness, although the estimated standard errors are quite large and the coefficient values relatively small for characteristic sample income levels.

When the total pattern of results is considered, an interesting response hierarchy suggests itself. The very poor countries which improved health conditions rapidly seem to have gotten the biggest productivity response. With movement up the income scale toward 150.00 per capita, there is a steady gain in relative importance for nutrition improvement, while some positive role for literacy is suggested at incomes higher than 300.00 per capita. Although any structural interpretation of these results must be tentative, to say the least, they may reflect the shift away from subsistence agriculture toward cash cropping and manufacturing with movement up the income scale. If literacy makes a significant contribution to productivity, it may well be at this higher level, while poorer countries experience the greatest output gains from improvements in basic health and nutrition.

Whether or not this structural interpretation is correct, the estimates in Figure 3 can be useful for comparing the apparent impacts of alternative investment and welfare policies. However, the total impact of exogenously-induced changes in the capital stock, health, nutrition, and education cannot be read from individual equations, because successive rounds of changes will result from any once-for-all shift in a particular policy variable. It is in this context that steady-state solutions become interesting. For purposes of illustration, two sets of total elasticities have been calculated using representative values for countries in the low and middle ranges of the

sample. The results are presented in Figure 4.

Figure 4

Total Response Elasticities

(A) Low-Range Country

(Q/P = 70.00, E = 3, H = 30 yea

Policy Variable

| | <dk> | <dh> | <dn> | <dd> | <dr> |
|---|---|---|---|---|---|
| <dq> | 1.964 | 4.738 | 3.381 | -.157 | -.242 |
| <dn> | .528 | 1.178 | 1.909 | -.042 | -.065 |
| <de> | 2.498 | 6.027 | 4.301 | -.200 | -.308 |
| <dh> | .151 | 1.365 | .705 | -.045 | -.070 |

(b) Mid-Range Country

(Q/P = 200.00, N = 85, E = 20, H = 40 years)

Policy Variable

| | <dk> | <dh> | <dn> | <de> | <dd> | <dr> |
|---|---|---|---|---|---|---|
| <dq> | 1.364 | 1.015 | 3.582 | .016 | -.010 | -.001 |
| <dn> | .187 | .139 | 1.491 | .002 | -.001 | -.0001 |
| <de> | .394 | .293 | 1.035 | 1.005 | -.003 | -.0003 |
| <dh> | .083 | 1.062 | .397 | -.030 | -.011 | -.0011 |

Three characteristics of the estimates in Figure 4 are of

particular interest. First, they demonstrate the simultaneous operation of the system. In set (A), for example, the predicted impact of a 1 percent decrease in population per nursing person is an increase in life expectancy (.07 ), but also increases in output (.24 ), nutrition (.065 ), and literacy (.308 ) as the resulting health change works its effect on the whole system. Similar patterns are evident for the other policy variables which have positive impacts.

The second evident characteristic of the estimates is a general decline in system responsiveness at higher welfare levels. A comparison of the results for (A) and (B) reveals only one elasticity which is higher in the latter set, while most are substantially smaller. This follows, of course, from the general pattern of declining elasticities which was evident in the individual estimated equations.

Finally, the two sets of estimates provide a good illustration of the changing response hierarchy which characterizes the welfare-productivity linkage (and all other linkages in the model, because of simultaneity). In set (A), both health and nutrition changes have substantially higher response elasticities than changes in the capital stock, and the measured responsiveness to changes in medical personnel is quite small. In set (B), nutrition changes generate higher proportional responses than any of the other variables, the positive effect of literacy changes is just beginning to appear, and the effect of changes in medical personnel has all but vanished.

Superior marginal benefits do not necessarily lead to higher marginal net benefits, of course, and it is of interest to touch briefly on the policy options from this perspective. If the numbers in set (A) are to be believed, the case for health promotion seems quite strong.

For a small developing country with a population of 10 million, a per capita income of 70.00, and a capital-output ratio of 4.0, a 1 percent shift in the capital stock would cost approximately 28 million and the final result (according to the estimated impact multipliers) would be a 1.96 percent increase in total output.

Given the response elasticity associated with health improvements (4.738), generalized investment would be preferable to health expenditures only if it cost more than 68 million to raise life expectancy by 1 percent. For a very poor country with a life expectancy of 30 years it seems doubtful that a marginal improvement in life expectancy would require resources of such magnitude. Even if this were the case, a decision in favor of generalized investment would imply a zero marginal valuation on the change in life expectancy itself, which would be disproportionately higher under the health expenditure option.

Similar comparative calculations can be done for policies which promote nutrition or educate medical personnel. In case (A), the general conclusion is that almost any welfare investments seem equivalent or superior to general capital accumulation at the margin. The relative desirability of the health option has clearly declined by the time stage (B) is reached, although the impact of nutrition changes is still important.

(3) Conclusion

The econometric estimates obtained by this study seem to reveal some strong patterns of interaction between growth and basic welfare. In a simple numerical exercise, the full set of equations has been used

to develop impact multipliers which illustrate the apparent economic desirability of health and nutrition improvements in very poor countries.

Thus, the available data are shown to be consistent with the notion that some basic needs expenditures can legitimately be regarded as investments in human capital. On the other hand, it must be remembered that the available data are still very poor. The general indices of health, nutrition, and education are subject to large measurement errors. In addition, they are only crude proxies for the measures of labor force quality which would ideally be used in such a study. The data will undoubtedly get better with the passage of time, and much more precise modeling exercises will become possible.

At this point in time, it seems fair to say that a detailed examination of the evidence gives considerable support to the notion that external benefits do flow from basic needs improvements. Furthermore, the results contain at least a suggestion that optimal targeting strategies may shift as the process of development continues. At the lowest level of development particular attention to basic nutrition and health may yield the highest net benefits, while relatively greater concentration on education may hold the key at highe levels. While the staging results which have emerged from this study are too weak to serve as a policy guide, they do suggest the need for a close look at the question of time-phasing in particular national contexts. If the most important meaning of recurrent cost is "recurren opportunity cost" and the level of net social subsidy is the measure of primary concern to the government, then it makes perfect sense to move resources toward the highest relative level of current return.

BIBLIOGRAPHY


elman, I., and C. T. Morris, Economic Growth and Social Equity in Developing Countries (Stanford: Stanford Press, 1973)

hluwalia, M. and H. Chenery, Redistribution with Growth (London: Oxford Press, 1974)

ustin, J., et. al., Poverty and Malnutrition in Third World Urban Areas, IBRD Report, 1976

arger, H., "Growth in Developed Nations," Review of Economics and Statistics, (51), 1969, pp. 143-48

asta, S. and A. Churchill, "Iron Deficiency Anemia and the Productivity of Adult Males in Indonesia," IBRD Working Paper No. 175, April 1974

hristensen, L., D. Jorgensen, L. Lau, "Transcendental Logarithmic Production Frontiers," Review of Economics and Statistics, (65), 1974, pp. 28-45

lark, E., "Socialist Development in an Underdeveloped Country: The Case of Tanzania," World Development, 3,4, April, 1975, pp. 223-228.

liffe, L., et. al., Rural Cooperation in Tanzania, (Dar es Salaam, Tanzania Publishing House, 1975)

orrea, H., "Sources of Growth in Latin America," Southern Economic Journal, (37), July 1970, pp. 17-31

enison, E.F., Why Growth Rates Differ, (Washington, D.C.: Brookings Institution, 1967)

" , Accounting for U.S. Economic Growth, 1929-1969 (Washington, D.C.: Brookings Institution, 1974)

ominguez, J., "Socialism in Cuba," in H. Desfosses and J. Levesque, eds., Socialism in the Third World, (New York: Prager, 1975)

ozier, C., Land Development and Colonization in Latin America, (New York: Praeger, 1969)

berstadt, N., "Has China Failed?" New York Review of Books, April, 1979

hai, D.P., et. al., The Basic Needs Approach to Development (Geneva: ILO, 1977)

rimes, O., Housing For Low Income Urban Families, (Baltimore: IBRD-Johns Hopkins, 1976)

arris, J., "Some Thoughts on a Housing Policy for Nairobi," Discussion Paper No. 78, (Nairobi: IDS, 1969)

Hartman, R.S., A Generalized Logit Formulation of Individual Choice, M.I.T. Energy Laboratory Working Paper, No. MIT-EL-010WP, February, 1979

Hartman, R.S., "A Note on the Use of Aggregate Data in Individual Choice Models, M.I.T. Energy Laboratory Working Paper (Draft), November, 1979

Hausman, J., and D. Wise, "A Conditional Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences, Econometrica, March 1978, pp. 403-26.

Hayami, Y., and V. Ruttan, "Agricultural Productivity Differences Among Countries," American Economic Review, (60), December 1970, pp. 895-911

Hirschman, A.O., Development Projects Observed, (Washington, D.C.: The Brookings Institution, 1967)

IBRD, World Tables (Washington, D.C.: IBRD, 1976)

IBRD, Social Indicators (Washington, D.C.: IBRD, 1978)

Intriligator, M.D., "Embodied Technical Change and Productivity in the United States, 1929-1958," Review of Economics and Statistics, (47), 1965, pp. 65-70

Katzman, M., "Colonization as an Approach to Regional Development," Economic Development and Cultural Change, (29), July, 1978, pp. 709-724

Lal, D., Methods of Project Analysis: A Review, IBRD Staff Occasional Papers No. 16, (Washington, D.C.: IBRD, 1974)

Leibenstein, H., Beyond Economic Man: A New Foundation for Microeconomics (Cambridge: Cambridge Press, 1976)

Leontief, W. "Notes on a Visit to Cuba," New York Review of Books, August, 1969

Lisk, F., "Conventional Development Strategies and Basic Needs Fulfillment," International Labor Review, (115), March-April, 1977, pp. 175-191

Little, I.M.D., and J.A. Mirrlees, Project Appraisal and Planning for the Developing Countries, (London: Heinemann, 1974)

Madansky, A., "On the Efficiency of Three-Stage Least Squares Estimation, Econometrica, (32), 1964, p. 55

Manski, C., R. Albright, and S. Lehrman, Report on the Development of an Estimation Program for the Multinomial Probit Model, Report to the U.S. Federal Highway Administration, October, 1977

cFadden, D., and T. Domencich, Urban Travel Demand: A Behavioral Analysis, (New York: North-Holland, 1975)

orawetz, D., Twenty-Five Years of Economic Development, (Baltimore: Johns Hopkins Press, 1977)

" , "Basic Needs Policies and Population Growth," World Development, (6), 1978, pp. 1251-1259

ouat, L., "Third World Parents More Receptive to Family Planning," The Christian Science Monitor, August 20, 1979, p. 10

usgrove, P., et. al., Consumer Behavior in Latin America: An ECIEL Study, (Washington, D.C.: The Brookings Institution, 1978)

adiri, M., "International Studies of Factor Imports and Total Factor Productivity: A Brief Survey," Review of Income and Wealth, June 1972, pp. 129-154

elson, M., The Development of Tropical Lands, (Baltimore: Johns Hopkins Press, 1974)

elson, P. and S. Winter, "Neoclassical vs. Evolutionary Theories of Economic Growth: Critique and Prospectus," The Economic Journal, December, 1974

yerere, J., "The Arusha Declaration Ten Years After," International Development Review, (1), 1977, pp. 2-7

Pigou, A.C., The Economics of Welfare (Cambridge: Cambridge Press, 1919)

Reutlinger, S., and M. Selowsky, "Undernutrition and Poverty," IBRD Staff Working Paper No. 202, December, 1975

Sabot, R., Economic Development and Urban Migration: A Study of Tanzania, (Oxford: Clarendon Press, 1977)

Saxonhouse, G.R., "Productivity Change and Labor Absorption in Japanese Cotton Spinning, 1891-1935," Quarterly Journal of Economics, May, 1977, pp. 195-219

Schumpeter, J.A., The Theory of Economic Development, (Cambridge: Harvard Press, 1934)

Schydlowsky, D., "General Disequilibrium Shadow Prices," Boston University Department of Economics Working Paper, 1973

Sen, A., and S. Marglin, Guidelines for Project Evaluation, (New York: UNIDO, 1972)

Squire, L., and H. van der Tak, Economic Analysis of Projects, (Baltimore: IBRD-Johns Hopkins, 1975)

Streeten, P., "Basic Needs: Premises and Promises," Journal of Policy Modeling, (1), January 1979, pp. 136-146

Tobin, J., "On Limiting the Domain of Inequality," Journal of Law and Economics, October 1970

USDA, "Bonus Food Stamps and Cash Income Supplements," Marketing Research Report No. 1034, USDA Economic Research Service, October, 1974

USAID, Evaluation Handbook, (Washington, D.C., USAID, 1974)

Weitzman, M., "Is the Price System or Rationing More Effective in Getting A Commodity to Those Who Need It Most?" M.I.T. Dept. of Economics Working Paper No. 140, October, 1974

Wheaton, W., "A Bid Rent Approach to Urban Housing Demand," Journal of Urban Economics," (4), 1977

Wheeler, D., "Do the Poor Need Nutrition Education?" Boston University Dept. of Economics Working Paper, 1978

Williamson, O., Markets and Hierarchies, (New York: The Free Press, 1975)

Zellner, A., and H. Theil, "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Relations, Econometrica, (30), 1962, pp. 54-78