MSU RURAL DEVELOPMENT SERIES

WORKING PAPER

SINGLE-EQUATION ESTIMATION OF FOOD CONSUMPTION CHOICES IN RURAL SIERRA LEONE

By

Victor E. Smith, John Strauss
and Peter Schmidt

Working Paper No. 13 1981

Department of Agricultural Economics
Michigan State University
East Lansing, Michigan 48824

MSU RURAL DEVELOPMENT WORKING PAPERS

Carl K. Eicher and Carl Liedholm, Co-editors

The MSU Rural Development Working Papers series is designed to further the comparative analysis of rural development in Africa, Latin America, Asia, and the Near East. The papers will report research findings on community development and rural development programs. The series will include papers on a wide range of topics such as alternative rural development strategies; off-farm employment and small-scale industry; marketing problems of small farmers; agricultural extension; interrelationships between technology, employment, and income distribution; and evaluation of rural development projects. While the papers will convey the research findings of MSU faculty and visiting scholars, a few papers will be published by researchers and policy-makers working with MSU scholars on cooperative research and active programs in the field.

The papers are aimed at teachers, researchers, policy-makers, donor agencies, and rural development practitioners. Selected papers will be translated into French, Spanish, and Arabic. Libraries, individuals, and institutions may obtain single copies of the MSU papers free of charge and may request their names be placed on a mailing list for periodic notifications of published papers by writing to:

MSU Rural Development Working Papers
Department of Agricultural Economics
206 International Center
Michigan State University
East Lansing, Michigan 48824
U.S.A.

SINGLE-EQUATION ESTIMATION OF FOOD CONSUMPTION CHOICES IN RURAL SIERRA LEONE

By

Victor E. Smith,** John Strauss,***
and Peter Schmidt**

*Published under Agency for International Development Contract AID/DSAN-C-0008, with Michigan State University.

**Professor, Department of Economics, Michigan State University East Lansing, Michigan.

***Graduate Research Assistant, Department of Agricultural Economics, Michigan State University, East Lansing, Michigan.

1981

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

PREFACE

In this report we use single-equation least squares regressions to study the determinants of food consumption choices in rural Sierra Leone. It is the fourth in a series of reports arising from a study of the effects of economic policy on food consumption choices and levels of household nutrient intake. The first report (Kolasa, 1979) described the nutritional situation in Sierra Leone. The second and third (Smith, et al., 1979, and Smith, et al., 1980) presented estimates of the quantities of foods consumed by rural households in Sierra Leone and presented tabular analyses of the effects of non-price factors affecting those quantities. The present report continues the analysis of the determinants of food consumption choices among these rural households.

The project as a whole is under the direction of Professor Victor E. Smith of the Department of Economics, Michigan State University. It is financed by the United States Agency for International Development (USAID) under Contract No. AID/DSAN-C-0008. The data were collected in Sierra Leone during 1974-75 by the Rural Employment Research Project at Njala University College, Sierra Leone (financed by a contract, AID/cds 3625, between the United States Agency for International Development and Michigan State University, and by the Rockefeller Foundation).

We are indebted to Dr. Dunstan S.C. Spencer and Dr. Derek Byerlee who collected the original data and who, along with many others, have been extremely helpful to us in our efforts to interpret the data. We especially appreciate the assistance given to us by three informants from Sierra Leone, Mrs. Agnes Becker, graduate student from Sierra Leone in the Department of Family Ecology at Michigan State University, Mr. Alimami Kargbo, graduate student in the Department of Agricultural Economics at Michigan State University, and Dr. Joseph Tommy, Acting Head of the Department of Agricultural Economics and Extension at Njala University College, Njala, Sierra Leone. To these and many others who have helped in many ways, we express our appreciation.

INTRODUCTION

Effective programs to improve nutrition in a developing country must be based on an understanding of how economic and other variables affect the food consumption patterns of (1) households at risk from malnutrition and (2) semi-subsistence households--those producing much of their own food. We cannot assume that the food choices of households at risk resemble those of other households, that low-income households behave like the well-to-do, and that rural households make the same food choices as urban ones, yet almost all the information that exists about food choice behavior is for countries as a whole, or for urban populations. This study examines the food consumption choices of a rural population consisting predominantly of low-income households that produce much of their own food. It will provide elasticities and predicting equations that are specific to low-income households and will pay special attention to the distinctive characteristics of semi-subsistence households.

To understand the food consumption choices made by semi-subsistence households, one must examine both the production and consumption sides of household activity. Purchases from the market are influenced by market prices and the amount of income in the form of money available for buying food and other items from the market, but what part of total income is received in money depends upon how the household divides its energies between production for home consumption and production for sale. For that reason we have used variables from both the production and the consumption side of the market in the single-equation least squares regressions used in this report. The regressions represent the combined effect of the production and consumption decisions that the household makes in response to the economic and demographic variables that partially define the situation in which the household finds itself.

Opposing views exist concerning the nutritional consequences of the process whereby households largely dependent upon their own productive efforts for food begin to enter the market economy, producing crops for sale and increasing the proportion of their food consumption obtained from the market. Nutritionists and others assert that the quality of a diet suffers when a household shifts from producing its own food to producing crops for sale. Economists, on the other hand, usually affirm that the production of cash crops occurs because larger incomes can be earned in that way, and that larger incomes result in improved diets. Which belief is valid is of major importance to any well-conceived program for nutritional improvement.

The fact is that each opinion finds some support in the empirical evidence. What is required is knowledge of what is likely to happen in any particular case. This report will examine the evidence for rural Sierra Leone, limiting the analysis to what can be accomplished by relatively simple and inexpensive means--the single-equation least squares regression. The following report will present the results from a system-of-equations analysis of the consumption side of a household-firm model, using the same data.

CHAPTER I

PURPOSE AND PROCEDURE

This study is the fourth in a series of reports dealing with food consumption among rural households in Sierra Leone. It presents the results of single-equation multiple regression analysis of the determinants of food-consumption patterns.

Relation to Earlier Work

In 1974-75 the Rural Employment Research Project, under the direction of Dunstan S.C. Spencer and Derek Byerlee, conducted a nationwide survey of rural household farm and nonfarm activities in Sierra Leone. The project was financed by a contract, AID/cds 3625, between the United States Agency for International Development and Michigan State University, and by the Rockefeller Foundation. It collected detailed data concerning the whole range of farm and nonfarm production activities through twice weekly interviews over a period of 12 to 14 months. Data on household expenditures were collected from half the households by interviews scheduled to occur twice during one week of each month. The sampling and interviewing procedure is described in Smith et al., 1980, pp. 4-11.

The sample was stratified in such a way as to provide equal representation of all the major agro-climatic or resource regions, which we shall call ecological zones. Two parts of Sierra Leone were excluded: the Western Area because it is primarily urban and the northern part of the Eastern Province because the patterns of agriculture behavior there were likely to be affected by the presence of diamond mining. The remainder of the country was divided into eight zones, Numbers 1, 3, 5 and 7 of which constitute the Northern Province and Numbers 2, 4 and 8 of which correspond closely to the Southern Province. (See Figure 1.) Zone six represents roughly the southern two-thirds of the Eastern Province.

The present research project is concerned with describing household food consumption patterns in rural Sierra Leone and measuring the influence of a number of economic and noneconomic variables that determine those patterns. Our first task was to estimate the quantities of foods available for consumption in rural households. This we did by using the expenditure survey data to estimate the quantities of foods purchased from the market and the

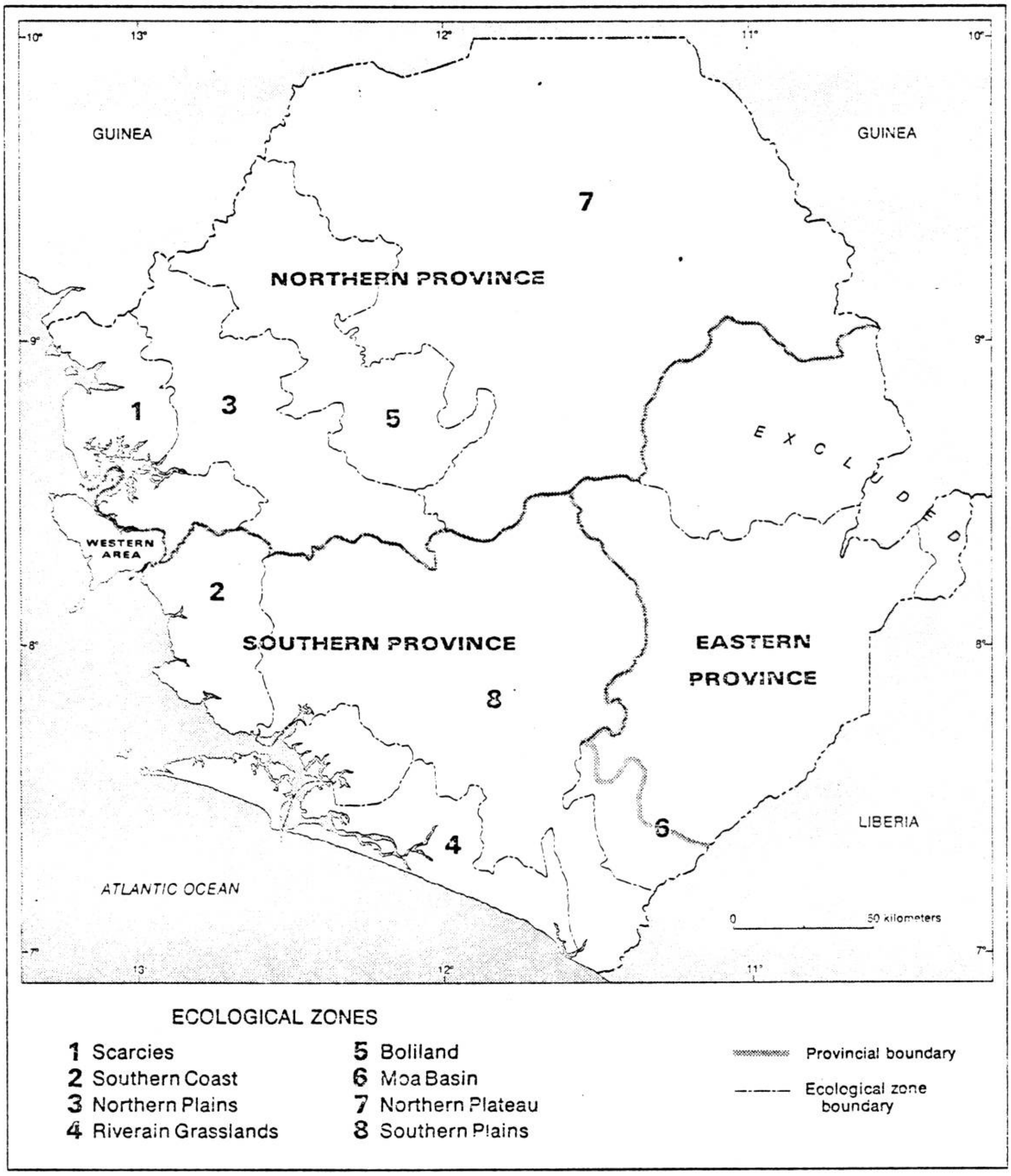

Figure 1

Sierra Leone: Ecological Zones

production survey data to estimate the quantities of foods available for home consumption. Details of the procedure and the resultant estimates of quantities consumed were presented in Smith et al., 1979, along with simple tabular analysis that relates the consumption per consumer equivalent of a variety of foods to such variables as income, the number of consumer equivalents in the household, the dependency ratio, region, market orientation, and the percentage of labor devoted to the production of upland rice.

A later study [Smith et al., 1980] extended the analysis by using cross-tabulation to analyze the effects upon food consumption per consumer equivalent of two variables acting simultaneously. Neither of these studies dealt with the influence of prices upon consumption patterns. The second of them examined the common belief among nutritionists and others that production patterns affect the quantity of food consumed. For cassava, palm oil and groundnuts producing large proportions of the food consumed encourages greater consumption. For rice, on the other hand, no such relationship appears. Likewise, production for the market is sometimes associated with smaller consumption levels, in particular for cereals other than rice, alcoholic beverages (almost entirely palm wine), and in some situations, cassava. At the same time, these beliefs are not supported for some foods.

Regression Analysis

While the cross-tabulation results are helpful, they do not take us far toward our ultimate goal of analyzing the effects upon food consumption choices of a wide range of variables. The size of the sample available effectively limited to us to using two independent variables at a time in the cross-tabulation work, but that left many uncontrolled variables. Moreover, in classifying each variable into only three or four categories, we lost the information contained in the variation among the individual observations, replacing those observations by a single value for each class, and sometimes an unrepresentative one at that.

Regression analysis, although less flexible with respect to functional form, not only controls for the levels of all variables included in the equation, but also treats each household as an individual observation rather than a member of a homogenized group, states relationships numerically (as regression coefficients), and measures the statistical significance of each coefficient, thus enabling us to distinguish between coefficients that may result from chance and those that reflect systematic relationships. Because regression analysis can handle large numbers of variables, we can study the effects of prices and expenditures as well as variables that describe household characteristics and others that represent different attitudes toward production for the market or different types of farming organization. One purpose of this research project was to see how effective single-equation multiple regression analysis could be in identifying variables that may effect consumption decisions and in measuring their effects. To be sure, a single-equation analysis will not disentangle all of the relationships involved. For that purpose we need econometric techniques which can deal with the system of equations as a whole. Results from such a systems analysis will be presented in the report to follow this one.

Our goals in this analysis are (1) to identify economic and non-economic factors that affect the quantities of food available for consumption, (2) to determine the effect of a change in any one of the independent variables on the quantity of any single food consumed by a household, holding the levels of the other variables constant, and (3) to obtain behavioral regressions that can predict the quantities of foods consumed by a typical household, given the values of the independent variables. To some extent these goals conflict. The coefficients of the best predicting equation will probably not be the least biased set of coefficients available, so that may not be the ideal regression for use if the second goal is preeminent. Yet unbiased parameter estimates are not always best for meeting the second goal. If the price of having an unbiased estimate is a large standard error, a certain amount of bias in the parameter estimate may be acceptable.

Given that goals may differ among readers, we present more than one regression for some foods. The user may choose the one best adapted to his needs.

Single-Equation Estimation

In this report, we present single-equation estimates of food consumption relationships for 14 foods and 6 groups of foods. Fitting each equation separately has advantages, even though it is inefficient in a statistical sense because it fails to use all the information available.

Food consumption decisions are interdependent; no single food decision can stand alone. (The equations for the different foods include many of the same variables.) Furthermore, disturbances are correlated across equations. Implicit in such a situation are two kinds of information that can be used with systems estimation but cannot be used in single-equation regression analysis: (1) cross-equation parameter restrictions that exist because all demand equations arise from a common utility function, and (2) correlations that exist among the disturbances across equations. Unfortunately, there is a trade-off between the use of these two kinds of information and the amount of detail concerning commodities and household characteristics that can be used. In single-equation estimation the number of parameters in the equation must, of course, be less than the number of observations (of households). In systems estimation the number of parameters in the system must be less than the number of households, and the number of parameters in the system is, at the least, a multiple of the number of commodities considered.

Given this fact, single-equation estimates are valuable. They are straightforward, inexpensive, and quickly obtained. They are easily understood and provide simple descriptive relationships that can readily be used to identify useful variables, including relevant socioeconomic characteristics, and groups of households that can effectively be set aside from the larger sample for separate analysis.

With single-equation estimates, we can have a great deal of commodity detail as well as narrowly defined independent variables. One can study many more commodities than would be possible either with tabular methods or with systems estimation. Single-equation estimates can also use effectively the information available from small numbers of observations, which tabular analysis cannot do.

Procedure

In general, we use only consuming households in fitting the food consumption regressions. We present the estimating equations in two forms: quantity equations, which estimate the quantity of a food available for consumption by the household; and share equations, which estimate the share of the total value of consumption devoted to the food. Quantity equations are also presented for specified groups of households in cases where the group behaves differently than the sample as a whole.

The procedure was to fit regression equations individually to each of 20 food categories, starting with a set of potential independent variables that we had reason to believe might affect food consumption. For each equation, a smaller set of variables was chosen from the potential set in such a way as to yield a regression equation that would minimize a predetermined measure of combined prediction error and bias.

While many forms are possible for the equations, we have chosen as our basic form a quantity equation that is homogeneous of zero degree in prices and incomes (thus embodying a characteristic of demand equations in static equilibrium theory). The independent variables enter arithmetically rather than logarithmically, so the values of the demand elasticities can vary with income.

CHAPTER II

VARIABLES

In this report, we examine 52 variables as determinants of the quantities of food available to the household. Two relate to total expenditure, 21 are price variables, and 29 measure characteristics of the household or the farm operation. Of the 14 variables employed in either of the two previous reports (Smith et al., 1979 and 1980), only population density and the number of adult male consumer equivalents are not included.

We replace household income, used in the earlier reports, by total expenditure, and we replace the quantity consumed per adult male equivalent (the dependent variable) by household consumption, unadjusted for household size or composition. Instead of using the number of consuming equivalents as an independent variable, we include household size (number of persons) plus variables for different sex and age groups within the household. Any measure of the number of consuming equivalents requires weighting the various household members in proportion to their importance as consumers, but all such weights are more or less arbitrary. In addition, the importance of a particular age-sex group as consumer differs among foods. In this report, each age-sex group constitutes a single variable, so the data themselves reveal how the group affects household consumption.

We include both region and ethnic group in our list of variables, although there is much correspondence between them. (The geographical distribution of ethnic groups is fairly well described by regional boundaries.) One purpose of the analysis is to determine which variable is the more useful or whether it may sometimes be well to use both. (Despite their high degree of correspondence, they do measure different things.) We also experiment with alternative methods of measuring production for the market. One of the variables used in the previous report (Smith et al., 1980) was the percentage of total value product that arises from certain specified activities. Such a broad classification may not be the best measure for any single commodity, but in the cross-tabulation analysis more narrowly defined classes would have contained too few households to be helpful.

Dependent Variables

Table 1 (see page 12) lists the variables. In the quantity regressions, the dependent variable will be the quantity of a specific food available for consumption in one year by an individual household.

Multiplying each quantity by its average price gives the "expenditure" on that food. In the share equations, the dependent variable is the share of total annual expenditure that the household devotes to a specific food where total expenditure is defined as the total value of food and non-food purchases plus the value of food and non-food consumption provided by the household from home production.

The Quantity Measures

In the tabulation analysis, quantities "consumed" were simply the sums of purchased quantities plus those available from home production. In the case of alcoholic beverages, however, treating a kilogram of palm wine, low in alcoholic content, as equivalent to one of omole (native gin) was not the best procedure. (It gave us a good measure of the total water consumed in these forms, however.)

In the regression analysis, the quantity consumed by each household is represented by an adjusted kilogram figure which takes into account the fact that market and home produced goods have different properties just as do onions and tomatoes or the components of any group of foods. For each food, the adjusted consumption quantity (q) was calculated as follows for each household:

$$q = \frac{q_h \cdot p_h}{p_a} + \frac{q_m \cdot p_m}{p_a},$$

where q_h and q_m, respectively, are the quantities the household produced at home and purchased from the market, p_h and p_m are the average home and market prices for those foods in the ecological zone where the household is located, and p_a is a weighted average of p_h and p_m. This definition of quantity consumed is consistent with theory, for $q.p_a$ equals the expenditure on the food being considered.

For a food group such as fruits or "other legumes," the procedure was the same, replacing $q_h \cdot p_h$ by $\sum_i q_{hi} \cdot p_{hi}$ and $q_m \cdot p_m$ by $\sum_i q_{mi} \cdot p_{mi}$, and summing over the i foods in the group, with p_a being the average price for the group of foods.

The Foods Studied

To understand thoroughly the nutritional situation of a household, one must study individual foods rather than broadly defined groups of foods. Important nutritional contributions are often rendered by individual items neither consumed in large quantities nor well represented by the remaining members of the conventional food groups frequently used in analysis. To study a diet in this detail can be expensive, but single-equation regressions make it feasible to examine at reasonable cost almost as many individual foods as one might wish. Hence, we present regressions for 14 individual foods: rice, cassava, palm oil, and groundnuts--examined by cross-tabulation analysis in our previous report (Smith et al., 1980), plus dried fish and nine other foods. (See the list of price variables in Table 1.) We treat dried fish as a single food even though the category is comprised of many species; we do the same for fresh fish. The foods considered include any food consumed by at least 61 percent of the households, in addition to palm wine (consumed by 37 percent of the households) and broadbeans (consumed by 38 percent). The broadbean is the most widely consumed legume other than the groundnut.

We also present regressions for six groups of foods--Other Cereals, Other Legumes, Vegetables, Fruits, Salt and Other Condiments, and Alcoholic Beverages.[1] These allow us to compare the results of analysis by groups and analysis by individual foods. (Except for Fruits, each food group contains at least one food that is also studied separately.) The broader definitions provide more complete coverage of the total set of foods consumed and include more consuming households within each group. The gains in coverage may be deceptive, however. It can be questioned whether the behavior of the group as a whole adequately represents the behavior of any of its components, particularly when the items in the group are diverse in character.

---

[1]Only Other Cereals and Alcoholic Beverages were examined in Smith et al., 1980.

TABLE 1

THE VARIABLES

Symbol	Definition
	Dependent
Food name - last two letters of the price variables below	Quantity of a specific food consumed by household (kilograms per year)
...	Household expenditure on a specific food (Leones per year)
...	Share of total annual household expenditure devoted to a specific food (expressed as a decimal)
	Independent
TEXP	Expenditure - total expenditure by household (Leones per year)[a]
$TEXP^2$	Expenditure squared
	Price (Leones per kilogram) -
PRB	Rice
PCA	Cassava
PPO	Palm oil
PDF	Fish, dried or tinned
PNF	Non-food
PGN	Groundnut[b]
POC	Other cereals (all cereals except rice)
PFF	Fish, fresh, frozen or iced
PSG	Sorghum
PBN	Broadbean
PON	Onions
PPC	Peppers and chillies

[a]In 1974/75 Le 1.00 equalled U.S. $1.10 [Spencer and Byerlee, p. 24].

[b]Peanut.

TABLE 1--Continued

Symbol	Definition
PSL	Salt
PMG	Maggi cubes[c]
PKL	Kola nut[d]
PPW	Palm wine[e]
POL	Other legumes (all legumes except groundnuts)
PVG	Vegetables
PFT	Fruits
PCN	Salt and other condiments (salt, sugar, Maggi cubes and condiments, unspecified)
PAB	Beverages, alcoholic
	Household characteristics
	Size and composition
SIZE	Size (number of persons)
INF	Children aged 0-5 years (number)
YCH	Children aged 6-10 years (number)
CH	Children aged 11-15 years (number)
MAD	Males aged 16-65 years (number)
FAD	Females aged 16-65 years (number)
...	Persons over 65 years (number)
DEPR	Dependency ratio [(number of persons aged 0-15 years and over 65) ÷ (number of persons aged 16-65 years)]
WIV	Wives (number)
AGEHD	Age of household head (years)

[c]Bouillon cubes, commonly referred to by the brand name, "Maggi".

[d]A stimulant, often used on ceremonial occasions.

[e]Made from the sap of certain palm trees.

TABLE 1--Continued

Symbol	Definition
	Other
	Ethnic group or area[f]
LIMB	Binary variable = 1 if Limba
TEMN	Binary variable = 1 if Temne
...	Each binary variable = 0 if member of the remaining group
	Region
REG1	Binary variable = 1 if Southern
REG2	Binary variable = 1 if Northern
...	Each binary variable = 0 if Eastern
	Production characteristics
	Percentage of the value of output plus labor sold out derived from
SHOOPT	Onions, peppers and chillies, and tomatoes
SHOCC	Cocoa and/or coffee
SHOLSO	Labor sold out
SHOSS	All specified sources (the three above plus oil palm products and non-farm activities including fishing)
SHLUR	Percentage of total labor devoted to upland rice

[f]The households are divided into three groups. One consists of 16 Limba households, a second of 31 Temne households, and a third of 83 Mende, 1 Loko and 7 Temne households. The 83 Mende households constitute 60 percent of the total sample.

TABLE 1--Continued

Symbol	Definition
	Market orientation
MKTOR	Total sales as a percentage of value of total output (not including the value of labor sold out)
SHCPH[g]	Percentage of household consumption of a specific food that is produced by the consuming household

[g]This represents nine variables, one for each of the foods for which it was calculated.

For a few foods (cassava, peppers and chillies, onions, and salt), data reliability is below the standard met by the other items. Quantity measurement for cassava is notoriously difficult. Harvested quantities are extraordinarily hard to measure accurately in physical terms, partly because cassava is harvested in small quantities as it is needed and partly because the weights of the units in which it is handled vary greatly. Similar problems exist with purchases from the market, where cassava is sold by the root and by the "heap."

Both salt and "peppers and chillies" are poorly defined categories. "Peppers and chillies" includes both dry and green peppers, while "salt" includes two distinctly different types--relatively expensive imported dry salt and relatively inexpensive locally produced wet salt. Among the food groups, vegetables and fruits also consist of foods for which accurate quantity measures are difficult to obtain. In each of these groups, sales by the piece or the heap were common, so converting these units to reasonable estimates in terms of kilograms was difficult.

Independent Variables

These consist of total expenditure and prices (the conventional economic variables), variables describing household characteristics (also called demographic variables), those having to do with the productive organization of the household, and a set of variables intended to measure the extent to which the household relies upon the market.

Expenditure

We use total household expenditure as a measure of the capacity for consumption that the household possesses. Total expenditure, the value of food and non-food purchases plus the value of food and non-food consumption provided from home production, correlates more closely with consumption decisions than does income. (It must, for total expenditure is the total of the individual consumption "expenditures.") Income figures vary more from year to year and from household to household than expenditures do; expenditures approximate more closely the "permanent income" to which the income concept in consumption theory normally corresponds.

There is also a practical consideration stemming from the nature of our data (a consideration relevant to many surveys). Total expenditure consists of the market expenditure for goods and services purchased for consumption plus the value of all home-produced consumption. While the market expenditure figures cover the period from May 1974 through April 1975, the data for food produced at home are based essentially on harvests that began in the fall of 1974. Market expenditures prior to the harvest season (October or later for most crops) were affected by the income earned during the previous harvest season, but we have no data on that income. The sum of market expenditures made between May and the 1974-75 harvest is clearly our best estimate of the income available for expenditure during that period.

The consumption of home-produced food between May and the 1974-75 harvests also depended upon the harvests of the previous year, but we have no reliable data on food inventories at the beginning of May, so we estimate the home-produced food available for consumption by subtracting total sales from May 1974 through April 1975 from the 1974-75 harvests. The quantity consumed during the pre-harvest months is estimated as equal to the quantities reserved from the 1974-75 harvest for use or sale in May 1975 and following. Our estimates of food availability are conservative, for rainfall during the 1974-75 crop season was later and of shorter duration than usual. Upland rice was affected more than swamp rice, as the swamps are generally planted later and hold their moisture longer (Spencer and Byerlee, 1977, p. 54).

Prices

The prices used are those of individual foods or groups of foods plus one that represents all non-food purchases. The price associated with the total annual consumption of each food is a value-weighted mean of quantity-weighted average prices for purchased goods (which we shall call market goods) and home-produced goods.

Prices for market goods (p_m) and for home-produced goods (p_h) are based upon household data for purchases and sales. In each case, an average price was calculated for each of eight ecological zones (combining zones when the number of households buying or selling a particular commodity was small). Had individual prices been computed for each household they would have exhibited spurious variation. For instance, suppose each household in a region faced the same seasonal pattern of prices but, because of different incomes, production patterns, and household characteristics, chose to sell or purchase at different points in time. Household-specific prices (averaged over one year) would then be different even though all faced the same prices. Worse yet, from a statistical point of view, these prices would be partially determined within the consumption decision-making process and would be endogenous to it. Using them as independent variables in demand analysis would yield biased estimates.

The price of the i^{th} market good in a given ecological area is a quantity-weighted mean of purchase prices paid by the households in that area:

$$p_i^m = \frac{\sum_{jk} p_{ijk}^m q_{ijk}^m}{\sum_{jk} q_{ijk}^m} ,$$

where p_{ijk}^m is the price paid by the j^{th} household in the k^{th} transaction for the purchase of q_{ijk}. This is, in short, the aggregate value for the area of all purchases of the i^{th} good divided by the aggregate quantity purchased (in kilograms). The sample used consisted of the 141 households in the expenditure sample.

The prices of home-produced goods are mean farm gate sales prices for the ecological zone, quantity weighted as above. They were calculated from the farm sales data for each commodity, using the full production sample of 328 households. The total value sold annually was divided by the total quantity sold.

The price, p_i^a, of the total annual consumption of a given food in a given area, is a value-weighted arithmetic mean of its sales and market purchase prices:

$$p_i^a = \frac{p_i(p_i^m q_i^m) + p_i^h(p_i^h q_i^h)}{p_i^m q_i^m + p_i^h q_i^h} ,$$

where p_i^m and p_i^h are the average market and farm sales prices for the area, $p_i^m q_i^m$ is the total value of area consumption of food bought from the market, and $p_i^h q_i^h$ is the total farm gate sales value of food consumed from each household's own production. The price, p_i^a, of the i^{th} food can also be described as a weighted average, where the weights are the shares of the total value of the i^{th} food consumed in a given ecological zone--the shares from the market and from home production. The share that was spent in the market is

$\frac{p_i^m q_i^m}{p_i^m q_i^m + p_i^h q_i^h}$, and the share consisting of the sale value of food produced at home but not sold is $\frac{p_i^h q_i^h}{p_i^m q_i^m + p_i^h q_i^h}$.

Where several foods are combined into a single group (aggregated) the average price for the group in a given ecological zone is calculated in the same way, except that both market purchases and home-produced consumption now must be totalled over all the foods in the group:

$$p_i^a = \frac{\sum_i p_i^m (p_i^m q_i^m) + \sum_i p_i^h (p_i^h q_i^h)}{\sum_i p_i^m q_i^m + \sum_i p_i^h q_i^h} .$$

Here p_i^m and p_i^h are the average market and farm sales prices, respectively, of the i^{th} food in the ecological zone, $p_i^m q_i^m$ is the value of total market purchases of the i^{th} food by households in the relevant ecological area, and $p_i^h q_i^h$ is the total value of the i^{th} food produced at home.

Calculating a price for all non-food expenditures poses a different problem. We did not have the information needed to convert all quantity data into kilograms, but we did have quantities purchased and expenditures for an exhaustive set of non-food items. For each important class of items, we chose one or a few commodities that represented the bulk of the value purchased. Then, choosing the quantity unit in which the majority (usually overwhelming) of transactions was carried out, we calculated an annual price for each ecological zone. Average prices for the class were then formed for each zone, using values purchased in the zone as weights. Finally, the various classes of non-food purchases were aggregated using the values of area expenditure on each class as weights for the area.

The commodities for which prices were calculated are listed in Table 1. A minor note: in Smith et al., 1980, sugar was not included with "salt and other condiments."

Household Characteristics

The variables describing household characteristics primarily concern the size and composition of the household, but some take note of ethnic group and location (region).

SIZE is the number of persons in the household. How these persons are distributed by age and sex is shown by a set of five variables that together include all persons aged 65 or less. No variable is included for persons over 65 because the number in this age group is already determined when the values of SIZE and the five other age-sex variables are known. If one independent variable is a linear combination of other independent variables, the regression computations become impossible because the matrix of independent variables becomes singular.

These size and composition variables also define the size of the farm in Sierra Leone. Land availability there is rarely a limitation on farm size, but quantities of particular types of land are clearly important in determining the type of farming activity followed.

The dependency ratio is a measure of household composition that is independent of household size. It is included to test the hypothesis that this single ratio is as effective an indicator of household consumption as the set of variables by age and sex, even though the set of variables gives more flexibility and recognizes that the relevant features of household composition are not necessarily the same for all foods.

Other variables that may influence food consumption behavior are the age of the household head and the number of wives he possesses. Note that the latter variable includes only wives of the household head; other wives in the household (i.e., wives of sons, brothers, or uncles) are not included.

The principal ethnic groups in Sierra Leone are the Mende, the Temne, and the Limba. The data for about half of our households specify the ethnic group of the head of the household; for the remainder, we relied upon the 1963 census, which listed the ethnic group for each location from which our sample was drawn. In every case where the datum for an individual household exists, the ethnic group reported agrees with the 1963 census, so we feel on safe ground. Any Susu households are included with the Temne group. The Limba and Temne groups are represented by binary variables.[1] (The value is one if the household is in the third group, which has 83 Mende, 7 Temne, and 1 Loko household.[2]

---

[1]Such binary variables are often called "dummy" variables because they are artificial variables, used as "dummies" or stand-ins to represent certain attributes or categories that may be important even though they have no natural numerical measurement or the level of the variable is unimportant except as it indicates the category.

[2]In Smith et al., 1980, the Loko household was included with the Limba group.

Households are also identified by region. Region 1 corresponds roughly to the Southern Province; Region 2 to the Northern. Each is represented by a binary variable which assumes the value of one when a household is in the specified region. If each of the regional variables has a value of zero, the household is in Region 3, the Eastern Region (approximately equivalent to the Eastern Province exclusive of the diamond mining area).

Although there are three ethnic groups to be considered, the variable for any one of them can be written as a linear combination of the other two, so only two can be used in any one equation. The same situation prevails with respect to the three regional variables. Hence, no symbol is assigned to the third variable in each of these cases.

Production Characteristics

The variables relating to production characteristics fall into two classes: one characterizes the production pattern by the value of output; the other by the allocation of labor time. The first class identifies certain products or activities likely to be directed toward the production of money income rather than food, the value of the output in each case being expressed as a percentage of the total value of output produced by the household, including labor sold out. Onions, peppers, and tomatoes (when produced in any volume) are for sale to the urban market (in particular, to Freetown), so SHOOPT is one of these variables. Another is SHOCC, the share from cocoa and/or coffee (both produced for export). SHOLSO (labor sold out) identifies households that obtain relatively large amounts of money income by providing labor services to other households. The fourth of these variables, SHOSS, represents a broad range of activities encompassing the three above, the production of oil palm products, and all non-farm activities, including fishing.[1] This variable identifies a group of households with this type of orientation, but does not identify a specific type of farming organization. (Producing oil palm products may be a questionable inclusion, for this activity produces both food and cash income.) While SHOSS provides less specific information than the three that preceded it, its counterpart (PCTOUT) was informative in some of the cross-tabulation analyses, so we have tried it here as an alternative to the more specific variables--but only for two foods, cassava and palm oil.

---

[1]In Smith et al., 1980, the comparable variable (PCTOUT) excluded fishing (pp. 36-37).

The percentage of labor devoted to upland rice, which also was useful in the cross-tabulation analysis, has likewise been tried here, and in the same two regressions. The variable can be useful for testing the hypothesis that the household producing upland rice has a better diet than households practicing other types of rice culture. (In contrast with rice grown under other systems of cultivation, upland rice is normally intercropped with other food crops.)

Market Orientation

The variables measuring the percentage of the total value of output produced (plus labor sold out) that comes from activities directed toward producing money income indicate both a certain type of farming organization and orientation toward the use of the market as an alternative to producing food for home consumption. We have, in addition, two variables that measure other aspects of market orientation. MKTOR is the ratio of the value of output sold to the value of total output produced (not including the value of labor sold out). It too measures the degree to which a household directs its activities toward obtaining money income through sales in the market, but has the advantage of not being tied closely to the production of any particular crop.

One last variable measures the extent to which a household is not market oriented with respect to the provision of each of nine of the food items under consideration. The one symbol, SHCPH, actually represents nine variables--one for each of the foods for which the datum was calculated.

CHAPTER III

EQUATIONS USED

Two major decisions are required concerning the equations to be used: (1) the functional form to use, and (2) the variables to include in each equation. Our approach has been to begin the analysis with straightforward single-equation regressions that can describe the influence of a wide range of potentially relevant variables, proceeding at later stages to more complicated analyses which may distinguish more effectively among the several mechanisms operating but which are less capable of handling large numbers of variables and many individual foods.

Functional Form

All the single-equation estimates derive from the following model:

$$q_i = f(y, p, h, v, r), \qquad (1)$$

where q_i, the annual quantity of the i^{th} food available for consumption by the household, is a function of y (household expenditure), p (a vector of prices), h (a vector of household characteristics), v (a vector of variables identifying certain types of production activities, and r (a vector of variables describing the relationship of the household to the market). Two major classes of regressions will be fitted: quantity regressions, with q_i as the dependent variable; and share equations, where the dependent variable, $\frac{p_i q_i}{y}$, is the share of total annual expenditure devoted to q_i. For the quantity equations and, usually, the share equations, we only use the data for the consuming households when fitting the regression for a specific food or group of foods. Some quantity equations are fitted to data from subgroups of households (i.e., households grouped by income or by region).

The Quantity Equations

For these, we use a functional form that is homogeneous of zero degree in the expenditure and price variables, in accordance with a standard result of the conventional economic theory of consumer choice. The function is linear in h, v, and r. The latter operate as shift variables, adjusting the average predicted relationships between quantity and the price and expenditure variables for shifts in the utility function, associated with the household characteristics variables (h), or for differences in production or market opportunities (or choices), reflected in the v and the r variables. Stated algebraically:

$$q_i = \sum_j a_j(p_j/p_i) + b_1(y/p_i) + b_2(y/p_i)^2 + \sum_k c_k h_k + \sum_m d_m v_m + \sum_n e_n r_n \, , \tag{2}$$

where q_i is the quantity of the i^{th} food consumed by the household, p_i and p_j are the respective prices of the i^{th} and j^{th} foods, y is the total expenditure variable for the household, the h_k, v_m, and r_n are the elements of the vectors h, v, and r, and the a_j, b_1, b_2, c_k, d_m, and e_n are the regression coefficients.[1] Doubling each price and expenditure variable has no effect on the quantity consumed. Isolating the term in the relative price of the i^{th} food leads to (3):

$$q_i = a_i + \sum_{j \neq i} a_j(p_j/p_i) + b_1(y/p_i) + b_2(y/p_i)^2 + \sum_k c_k h_k + \sum_m d_m v_m + \sum_n e_n r_n \, . \tag{3}$$

As the relative price of the i^{th} food is always unity ($p_i/p_i = 1$), its regression coefficient, a_i, appears as the constant term in (3), and the own-price variable does not appear explicitly as an independent variable. Its influence on quantity operates through all the relative price and expenditure variables, as well as the constant term.

If we drop the terms in v and r, (3) becomes a conventional demand regression for a household that receives all its income in money and buys all its goods in the market. Its selection of goods depends upon market prices and the amount of income, but not upon the form in which income is received or how it is produced. To be sure, new car salesmen buy different clothing than construction workers, but such differences usually enter the theory only as differences in the utility functions. But when a household produces much of its own food, many individuals (perhaps mostly non-economists) believe that both the amount and the form of income affect food consumption choices. When a household produces part of its food, consumption decisions are affected by

---

[1]The regression coefficients are specific for the i^{th} food; they change from food to food.

both production and market purchase opportunities. In addition, control over the allocation of income may reside in different hands, as when the male household head controls the expenditure of the income from cash crops but the women of the household allocate some or all of the food crops.

To test the hypothesis that the form or source of income matters, we include the terms in v and r. If the hypothesis is correct in a least squares demand regression that ignores the form or source of income, the coefficients are biased[1] whenever the regression is fitted to data from households that produce significant portions of their own food.

Equation (3), useful as a test of the hypothesis that production characteristics or decisions affect consumption choices, is not a demand regression in the sense of a regression that concerns only responses that occur on the demand side of the household's calculations (under the assumption that the form or source of income does not matter). It is a behavioral regression, which predicts the net effect on consumption of both production and consumption responses to the situation faced by the household. This is what is required by the student of food consumption and nutritional well-being, but the single-equation estimate is not the sharpest tool for sorting out clearly all the interconnections among the variables.

<u>The Share Equations</u>

Multiplying each term in (3) by p_i yields the expenditure regression:

$$q_i p_i = a_i p_i + \sum_{j \neq i} a_j p_j + b_1 y + b_2 (y^2/p_i) + p_i(\sum_k c_k h_k + \sum_m d_m v_m + \sum_n e_n r_n). \qquad (4)$$

The expenditure equation passes through the origin. (It has no constant term.) The expenditure for q_i is zero for values of the independent variables that make $q_i = 0$.

Dividing (4) through by the total expenditure, y, yields (5), the share equation for the i[th] food:

$$\frac{q_i p_i}{y} = \frac{a_i p_i}{y} + \sum_{j \neq i}\left(\frac{a_j p_j}{y}\right) + b_1 + b_2 \frac{y}{p_i} + \frac{p_i}{y}(\sum_k c_k h_k + \sum_m d_m v_m + \sum_n e_n r_n). \qquad (5)$$

---

[1]Because relevant variables have been omitted from the regression.

The constant term in (3), a_i, appears here as the coefficient of the price variable for the i^{th} food. The constant term in (5), b_1, corresponds to the regression coefficient for the linear expenditure term in (2), y/p_i, where expenditure is expressed in real terms as the capacity to purchase the i^{th} food--the number of units of the i^{th} food that could be purchased if all expenditures were used for that purpose. Regression coefficient b_2 in (5) corresponds to the coefficient of the squared expenditure term in (2), $(y/p_i)^2$.

However, estimates of b_1, b_2, or the other regression coefficients will not usually be the same whether obtained from (2) or from (5). Whereas the ordinary least squares regression for equation (2) is fitted by minimizing the sum of the squared deviations of the observed from the predicted values of q_i, that for equation (5) minimizes the sum of the squared deviations of the observed from the predicted values of $(q_i p_i)/y$. Because different measures are being minimized, and each one has a random component, the parameter estimates will usually be different. Moreover, the influence of the observation for a single household differs in the two forms: in the share equation form, households for which y is large have less influence than in the quantity equation form.

Still another difference exists between the two equations as we have used them, for we have not permitted certain variables to enter the share equations that were available for the quantity regressions.

We calculated share equations as part of the exploratory work required to prepare for the systems estimation. We needed to know which demographic variables would be the more important in a model without the v and r variables (the variables relating to production activities and market orientation). When fitting regressions without those variables, we chose the share form because it seemed probable that heteroskedasticity[1] would be less troublesome in the share form.

As it turned out, both the quantity and the share regressions suffer from heteroskedasticity. Tests on the residuals indicate that the heteroskedasticity in the quantity form can be dealt with by fitting weighted least squares regressions, using 1/q as the weight, where q is the quantity predicted for each household by ordinary least squares. An analogous procedure would

[1] A situation in which the variances of the error terms are not the same.

have been effective for the share equations. No weighting was done, given time and money limitations. However, the unweighted regressions give unbiased point estimates of the parameters, even though they do not yield the most accurate measures of their sampling variation that are possible.

Alternative Forms

Many other functional forms could have been used, or we could have experimented with a different functional form for each commodity regression. Yet, even though we calculate each single-equation estimate independently, each is part of a set of relations derived from the same utility function. Under these circumstances, using the same form for each regression is reasonable.

One alternative to the linear equations used here, the log-log equation, is often used with considerable success, but the form has two disadvantages that make it inappropriate for our purposes. First, this form forces the price elasticities to be constant with respect to income, but we want to know whether price elasticities change with income, and if so, how they change. Secondly, the log-log equation cannot handle negative values of the dependent variable. To be sure, quantities consumed cannot be negative, but estimates of them can. Estimates of household consumption are affected by both positive and negative errors. In some cases, negative errors are large enough to create negative estimates for the consumption of particular commodities by individual households. To eliminate those negative terms from the data without at the same time removing equivalent errors on the positive side (errors that cannot be identified with any degree of accuracy) would change the error distribution and bias the estimates of the regression coefficients. The arithmetic form of the equation can be used without requiring us to tamper with the distribution of the errors in the variables.

In summary, we use equation (3), homogeneous of zero degree in expenditure and prices, to estimate the quantity regressions. Price elasticities are free to vary with income as the data may indicate, and the equation is easily converted into the expenditure and share forms. Also, these single-equation estimates have easily understood analogs in the components of the systems estimation to be presented later.

Elasticities

The expenditure and pricė elasticities from these equations vary with expenditure as well as with prices and, in the case of the own-price elasticity, with variables other than price and expenditure. The own-price elasticity, given for the i^{th} food, is:

$$\frac{\partial q_i}{\partial p_i}\frac{p_i}{q_i} = -1 + [a_i - b_2(y/p_i)^2 + g]/q_i, \text{ where } g = \sum_k c_k h_k + \sum_m d_m v_m + \sum_n e_n r_n.^{1} \quad (6)$$

The own-price elasticity will be independent of total expenditure and equal to -1 if a_i, b_2, and g are equal to zero. The own-price elasticity will be independent of expenditure, but not necessarily equal to -1, if b_1 and b_2 are equal to zero, for in that case the income term in the enumerator of the second term of the expression will have a value of zero and the q_i in the denominator will itself be independent of income. The value of the constant term in the regression for equation (3), a_i, is important in determining the value of the own-price elasticity, but neither its magnitude nor its sign is related in a simple way to that own-price elasticity.

The cross-price elasticities are:

$$\frac{\partial q_i}{\partial p_j}\frac{p_j}{q_i} = \frac{a_j}{p_i}\frac{p_j}{q_i} = \frac{a_j}{q_i}\frac{p_j}{p_i}. \quad (7)$$

They vary with total expenditure whenever q_i does (when b_1 and b_2 are not equal to zero).

The income (expenditure) elasticity for the i^{th} food is:

$$\frac{\partial q_i}{\partial y}\frac{y}{q_i} = [b_1 + 2b_2(y/p_i)]\frac{y}{p_i q_i}. \quad (8)$$

---

[1]From (3), we have $\frac{\partial p_i}{\partial p_i} = \frac{-1}{p_i}[\sum_{j\neq i} a_j \frac{p_j}{p_i} + b_1\frac{y}{p_i} + 2b_2(y/p_i)^2] =$

$-\frac{1}{p_i}[q_i - a_i + b_2(y/p_i)^2 - g]$. Hence $\frac{\partial q_i}{\partial p_i}\frac{p_i}{q_i} = -[\frac{q_i - a_i + b_2(y/p_i)^2 - g}{q_i}] =$

$-1 + [\frac{a_i - b_2(y/p_i)^2 + g}{q_i}]$.

It too is a function of income unless b_1 and b_2 both equal zero or $p_i q_i$ is equal to ky^2 for constant p_i.

A representative elasticity with respect to the other variables is given by $\frac{\partial q_i}{\partial h_k} \frac{h_k}{q_i} = c_k \frac{h_k}{q_i}$. This is also a function of income whenever q_i varies with income.

The Variables Available

In principle, the number of variables that affect decisions about the quantities of foods to be consumed is limited only by the curiosity of the investigator. In practice, considerations of feasibility arise--we ask ourselves how much time and money are really worth spending on experimentation with variables that have some plausible connection with the consumption decision. In this case, we set an upper limit of 27 (the maximum number that could be handled by the computer program we planned to use) upon the number of independent variables to be made available for use in any one of the quantity equations. The variables fell into three classes: price and expenditure, household characteristics, and those relating to the source of income.

If a household must allocate a fixed monetary income among many consumption goods, economic theory concludes that income (or total expenditure) and the prices of all goods are relevant variables. We include total expenditure and its square plus the prices of rice, cassava, palm oil, dried fish, and non-food goods as variables available to each of the food consumption regressions. The list includes the prices of the four most important widely-consumed foods in rural Sierra Leone. In addition, each food consumption regression includes as an available independent variable the price of that specific food (the own-price variable) and the prices of such other foods as one would expect to be rather closely related in consumption to the dependent variable. The most frequently used of these additional prices is the price of groundnuts, but the prices of fresh fish and of "other cereals" also appear in a number of equations.

The variables relating to household characteristics--size, composition, ethnic group, and region--identify influences that may affect the utility function of the household. Variables relating to size and composition represent household members' physiological needs for food and the effects of any consumption preferences (food or non-food) that may differ by age and sex among the subgroups that comprise the household. These variables also represent the amount and type of labor available within the household.

Ethnic group and region represent differences in customs and taste, differences in ecological characteristics, or differences in the economic opportunities available (including access to the market, to saltwater or freshwater fishing locations, and so forth). The entire set of household characteristics variables was included in the available set for each of the quantity regression equations.

As we have already indicated, students of food consumption behavior often argue that the quantity and quality of food that a household consumes is affected by the source of household income as well as by its amount. The economist, in contrast, often argues that if the time and effort spent in earning the income is held fixed, only the amount of income affects the consumption decisions made at any given set of relative prices. A partial explanation of these different points of view lies in the fact that noneconomists examining food consumption behavior frequently do not make adequate observations of incomes and relative prices, and that economists tend to arrive at their conclusions by using a theory that assumes perfectly competitive markets, a clear distinction between production and consumption decisions, and a household that can be thought of as an integrated decision-making unit.

As we have said before, the decision to consume food produced at home is likely to be affected by both the production and the consumption opportunities available. Furthermore, the kind of production chosen (for market or for home consumption) may alter the locus of consumption decisions within the household and thus the nature of those decisions. To test the hypothesis that the source or form of income has an effect on food consumption choices, we include several variables relating to source of income.

In general, these variables fall into two categories: (1) production characteristics--the type of production activity, and (2) market orientation--the extent to which (a) crops are produced for the money income they provide, or (b) the household relies upon the market as a source of food. Three variables identify the extent to which a household engages in certain activities often chosen primarily, if not exclusively, as sources of money income. Each measures the share of the value of total output plus labor sold out that is obtained from a single activity: (1) SHOOPT--the production of onions, peppers, and tomatoes (if on a large scale, this output is normally intended for sale in urban markets); (2) SHOCC--the production of cocoa or coffee; and (3) SHOLSO--labor sold out for use by other households. These three variables do not comprise all activities engaged in primarily for money income, but they are examples that allow us to examine the hypothesis of interest. They are included in the available set for each food consumption regression.

In two regressions, those for cassava and for palm oil, we also use a more inclusive variable, which is SHOSS--the share of the value of output plus labor sold out which is derived from the three specific sources identified above plus the production of palm oil products and/or any non-farm activity, including fishing. Both the production of oil palm products and fishing are activities that may or may not be primarily devoted to the provision of money income, but when either of these comprises an unusually large fraction of the value of the output of the household, we may reasonably conclude that money income was an important objective.

SHLUR--the share of household labor that is devoted to the production of upland rice--characterizes the type of farming activity from a different point of view. This variable is of interest because intercropping is commonly associated with the production of upland rice. Again, we have experimented with the variable only in the equations for cassava and for palm oil.

The previous five variables distinguish among households in terms of potentially relevant characteristics of their cropping patterns. The first four identify households that apparently have a particular interest in the production of money income, but they do not necessarily identify all such households. A measure of market orientation that applies to all households, but gives no specific information with respect to type of activity, is MKTOR--the total value of sales as a percentage of the value of total output, including the output from non-farm activities. Income from labor sold out or from trading activity is not included in either the numerator or the denominator of this fraction.

The last variable, the share of the household consumption of a given food commodity that the household itself produces, approaches market dependence from a different point of view. In this case, we measure the extent to which the household is free of dependence upon the market in obtaining the food it consumes.

Permitting 27 variables to be available for use in a given commodity regression may be regarded as testing the hypothesis that each variable affects the quantity of food consumed. The test is not as sharp as one would like because in some cases several variables are alternative measures of the same underlying factor. In these cases, the data will determine which of the alternative measures are the more useful as predictors of food consumption.

Multicollinearity

Variables were selected for possible use in each equation on theoretical grounds, as explained above. It turned out, however, that for each commodity at least one variable was an almost exact linear combination of other variables in the set--the multiple correlation (R^2) between this variable and that combination exceeded 0.9999. In this situation, at least one variable had to be deleted if the necessary matrix inversion operation were to be carried out satisfactorily.

The variables most commonly identified as being substantially linear combinations of the other variables were Region 1, Region 2, and the prices of palm oil, non-food, and cassava. Some of the multicollinearity exists because food prices are calculated for areas which are subdivisions of the regions. There can be at most eight different values for a single price variable, one for each ecological zone. Each region consists of a set of these zones, so it is not surprising that some combination of one of the price variables should exist that could replace the regional variable. If a regional variable is omitted in this situation, whatever influence the regional variable might have had can be picked up by an appropriate combination of variables that was not deleted. Similarly, if the palm oil price is deleted, the regional and other variables may pick up part of its influence.

Choosing the Final Set

While in principle each variable in the available set may contribute to an explanation of the dependent variable, in practice estimation is often improved by omitting the less useful of the potential independent variables. Thus the final regressions for each commodity contain only a selection of variables from the available set.

To obtain the optimal subset of variables, we use a procedure that minimizes C_p, the ratio of (1), the expected value of the sum of the squared differences between (a), the predicted values of the dependent variable when predicted from a specific subset of independent variables, and (b), the expected values of the dependent variables (which depend upon the entire set of independent variables), to (2), the variance of the disturbances when all variables are present.[1] That is, we minimize the expected value of $\frac{\Sigma(\hat{q}_i - q_i^*)^2}{\sigma^2}$ where $\hat{q}_i$ is the predicted value of q_i using the specified set of variables, q_i^* is the expected value using all variables in the available set, and σ^2 is the variance of the disturbances when all variables are present.[2] The summation is across households. It is as though we calculated predicted values of the dependent variable for each household from every possible combination of independent variables, compared each set of those predicted values with the expected values (q_i^*) obtained from the full set of variables and chose the regression for which the sum of the squared differences between these two estimates was the smallest. Dropping variables from the full set of independent variables reduces the variance of the $\hat{q}_i$ but introduces bias into the regression coefficients if the variables really belonged in the true model. By minimizing C_p, we choose a subset of independent variables that is optimal in the sense that the expected gain from reducing the variance of the predictions is balanced against the expected loss from having more bias in the regression coefficients. In general, the regressions we present will minimize the value of C_p.

If selecting a regression equation on this basis resulted in omitting variables particularly important for our study (the price or expenditure variables, perhaps), we turned to an alternative criterion, choosing the regression that maximized $\bar{R}^2$ (the adjusted multiple correlation coefficient)--a measure of the goodness of fit that is based upon the squared

---

[1] Computationally, $C_p = \frac{RSS_p}{\sigma^2} + 2p - n$, where p is the number of terms retained in a particular equation (including the intercept term if present), RSS_p is the error sum of squares for a particular regression, $\hat{\sigma}^2$ is the estimate of the variance of the disturbances when all variables are present, and n is the number of observations (Hocking, 1976, pp. 5, 10, 17-18; Gaver and Geisel, 1974, pp. 59-61). C_p was calculated by the Furnival and Wilson algorithm (1974), using the BMDP computer program.

[2] As σ^2 is always the same for a given dependent variable, we can think simply of minimizing the numerator, $\Sigma(\hat{q}_i - q_i^*)^2$.

differences between the actual values of the dependent variable and the values predicted by using the set of independent variables included in the regression equation under consideration. As a practical matter, minimizing C_p normally leads to the use of a smaller subset of variables, for it includes only variables for which $|t| \geq \sqrt{2}$ (where t is the ratio of the regression coefficient to its standard deviation). Maximizing $\bar{R}^2$ means including all variables for which $|t| \geq 1$.

One could justify selecting the regression that maximizes $\bar{R}^2$ on the ground that, on the average, the true model would be the one for which $\bar{R}^2$ would be the largest. Most of our choices, however, will be based upon the value of C_p. This means somewhat more bias in the regression coefficients but fewer variables with coefficients that are statistically insignificant.[1]

---

[1]The optimality properties associated with choosing a regression according to either the C_p or the $\bar{R}^2$ criterion follow in part from the assumption of homoskedasticity. Unfortunately, the data do not satisfy this assumption.

CHAPTER IV

QUANTITY REGRESSIONS: ENTIRE SAMPLE

In this study we examine the semi-subsistence household as a whole to test the hypothesis that production and consumption decisions in households producing large portions of their own food are so interrelated that no satisfactory understanding of food consumption choices is possible unless full account is taken of both production and consumption alternatives. This interrelationship is of particular importance because it deals with the effects of the production of non-food crops on household food consumption, or, more generally, the extent to which production choices or market orientation affect household food consumption levels.

In this report we use least squares single-equation estimation--a simple but powerful technique capable of handling a great many independent variables. Chapters V and VI present variations on this procedure, while Chapter VII presents the expenditure and price elasticities obtained from the regressions presented in all three of these chapters.

Non-Consuming Households

A major problem arises when estimating demand regressions for commodities defined as specifically as those in this study. For most commodities there will be households, sometimes many, whose consumption levels are zero, but zero values of the dependent variable can lead to biased estimates of the regression coefficients (see below). For the major food, rice, this is no problem; every household in the sample consumed rice. Almost all households consumed palm oil, dried fish, and salt. But for some foods, the number of non-consuming households was very large. The number of consuming households ranged from 138 for rice to 41 for broadbeans.

Non-consuming households create a problem because when the dependent variable has few if any negative values but many zero values, it is not well represented by the standard assumption that deviations are normally distributed. Instead, it is as though the lower tail of the normal distribution, which in principle extends to negative infinity, had largely been cut off at zero. Furthermore, the presence of non-consuming households creates a sudden jump in the frequency count at zero consumption level instead of a smoothly declining curve. Excluding non-consuming households removes this piling up of observations at zero, but still leaves us with a distribution that has lost part of its lower tail. As a consequence, estimates of the regression coefficients are likely to be biased. Tobit analysis yields unbiased estimates of the regression coefficients but becomes very expensive, both in time and money, if done for a large number of regressions. As a second best approach, we have excluded non-consumers of a particular food from the data used in estimating the regression for that food, admitting that this procedure involves some risk of biased coefficients.

To be sure, non-consuming households may have different utility functions than consuming households.[1] If there are no values of the independent variables for which a household would consume positive quantities of the commodity, eliminating those households would be satisfactory, for they are not a part of the population that interests us. Using the remaining households we would obtain a regression that describes the behavior of all households that might consume the commodity, which is what we are after. But probably some non-consuming households would appear as consuming households at some set of values of the independent variables; they ought not to be excluded, but we cannot tell which households they are.

Non-consuming households create fewer difficulties where the analysis of consumption purchases is limited to the study of broadly defined groups of commodities. If the group is defined broadly enough, the non-consuming household may nearly disappear from the data set, so the statistical problem disappears. Still, though one might have more confidence in the coefficients obtained, the behavior defined is behavior with respect to an artificial "average" commodity which is not necessarily representative of any commodity actually purchased by the household. The less homogeneous the commodity group the less clear it is what the regression coefficient describes.

The Regressions as a Group

Table 2 (see page 38) contains single-equation results for fourteen foods[2] and six groups of foods. (Some of the fourteen foods are included again within the groups.) The six groups of foods plus the single foods not included within them comprise almost the whole of the rural Sierra Leonean diet. We fitted no regression for animal foods other than fish because such foods constitute a very small part of the diet.

---

[1]The evidence in Smith et al., 1980, [pp. 72-74] could be taken as supporting either view.

[2]Dried fish and fresh fish are regarded as single foods, although each one consists of many varieties of fish.

Food consumption regressions were calculated for a sample of 138 households (900 persons).[1] Two of these households, the two largest, with 17 and 22 members respectively, are outliers in several regression equations. (The next largest, three of them, have 15 members each.) The dependency ratios for these two largest households are 1.83 and 1.20, not at all out of line with those for other large households. Despite their unusually large size, we include these two households because, the information we have about them appears valid.

In general, the results are quite good. The values of $\overline{R}^2$, the proportion of the variation in the dependent variable accounted for by the influence of the independent variables, adjusted for degrees of freedom, range as high as .76, with most values between .30 and .50. Most regression coefficients have plausible signs, although the number of negative signs on price coefficients (indicating that gross complementary exists) is larger than one would expect from households that purchase all their food. When a food is produced at home, however, what appears to be demand-side complementarity may reflect either a demand-side or a supply-side relationship and these may be opposite in nature and in effect.

Although these results are generally satisfactory, examination of the residuals suggests that heteroskedasticity is a problem. (One of the assumptions that must be satisfied if least squares estimation is to yield best unbiased estimates is the assumption of homoskedasticity--that the variances of the error terms are the same for all observations. That assumption is not fulfilled for the quantity regressions.)

The data suggest also that heteroskedasticity could be substantially reduced by using weighted regressions, weighting the data for each household by $1/\hat{q}_i$, where $\hat{q}_i$ is the predicted consumption of the i^{th} food for that household. We examined only the regressions for rice, fish, cassava and palm oil. In each case $1/\hat{q}_i$ appeared to be the appropriate weighting.

Because of the time constraint under which this project is operating we did not carry out this modification of the least-squares analysis. We suggest, however, that weighted regressions should be fitted if the single-equation estimates are to be used as a basis for policy decisions. We suggest also that in principle not all commodity regressions need be alike in this respect; the best weighting for one commodity may not be the best for another. Nonetheless all the commodities we have looked at are alike in this respect.

---

[1] Two households included in our tabular analysis [Smith et al., 1980] were excluded because data were not available on their non-food expenditures.

TABLE 2

THE QUANTITY REGRESSIONS

Commodity and Mean Consumption of Consuming Households (Kilograms)		Number of Consuming Households	$\bar{R}^2$	C_p	Constant Term	Independent Variables: Expenditure		Prices							
						TEXP	$(TEXP)^2$	PRB	PCA	PPO	PDF	PNF	PGN	POC	Other
(1) Rice 589	(1.1)	138	.599	7.02	-268.7 (-1.86)	.259 (5.24)	-.176E-4 (-3.02)	a	436.0 (1.04)	-108.3 (-1.93)	204.2 (3.43)				
	(1.2)	138	.603	5.98	-285.6 (-2.31)	.273 (5.63)	-.194E-4 (-3.46)	a			192.9 (3.95)		-141.5 (-2.13)		
(2) Sorghum 60		90	.355	6.72	115.6 (2.40)	.253E-1 (1.81)		202.2 (2.10)	-496.4 (-2.32)	b		-126.7 (-2.66)		c	PSG a
(3) Cassava 394	(3.1)	114	.529	6.20	-374.6 (-1.72)	.155E-2 (0.11)	-.136E-6 (-0.51)	-99.6 (-4.02)	a		-51.6 (-1.95)		122.2 (6.62)	26.0 (4.59)	
	(3.2)	114	.534	4.29	-371.1 (-1.73)		-.109E-6 (-1.08)	-99.0 (-4.11)	a		-52.1 (-2.01)		122.5 (6.75)	26.2 (4.94)	
	(3.3)	114	.496	-5.37	-91.2 (-0.58)				a			-63.2 (-5.46)	144.7 (6.91)	30.6 (8.10)	
(4) Palm oil 85	(4.1)	133	.481	6.62	-74.7 (-1.38)	.196 (6.52)	-.328E-4 (-4.75)			a	105.5 (2.11)	-269.1 (-4.31)	200.4 (2.33)		
	(4.2)	133	.444	5.96	-86.3 (-1.61)	.204 (6.59)	-.358E-4 (-5.04)		682.3 (2.05)	a	98.4 (2.07)	-304.4 (-4.05)	299.7 (3.52)		
(5) Groundnut[d] 81	(5.1)	103	.406	12.32	-235.2 (-3.35)		.782E-6 (0.85)		241.2 (1.35)	b	-96.7 (-2.47)	90.5 (2.09)	a	c	PFF e
	(5.2)	103	.379	5.58	-123.4 (-3.81)		.146E-5 (1.99)			b			a	c	PFF e

NOTE: With a few exceptions, the regressions presented here were chosen from all possible subsets of the available set by minimizing C_p (where p is the number of terms in a given equation) or maximizing $\bar{R}^2$. The underlined value in the $\bar{R}^2$ or C_p column indicates which criterion was used in the case at hand. There is no underlining for the first cassava equation because that is merely a variant of the second with the linear TEXP term required.

Each expenditure and price variable has been divided by the price of the dependent variable.

The t-ratio for each regression coefficient is given in parentheses. The probability of obtaining by chance a t-ratio that lies outside $\pm$ 1.697 from a regression with 30 degrees of freedom is only 1 in 10. (The smallest number of degrees of freedom in these regressions is 28.)

An entry such as -.176E-4 is to be read as $-.176 \times 10^{-4}$; -.238E-1 = $-.238 \times 10^{-1}$, etc. See TEXP, $(TEXP)^2$ and AGEHD.

[a] No entry is possible in this cell because it is the own-price cell. The own-price variable, p_i/p_i, always has a value of one. The constant term of the regression is the coefficient of the own-price variable; but the significance of the constant term does not indicate the level of significance of that variable.

[b] This variable was deleted from the available set because it was almost an exact linear combination of the remaining variables in the set.

[c] Not included in the available set.

[d] One household has much influence on the estimates of these coefficients.

[e] Variable available but not included in the regression selected.

TABLE 2--Continued

Commodity		Independent Variables: Household Characteristics													Production and Market Factors						
		SIZE	INF	YCH	CH	MAD	FAD	DEPR	WIV	AGEHD	LIMB	TENN	REG1	REG2	SHOOPT	SHOCC	SHOLSO	SHOSS	SHLUR	MKTOR	SHCPH
(1) Rice	(1.1)		69.6 (2.97)	-33.5 (-1.51)	54.3 (1.87)				59.5 (1.87)		-473.4 (-3.68)	-301.9 (-2.89)	b	741.4 (7.03)				c	c		
	(1.2)		70.6 (3.04)	-34.9 (-1.57)	58.0 (2.02)				63.6 (2.03)		-435.5 (-3.53)	-249.6 (-2.61)	b	586.3 (6.00)	5.7 (1.46)			c	c		
(2) Sorghum		48.7 (3.38)	-44.7 (-2.31)	-38.9 (-2.05)	-67.1 (-3.02)	-45.9 (-1.99)			29.4 (-1.41)		312.0 (3.44)	232.5 (3.94)	b	b				c	c	-3.5 (-4.03)	
(3) Cassava	(3.1)	35.6 (2.18)		-62.9 (-1.46)				82.8 (1.23)					b	b			-5.6 (-1.33)	7.4 (2.47)	5.9 (3.45)	-7.2 (-2.15)	-2.0 (-1.91)
	(3.2)	36.1 (2.31)		-63.5 (-1.49)				82.3 (1.23)					b	b			-5.7 (-1.46)	7.5 (2.65)	6.0 (3.50)	-7.3 (-2.28)	-2.0 (-1.92)
	(3.3)	25.3 (2.55)											b	b	7.0 (1.23)			c	c	-3.2 (-1.36)	-2.8 (-2.55)
(4) Palm oil	(4.1)	-6.6 (-1.49)	25.8 (2.78)				35.0 (3.03)		-32.9 (-2.97)	2.03 (3.50)			b	b		-1.6 (-2.01)		1.5 (3.74)			.8 (3.78)
	(4.2)		14.8 (2.09)				34.1 (3.04)		-43.9 (-3.91)	1.47 (2.55)			b					c	c		.8 (3.80)
(5) Groundnut[d]	(5.1)	-49.7 (-2.73)	53.3 (2.68)	42.5 (2.42)	57.6 (2.51)	67.4 (3.37)	66.1 (2.75)		26.3 (1.89)		78.6 (1.49)	155.1 (3.63)		-149.2 (-3.19)			-2.2 (-1.86)	c	c	-.9 (-1.43)	1.2 (3.88)
	(5.2)					17.1 (1.56)			32.6 (2.89)		112.2 (2.34)	161.2 (4.17)		-93.2 (-2.44)				c	c		1.3 (4.37)

TABLE 2--Continued

Commodity and Mean Consumption of Consuming House-holds (Kilograms)	Number of Consuming Households	$\bar{R}^2$	C_p	Constant Term	Independent Variable										
					Expenditure		Prices								
					TEXP	$(TEXP)^2$	PRB	PCA	PPO	PDF	PNF	PGN	POC	Other	
(6) Broadbean 27	41	.759	9.23	390.7 (2.15)	.396E-1 (5.69)	-.332E-5 (-5.06)	b	b	b	-6.6 (-2.47)	b		c	PBN a	
(7) Fish, fresh[f] 65	94	.392	0.07	80.6 (4.23)	.058 (4.71)	-.443E-5 (-3.64)	-332.0 (-6.31)	-506.7 -3.33	96.2 (5.4)	80.0 (5.74)	b		c	PFF a	
(8) Fish, dried[f] 127	124	.442	5.66	110.3 (3.91)	.048 (6.09)		131.6 (2.92)	-784.5 (-4.73)		a	b	168.7 (2.68)	c	PFF -102.5 (-5.69)	
(9) Onions[d] 55	91	.606	-3.42	1.3 (0.10)		.489E-6 (2.53)			b			c	c	PON a	PPC e
(10) Peppers and chillies 23	115	.418	-5.81	45.8 (2.07)		.128E-5 (3.09)		-460.3 (-2.73)	b			c	c	PON 26.6 (2.66)	PPC a
(11) Salt 15	128	.347	4.39	-3.9 (-0.76)	.436E-2 (3.06)	-.329E-6 (-1.70)	-16.1 (-2.34)	-56.5 (-2.57)			13.2 (2.86)	c	c	PFF 4.4 (1.70)	PSL a
(12) Maggi cubes 1	109	.333	5.38	.8 (1.06)	.125E-1 (3.78)	-.898E-5 (-2.25)	35.9 (4.65)	-145.6 (-6.33)	b			c	c	PSL e	PMG a
(13) Kola nut 13	81	.319	-3.10	11.2 (1.01)	.241E-1 (3.28)	-.169E-5 (-2.89)	-73.7 (-2.17)		43.6 (2.08)	-38.6 (-1.84)		c	c	PKL a	PPW b
(14) Palm wine 292	46	.310	-13.64	13.6 (0.12)								c	b	PPW a	
(15) Other cereals 35	113	.342	5.85	17.9 (0.49)	.073 (2.65)	-.631E-5 (-2.38)		933.5 (2.96)			-311.3 (-4.07)	380.4 (2.77)	a		
(16) Other legumes 47	62	.410	1.70	11.6 (0.26)	.050 (3.41)	-.464E-5 (-2.92)			b	b	48.3 (1.62)	-289.3 (-2.02)	c	POL a	
(17) Vegetables[d] 45	132	.656	-6.71	32.1 (1.90)		.343E-5 (6.45)		b			b		c	PVG a	PFT e
(18) Fruits 14	78	.189	0.67	-11.4 (-0.49)	.174E-2 (1.22)			-56.7 (-3.45)		b	b		c	PVG 4.3 (1.92)	PFT a
(19) Salt and other condiments 9	136	.412	1.74	6.5 (2.72)	.166E-1 (4.38)	-.358E-5 (-2.09)		-76.4 (-3.70)				c	c	PFF b	PCN a
(20) Beverages, alcoholic[d] 150	72	.439	-9.98	-157.1 (-1.80)	.132 (1.52)	-.127E-4 (-1.47)						c		PAB a	

[f]The ten households in Enumeration Area 13 were excluded when calculating this regression.

TABLE 2--Continued

Commodity	Independent Variables																			
	Household Characteristics													Production and Market Factors						
	SIZE	INF	YCH	CH	MAD	FAD	DEPR	WIV	AGEHD	LIMB	TEMN	REG1	REG2	SHOOPT	SHOCC	SHOLSO	SHOSS	SHLUR	MKTOR	SHCPH
(6) Broadbean	13.1 (3.16)		-13.6 (-2.43)	-47.0 (-4.93)	-15.4 (-2.18)	-23.3 (-3.68)	-21.4 (-2.21)		1.14 (3.93)	b	83.3 (4.83)						c	c		-4.6 (-2.56)
(7) Fish, fresh[f]					-18.3 (-2.37)							b	b				c	c		c
(8) Fish, dried[f]	-19.9 (-3.63)	33.2 (3.15)		22.9 (1.77)						-117.5 (-3.64)		b	b	-16.8 (-4.06)			c	c		c
(9) Onions[d]														3.7 (2.12)			c	c		2.9 (4.68)
(10) Peppers and chillies														2.6 (4.06)			c	c		.3 (3.09)
(11) Salt		1.3 (1.98)											b				c	c	.2 (2.62)	c
(12) Maggi cubes	-.2 (-2.78)								-.24E-1 (-2.28)	1.1 (1.64)	2.3 (3.92)						c	c		c
(13) Kola nut		-5.4 (-1.95)				-8.0 (-1.73)									1.1 (3.28)		c	c		
(14) Palm wine		-55.3 (-1.19)						92.3 (1.42)		537.2 (3.75)							c	c		c
(15) Other cereals						47.5 (3.36)				366.3 (4.09)	208.9 (4.21)	b	b				c	c	-3.0 (-3.23)	c
(16) Other legumes		17.9 (2.97)						-15.1 (-1.85)	1.24 (2.81)	100.4 (4.53)		b	b				c	c		c
(17) Vegetables[d]					-5.1 (-1.11)				-.35 (-1.2)					6.5 (10.31)			c	c		c
(18) Fruits	3.5 (1.6)				-9.5 (-1.41)			-16.9 (-2.12)	.58 (1.68)					3.3 (3.73)	1.1 (2.11)	.9 (1.62)				
(19) Salt and other condiments				1.7 (2.67)	-1.4 (-2.18)				-.08 (-2.03)		4.5 (2.52)						c	c	.2 (4.71)	c
(20) Beverages, alcoholic[d]		-41.2 (-1.43)						95.3 (1.91)		319.0 (1.94)		b				c	c		c	

Table 2 shows clearly that food consumption patterns are affected by many influences and that the set of variables most important for explaining the consumption of a particular food varies among foods. Useful variables were found in each major category of the independent variables.

The economic variables, expenditure and prices, clearly affect consumption behavior. Expenditure and/or expenditure squared almost always appear and are significant at the 10 percent level or better in one or more regressions for 80 percent of the foods considered. Price relationships are also important, and often quite strong. Household size and composition have material effects, but not always what one would have predicted. The role of the number of wives of the household head, for instance, was quite unanticipated. Region and ethnic group are of consequence for various foods.

One purpose of this study was to determine whether market orientation and/or production characteristics were related to food consumption levels. It is now clear that either may be, but in different ways for different foods.

Interpreting the Coefficients

We turn from this overview of the variables that help explain food consumption choices to a more detailed examination of Table 2. In order to evaluate the magnitudes of the regression coefficients we must know in what units the variables are expressed.[1] Each regression concerns a specific food or group of foods. The dependent variable is the quantity (in kilograms) of the food consumed (available for consumption) annually by the household. Each price and expenditure variable in a given equation has been divided by the price of the food to which the equation refers. Thus each price variable is a relative price: *i.e.*, $\frac{p_{df}}{p_r}$, where p_{df} and p_r are the prices of dried fish and of rice, in Leones per kilogram. This variable measures the price of dried fish in rice, in real terms--the number of kilograms of rice required to buy one kilogram of dried fish at the prices p_{df} and p_r.

The relative price of the dependent variable, p_r/p_r, for instance, is always unity. Its coefficient is the constant term of the regression, but the statistical significance of the constant term does not indicate the significance of the own-price variable. The latter variable also enters the regression as the denominator of all price and expenditure variables.

---

[1]Later we shall present elasticities, which are independent of units, but if we think only in terms of elasticities we lose touch with some of the concrete realities of the situation.

The expenditure variables are also measured in terms of the price of the dependent variable. Thus $\frac{TEXP}{p_r}$, where TEXP is in Leones and p_r, the price of rice, is in Leones per kilogram, is measured in real terms, as the number of kilograms of rice that can be purchased if a household devotes its whole expenditure to buying rice. In terms of its purchasing power, the mean total expenditure figure for our sample is equivalent to 2,481 kilograms of rice.[1] The mean household purchase of rice, 589 kg per year, is 24 percent of this.

Given this definition of TEXP, how does one interpret its coefficient(s)? See Regression (1.1), the first regression for rice in Table 2. Let A and B be identical households, except that the total expenditure of B can purchase one more kilogram of rice than that of A. The effect of this added power to purchase rice must be calculated from the joint action of TEXP and $(TEXP)^2$. From Equation (3) of Chapter IV, $\frac{\partial q_i}{\partial (y/p_i)} = b_1 + 2b_2\left(y/p_i\right)$

$$= .259 - 2(.0000176)(2481) = .172,$$

when evaluated at the mean total expenditure figure for the sample. At this level of real income (in terms of rice), the household with an income capable of buying one more kilogram of rice chooses, on the average, to consume only 0.17 kg more. The coefficient of (TEXP) , which appears to be very small (-.176 E-4), actually has an appreciable effect, reducing marginal consumption by 35 percent (i.e., by 0.09 kg). The coefficient is small, but it is multiplied by total expenditure, which is large.

At a total expenditure level of 1240 kg of rice, B would purchase 0.215 kg of rice more than A. The lower the expenditure level, the greater the amount of added consumption by the household with the slightly larger TEXP.

For all but two of the foods in Table 2, quantities consumed rise with real income (measured in terms of the dependent variable), at least at low income levels. The amount of the increase may fall, rise or remain constant. For palm wine the data reveal no consumption-expenditure relationship. Nor do the data provide conclusive information about the widespread belief that cassava consumption

---

[1]Table 3 gives the sample mean values of TEXP for the consuming households in terms of each of the dependent variables. The values vary both because the prices of the dependent variable differ and because the group of consuming households differs among foods.

TABLE 3

RATIO OF TOTAL HOUSEHOLD EXPENDITURE TO PRICE OF COMMODITY--MEAN VALUE OVER CONSUMING HOUSEHOLDS (Kilograms)

Rice	2481	Salt	1805
Sorghum	2046	Maggi cubes	184
Cassava	13645	Kola nut	1742
Palm Oil	1018	Palm wine	7861
Groundnut	2451	Other cereals	2009
Broadbean	3403	Other legumes	2432
Fish, fresh	2202[a]	Vegetables	1969
Fish, dried	2246[a]	Fruits	8021
Onions	4940	Salt and other condiments	526
Peppers and chillies	1935	Beverages, alcoholic	1813

[a]Excluding the households in Enumeration Area 13.

declines with income, for the expenditure coefficients are statistically insignificant at the 10 percent level. Clearly they do not refute that belief.[1]

Whereas the coefficients of $(TEXP)^2$ are very small and the variable is very large, the coefficients of the relative price variables are large while the variable is small. The mean values of the relative price variables range from 0.018 (the price of cassava in terms of Maggi cubes) to 19.6 (the price of "other cereals" in terms of cassava).[2] In general, relative prices are highest when cassava is the dependent variable and quite high when the dependent variable is palm wine, fruits or onions. Dependent variables that result in the smallest relative prices are Maggi cubes, palm oil, and the group labeled salt and other condiments.

Given this range in relative prices, a "large" regression coefficient can be misleading. Consider the first rice regression, Equation (1.1). The mean price of cassava in rice is 0.205 kilograms. A one-unit change in this price (a five-fold increase, to 1.205 kg) is associated with extra consumption of 436 kg of rice, 74 percent as much as the mean household consumption of 589 kg.[3] Of course, such a change in price is far outside the range of the data. (The actual range of the price of cassava in rice was from 0.09 to 0.34 kg.) Regression coefficients for the cassava price are large in most equations where the variable appears--often for this reason.

The unit of each household characteristic variable from SIZE through WIV is one person. The first six of these, SIZE through FAD, must be considered jointly. Consider the first groundnut regression, Equation (5.1), in which each of these six regression coefficients is significant at less than the 5% level. The coefficient of FAD tells us that if A and B are otherwise identical households, but A has one more female adult than B (even though A and B are equal in size), A will be expected to consume 66 more kilograms of groundnuts per year than B.

Note that SIZE is included in the regression. The coefficient, then, of any other variable measures the effect of that variable when SIZE and other included

---

[1]From Regression (3.1) the marginal increment of consumption associated with higher TEXT is - 0.002 kg at the mean TEXP level for the sample (13,645 kg of cassava). The increment is positive only at TEXP levels below 5698.5 kg.

From Regression (3.2) the marginal increment of consumption is - 0.003 kg at the mean TEXP level for the sample; it remains negative at all income levels.

[2]Table 4 gives the absolute prices (in Leones per kg).

[3]As the t-statistic indicates, this coefficient is not statistically significant.

TABLE 4

COMMODITY PRICES
(Leones per kilogram)

Commodity	Region							
	Northern				Southern			Eastern
	Scarcies	Northern Plains	Bolilands	Northern Plateau	Southern Coast	Riverain Grasslands	Southern Plains	Moa Basin
Rice	.37	.28	.21	.30	.25	.15	.23	.31
Cassava	.13	.02	.06	.04	.05	.03	.07	.04
Palm oil	.76	.76	.74	.78	.49	.49	.60	.61
Fish, dried	.25	.26	.13	.33	.30	.28	.62	.22
Non-food	1.04	.75	.56	.61	.69	.55	.73	.58
Groundnut	.43	.24	.23	.20	.24	.31	.18	.30
Other cereals (all cereals except rice)	.24	.47	.07	.07	1.46	.82	1.88	.20
Fish, fresh	.14	.08	.18	.18	.28	.61	.26	.64
Sorghum	.70	.17	.11	a	1.53	.82	.44	.20
Broadbean	a	.40	a	a	.29	.68	.09	.62
Onions	.10	.19	.19	.10	.11	.25	.13	.12
Peppers and chillies	.58	.21	.49	.35	.46	.19	.44	.34
Salt	.16	.37	.30	.24	.42	.46	.30	.45
Maggi cubes	1.57	3.29	4.74	6.57	4.26	4.25	4.40	2.95
Kola nut	1.42	1.21	1.44	1.68	.48	.23	.77	.12
Palm wine	.15	.07	.08	.08	.07	.05	.06	.16
Other legumes (all legumes except groundnuts)	a	.47	.40	.40	.29	.68	.11	.64
Vegetables	.22	.27	.73	.33	.41	.21	.35	.29
Fruits	.06	.08	.06	.09	.10	.14	.10	.06
Salt and other condiments (salt, sugar, Maggi cubes and condiments, unspecified)	.54	1.70	1.18	1.46	1.22	1.02	1.56	1.80
Beverages, alcoholic	.15	.56	.08	.11	.99	.40	.67	1.22

[a]No price was computed, for no sample household in this ecological zone consumed this commodity.

variables remain constant. Thus a one-unit increase in the level of FAD (females from 16 to 65 years old) can only occur if there is simultaneously a one-unit decrease in the number of persons over 65, a category not represented by a variable and therefore not held constant in fitting this regression. The coefficient of FAD tells us that if A has one more female adult and one less person over 65 than does B, A consumes an extra 66 kg of groundnuts. (The younger of the two women may eat a little more; she almost certainly contributes more labor to the household, and women provide most of the labor for groundnut production [Spencer, Byerlee and Franzel, 1979, p. 24]). If the female adult is also a wife of the household head (WIV), another 26 kg is added to groundnut consumption.

If two households differ only by one unit in the level of the SIZE variable, the SIZE coefficient shows the effect of adding a person over 65. On the average this reduces consumption by 50 kilograms, other variables being held constant.

In the broadbean regression a one-unit higher level for the SIZE variable represents <u>either</u> one extra infant <u>or</u> one more adult over 65; only the remaining age-sex variables were held constant in fitting this regression. Average consumption in this case rises by 13 kg per year.

In the rice regressions only three age-sex variables are included; a positive one-unit difference in the number of infants (INF) is associated with household consumption of an extra 70 kg of rice, with no specification as to what is happening to household size or the levels of MAD, FAD or people over 65. Obviously these variables matter, but the relationships are not consistent enough to be useful in improving the estimates to be made with this regression.

For the full sample the mean level of the variable DEPR (roughly, the ratio of non-workers to workers) is .927. A one-unit change in the level of this variable is equivalent to increasing the numerator by adding to it the number of workers--which would be a large change. The variable is significant only for broadbeans, where it seems to replace INF, the number of infants.

The meaning of a one-unit change in the number of wives depends on which of the size and age-sex variables are in the regression. In the first groundnut regression, if the head of Household A possesses one more wife than the head of Household B, and all other variables are alike, the additional wife in A is probably one of the FAD (female adults) or over 65. If the former, the total effect of her presence is the effect of the 26 kg of consumption associated with an extra wife plus the 66 kg associated with FAD. If she is over 65, only the coefficient of the WIV variable applies.

In the case of the first rice regression, however, the presence of an extra wife can be associated with an extra member in the female adult group and an increase of one in the household size, for these variables are not held constant in the regression.

The variable AGEHD is measured in years. It creates no problems of interpretation.

LIMB and TEMN are binary variables that can take on only two values, 0 or 1. They must be read in conjunction with the two other 0-1 variables, REG 1 and REG 2. If we limit ourselves to the combinations of these variables that exist in our sample, we can say that if LIMB = 1 then REG 2 = 1, for all 16 Limba households are in the Northern Region. If TEMN = 1, REG 2 probably equals 1, but REG 1 = 1 is also possible, for of the 31 households in the Temne group, 25 are in the Northern and 6 in the Southern Region. If neither LIMB nor TEMN = 1, probably REG 1 = 1, but either REG 1 = REG 2 = 0 or REG 2 = 1 is possible, for households in this third group may be either in the Southern or the Eastern Region, or, in a few cases, even in the North.

If LIMB = TEMN = REG 2 = 0, the predicted value applies to a household in the Southern or Eastern regions that is one of the group of 91 households (mostly Mende) not represented by an explicit ethnic variable. If LIMB = 1, REG 2 also equals 1, and the predicted value for rice consumption given by Equation (1.1) will be increased by 268 kg per year (741-473),[1] given a specified set of values for all other variables. In general, Northern households consume more rice than others, but the increase is less for Limba than for Temne households[2].

The Temne people, incidentally, are reported to be rice farmers, while the Limba people farm less and work as palm wine tappers, tapping their own trees and travelling about the country tapping palm trees for others. As equation (14) indicates, other things equal, Limba households clearly consume more palm wine than others.

---

[1]These coefficients would differ somewhat (the net effect would be 68 kg larger) if an alternative ethnic group classification were used that put all Temne households together. This experiment was tried for rice, the only important commodity for which the Temne variable was useful. The results did not differ significantly from those obtained with our other classification. (The changes in most parameters were within one standard error of the new parameter; no parameter changed by as much as two standard errors.)

[2]But consumption per consumer equivalent is highest in the South and, per consumer equivalent, Mende and Temne households consume about the same amounts of rice [Smith et al., 1980, Tables 3.4.A and B, p. 44].

The coefficients of the variables describing production and market factors are small in comparison with those we have just examined. Much of this is explained by the fact that each production and market factor variable is expressed as a percentage. A one-unit change in the value of the variable is a small change. Consider SHOSS in the second cassava equation (3.2). If one household, A, obtains 11 percent of the total value of its output (plus labor sold out) from the list of specified sources (onions, peppers and chillies, and tomatoes; cocoa and/or coffee; oil palm products; non-farm activities; and labor sold out), while B obtains 12 percent from the same sources, we expect B to consume 7 1/2 kg more cassava per year than A if both face the same values for the other variables in the regression. The effect seems small, but the coefficient is highly significant and the range of the variable within the sample is from zero to 100 percent. At the mean value for the sample, 29.6 percent, this variable contributes 220 kg to the total predicted value for cassava consumption. (Mean household consumption for the 114 consuming households in the sample is 394 kg.) Households that apply larger shares of their productive activities in these ways (usually in order to earn cash income) tend to consume more cassava than others. But note that this coefficient is an estimate for given values of four other production and market factor variables. The increase in these production activities must occur without changing the level of labor sold out, the proportion of labor devoted to upland rice, the percentage of output (by value) that is sold, or the percentage of cassava consumption that is grown at home. If we were to sort the households into those with large and small percentages of output value from these sources, we should undoubtedly find that the values of a number of other variables were quite different for the two groups of households.

If MKTOR is high (a household sells a large percentage of its value output), given no change in one of the five production pattern variables, the average household consumes <u>less</u> cassava than others, at the rate of 7.3 kg per percentage point increase in MKTOR. Should MKTOR and SHOSS increase simultaneously their effects would essentially offset each other.

The Commodity Equations

We turn our attention now from interpreting individual regression coefficients to examining the commodity equations as wholes. We shall study the rice equations in detail before presenting a summary view of the remaining regressions.

Rice

Table 2 presents two equations, both chosen in accordance with the C_p criterion. The two-digit multiple correlation coefficient (.78) is the same for both, but only the first equation (1.1) provides information about relationships between rice and two major foods, cassava and palm oil. In other respects the equations differ little except that (1.2) furnishes information about the influence of groundnut prices and household production of onions, peppers and tomatoes. To be sure, the coefficient of PCA (cassava) in (1.1) is not statistically significant even at the 10 percent level.[1] Either regression can be used, depending upon which type of information is needed.

Each rice equation contains two variables with coefficients that are not significant at the 10 percent level or less. In (1.1) these are the price of cassava and the number of young children in the household, and in (1.2) the number of young children and the share of the value of output (including labor sold out) that is derived from the production of onions, peppers and tomatoes. In these four cases we cannot regard the data as supporting the hypothesis that the coefficient is indeed different from zero, because its value could have resulted from chance with a probability of more than 10 percent. If, however, we believe that such a relationship exists, the regression coefficient measures the relationship that is found in the data. Though statistically not significant, these coefficients improve the predictive power of the regression as a whole and provide information about the quantitative nature of the relationship. It may be desirable to use them if we have a strong enough *a priori* belief that the relationship exists and is likely to be important.

Expenditure and Price Relationships

As we have seen, households with larger total expenditures consume more rice than others, but the higher the expenditure figure the smaller the additional effect. Households facing high relative prices of palm oil or groundnuts consume less rice on the average than others and those facing high relative prices of dried fish or perhaps cassava consume more rice, if the households are otherwise similar in all respects. In these regressions, the effect of a

[1] If all Temne households are placed in a single class, the coefficient becomes -44, but this is not significantly different from zero at any acceptable level. (The new t-ratio is only 0.11.)

change in income or prices is the same regardless of the size or composition of the household or of the levels of other variables. The effect is estimated as though the other variables were at their average levels.

If we regard these equations as demand regressions, in the sense of being affected by the amount of income but not by its source or form, dried fish and perhaps cassava appear to be gross substitutes for rice, while palm oil and groundnuts appear to be gross complements. That is, where relative prices for dried fish are high, the household consumes less dried fish and therefore more rice than otherwise; where palm oil prices are high it consumes less palm oil and therefore less rice.[1]

However, these are not demand regressions in the narrow sense. They are behavioral regressions which take account of forces operating on the production as well as the consumption side of the household activities. Correlations with uncontrolled supply side variables may account for part of the influence attributed to the prices of palm oil and groundnuts in these equations. With single-equation estimates it is not possible to know whether the relationship that appears is the result of mechanisms operating on the expenditure side or the production side or some combination of the two, since the influences of both sides are present. Taking account of production as well as consumption effects can change not only the magnitude but the sign of response to a changed price situation. Moreover, a coefficient resulting from a substitution relationship on the supply side may have the same sign as one resulting from a complementary relationship on the demand side.

The Palm Oil Connection

Both rice equations show that households facing higher prices of dried fish consume more rice than others and the first (1.1) shows a similar relationship with respect to the price of cassava. These are the standard relationships one would expect from theory that looks only at the consumption side of household decision-making. The household facing a higher price for dried fish is expected to consume less of it than other households and more of other foods in general, including rice. There would be less fish in the sauce and perhaps less sauce with the rice.

---

[1] There are relationships within customary eating patterns that could lead to the latter result, but it is unlikely that they are quantitatively important. See Codicil A to this chapter.

The mechanism is different with respect to cassava, but the relationship is similar. Rice is the basic staple, but cassava is consumed as a less desired alternative. Where the price of cassava is relatively high one would expect households to reduce their use of cassava in favor of more rice.

Each rice equation identifies one food for which the previous relationship appears to be reversed. According to (1.1), households facing high prices of palm oil consume less rice than other households, and according to (1.2) households facing high prices for groundnuts also consume less rice than others. Let us look first at the relationship between the price of palm oil and the consumption of rice.

One possibility is that the result is simply wrong, because the model is misspecified, the data are unreliable, or the sample is unrepresentative in this respect. In this instance, that does not appear to be the case. In behavioral regressions like these there are mechanisms other than demand-side complementarity that can result in a negative association between the price of palm oil and the quantity of rice consumed by the household. The single-equation estimate can identify and measure the net effect of these relationships, but it is not well designed for sorting out the influences of several mechanisms operating simultaneously; it shows the net result of the whole set of mechanisms.

The principal relationships relating the price of palm oil to the consumption of rice appear to be the following. Other things equal, a high price for palm oil is associated with high palm oil production; high production is associated with high consumption; and palm oil consumption is a substitute for the consumption of rice.

Processing the fruit of the oil palm produces both palm oil and palm kernel, so the level of palm oil production is affected both by the price of palm oil and the price of palm kernel (or of kernel oil).[1] At this stage of the analysis we have not taken the palm kernel price into account, but, other things equal, a high price for palm oil should induce greater output of both palm oil and palm kernel. A high market price for palm oil is an incentive for expanding production for home use; a high sales price is an incentive for expanding output for sale.

---

[1] Palm kernels may also be obtained in a separate operation by collecting dried fruit that has fallen from wild oil palm trees.

Palm oil consumption is largest among households that produce palm oil. The principal producers of oil palm products are households in the South and East [Spencer and Byerlee, 1977, p. 53], and palm oil consumption per consumer equivalent is largest in households in those regions [Smith, et al., 1980, Table 3.7.A, p. 58].

Furthermore, producing a large part of the palm oil that is consumed is conducive to high consumption per consumer equivalent. Annual palm oil consumption is 13 kg per adult male consumer equivalent among the 47 percent of the consuming households in our sample that derive less than 19 1/2 percent of their consumption from their own production. Households that produce between 19 1/2 and 94 percent of their consumption (18 1/2 percent of the consuming households) consume 25 kg per consumer equivalent. For the remaining 22 percent, households that produce 94 percent or more of what they consume, the average consumption figure is 30 kg [Smith et al., 1980, Table 3.7.E, p. 60.] Both palm oil regressions (4.1 and 4.2) in Table 2 of this report reveal the same relationship. Household palm oil consumption rises with the share of that consumption that is produced at home.

Equation (4.1) also shows that household palm oil consumption rises with the share of the value of output plus labor sold out that is derived from the production of oil palm products, non-farm activities other than fishing, labor sold out, and the production of either cocoa and/or coffee or onions, peppers and tomatoes. Of course we cannot be sure that this positive coefficient results primarily from the presence of oil palm products in this list of activities. The same regression, however, shows that the share of this value sum which comes from the production of cocoa or coffee is negatively related to palm oil consumption, so the positive effect of the other components of the list must be even stronger than the regression indicates.

Finally, Equation 4.1 (the palm oil regression) yields positive own-price elasticities of demand for palm oil, presumably because of the effect through production.[1]

When we consider the production side of the problem it should not be too surprising if a high price for palm oil leads to greater consumption by producing households. Selling some of the palm oil one produces is in effect

---

[1]These elasticities will be presented in Table 11 (Chapter VII).

exchanging palm oil for the other things (call them Good Z) one can buy with the money received from palm oil sales. If the relative price of palm oil is high the price of Good Z is low in terms of palm oil. This low price of Z causes more Z to be purchased, but does not necessarily cause more palm oil to be given up in exchange for it. If the elasticity of demand for Z in terms of palm oil is less than unity, a household can obtain the larger quantities of Z it chooses to consume for less palm oil than it would have taken to buy the smaller quantities purchased at the higher price of Z in palm oil. In short, where the sales price of palm oil is high households may choose to take advantage of this price by retaining more palm oil for their own consumption than would be done otherwise, even in the absence of any change in the quantity of palm oil produced. If they also react to a high price of palm oil by expanding palm oil output, the household may be able to increase both the quantity of palm oil consumed and the quantity exchanged for money income to be used to purchase other things.

Another aspect of the situation is that producers of palm oil have access to at least part of their consumption at a price below what must be paid for oil purchased from the market. Producers of large amounts of palm oil can obtain a larger fraction of their total consumption at the relatively low farm sales price than can small producers. (Forty-seven percent of all consuming households buy some and produce some of what they consume; another 15 percent produce all they consume.) In addition, though the price data do not reveal this, the oil produced at home may be fresher and of better quality than that obtained from the market. Thus the effective average consumption price of palm oil can be lower for large producers than for others.[1] While the opportunity cost of consuming home-produced palm oil rises if the selling price for the oil rises, that opportunity cost may still be lower than the market price of palm oil that otherwise would have to be purchased from the market.

Note, however, that the sequence operates through the effect of the high price of palm oil upon production, so we would not expect it to be significant except among the 62 percent of the households that produce some or all of the

---

[1]Such household-to-household variations in the effective consumption price are not fully shown by our data, for our data are averages based on the proportions obtained from home production and from the market for all households in a given ecological zone.

palm oil they consume. This is consistent with findings obtained when we fitted the rice regression separately to houeholds in the Southern and Eastern Regions (where much palm oil is produced) and in the Northern region (where relatively little palm oil is produced). In the Southern and Eastern Regions a high price of palm oil reduces rice consumption, presumably through such mechanisms as we are discussing, but in the Northern region it increases rice consumption. In the North, of course, most of the palm oil consumed has to be obtained through the market, while in the Southern and Eastern Regions much larger fractions of palm oil consumption are produced within the household.

In addition to the fact that producers of palm oil have access to part of their consumption at prices well below the market price, there are relevant supply-side mechanisms associated with changes in production levels for palm oil that are excluded from standard consumption theory because that theory excludes supply-side phenomena. On the supply side palm oil production competes for labor and resources with a variety of other activities. Any expansion of a productive activity in response to a relatively high price for the product is likely to be at the expense of some other activity, unless additional labor and other resources are acquired by the household. When households devote as much of their total energy to the production of food crops as is done in rural Sierra Leone, expanding any activity is likely to be at the expense of a crop that can be used for food. It is possible for reductions in food crop production to occur without impinging upon the quantity retained for food, if the full effect is taken in the form of a reduction in sales, but this is not likely. In any case, once sales fall to zero, further output reduction must force a lower level of household consumption. The smaller the proportion produced for sale, the smaller the margin for reducing output without reducing consumption and the greater the likelihood that home consumption will be reduced. As the households in our sample produce only 74 percent of their own rice consumption, on the average, there must be many that do not produce enough for sale to permit them to maintain their levels of rice consumption in the face of appreciable expansion in other household activities.

To be sure, peak periods of labor use in the production of oil palm products, in the South at least, do not coincide with the peak periods of labor use in rice production. (There is some overlapping of periods of high labor use with inland swamp and hand boliland rice production in the North.)

[Spencer, Byerlee and Franzel, 1979, pages 15-24, 32]. However, this pattern of labor use has been adopted, at least in part, because there is competition for labor between processing oil palm products and other activities. In the 1974-5 survey, palm oil prices were highest in December and January when little palm fruit was being harvested and female labor was heavily engaged in rice harvesting [Spencer and Byerlee, 1977, pp. 56-8]. Prices were more than twice as high in January as they had been the previous May. Furthermore, the principal processing season for palm fruit grown on plantations extends over a longer period than the season for households that also engage in the production of rice and other types of crops.

These substitution relationships between palm oil production and the production of other food crops (the most important of which is rice) mean that anything that encourages the production of palm oil raises the internal marginal opportunity cost to the household of consuming other foods produced at home and thus has a negative effect on the consumption of such foods. This is in addition to the demand-side relationships considered in conventional demand theory, for the latter considers the relevant opportunity costs to be the market prices of alternative foods and regards these as independent of the price of palm oil.

In Sierra Leone (as in other parts of West Africa) there is an additional supply-side relationship that grows out of competition with the use of women's labor in domestic activities. Pounding and cooking rice compete with the production of palm oil for the time of the women of the household. Seventy-four percent of the rice consumed by the households in our sample was produced by the household [Smith, et al., 1980, Table 3.1, p. 27]. Before rice can be cooked it must be pounded--a time-consuming occupation carried out by women and children. Yet women and children also do much of the work in processing oil palm fruit to produce palm oil. If a household in an area where the price of palm oil is relatively high responds to that price by producing more palm oil than other comparable households, the extra labor spent on palm oil processing must be withdrawn from some other activity (including leisure), unless extra labor is hired. To spend less time pounding rice for consumption would be a normal response, particularly since rice pounding takes such a large fraction of womens' time[1] and there are foods available (cassava, for instance)

[1]The importance of this mechanism will diminish sharply as the use of small mills to clean rice increases, but such mills are still relatively unimportant in rural Sierra Leone.

that require less work to prepare.[1] The fact that palm oil processing competes with rice pounding for women's labor means that anything that makes palm oil production more profitable raises the internal marginal opportunity cost of consuming rice, even though there is no change in either the sale price or the market purchase price for rice. This too is a mechanism that finds no place in conventional demand theory, for the standard form of that theory takes no account of the facts that (1) consumption activities require time and (2) the time available for such activities may be affected by the kind of production activities selected.

If palm oil production is positively associated with the high price of palm oil and palm oil consumption with palm oil production, then the response of rice consumption to the price of palm oil will be positive or negative according as rice and palm oil are complements or substitutes in demand. On the demand side one can argue either that rice and palm oil are complements (because they are used together in the diet) or substitutes (because the proportions of palm oil to rice may vary). Which relationship dominates is an empirical question. Before considering the possibilities more closely we describe briefly the predominant eating patterns.

Most households in Sierra Leone eat one or two meals a day. Hamilton [1977] reported that 58 percent of the villagers had eaten one meal on the day before his interview and 40 percent had eaten two. Seventy-five percent of the villagers in the Northern Province reported eating two meals, but only 12 percent of the Southern Province villagers had two meals. See also Kolasa [1979, p. 36, and Table 3.6, p. 18].

The basic staple is rice. Kolasa reports, "Cassava is viewed as a food to be eaten when rice is not available." [1979, p. 45]. Rice is normally eaten with "sauce", also known as plasas or palaver sauce. Occasionally it is eaten with stew. Cassava, on the other hand, is normally eaten with stew or soup but is sometimes accompanied by sauce. The plasas (sauce) contains oil, usually palm oil, various types of vegetables leaves and "condiments"--vegetables, meat or fish, and seasonings. Dried fish is normally used. Stew, not eaten often, consists of fish (normally fresh), chicken, or meat, cooked in oil with onions, peppers or tomatoes.

---

[1]Most cassava grown in Sierra Leone is the type sometimes called "sweet" cassava. It does not require fermentation and pounding to make it edible, but can be prepared by boiling.

Soup usually has little or no oil, but contains meat, fish or chicken and a good deal of water. Palm oil soup, however, is made with fish and palm oil.

Sauce is served most frequently, with soup and stew less often. Soup is normally served with cassava but not with rice. Stew is more expensive than sauce so the frequency with which it is served depends upon the family income and its access to fish and or meat.

If palm oil is scarce or expensive it can be replaced by some other oil (perhaps groundnut or palm kernel oil) without major disruption of the customary meal of rice with sauce. Alternatively, the household might eat fewer sauces and stews, which use palm oil, and more soup, which uses little or no oil. Soup is normally eaten with cassava rather than with rice, so the latter adjustment implies that palm oil and rice are complements on the demand side while palm oil and cassava are substitutes. If this were the dominant relationship we might argue that demand side complementarity is sufficient explanation of the fact that high palm oil prices are associated with low rice consumption. That is, a high price for palm oil reduces the consumption of palm oil and therefore of the rice eaten with it. It is unlikely, however, that this is a dominant relationship, for soup is not a major item of consumption, probably because the meat, fish or chicken it contains is relatively expensive.

It is more likely that the dominant demand-side relationship between palm oil and rice is substitution. According to informants in Sierra Leone, when palm oil is scarce or poor in quality people eat sparingly of the sauce and increase the proportion of rice they consume. When palm oil is abundant or of good quality (fresh and "heavy"), they take more sauce and less rice. There is also a physiological basis for this substitution, for both palm oil and rice are important sources of food energy. The more palm oil one eats, the less rice he needs. Likewise, because fats and oils have higher satiety values than carbohydrates, a person feels satisfied for a longer time at a given level of caloric intake when there is a high proportion of palm oil to rice in the diet.

In summary, Regression (1.1) shows that households facing high palm oil prices consume less rice than other similar households. The explanation may be that (1) high relative prices for palm oil are associated with high production levels, (2) that high production is associated with high consumption of palm oil, and (3) that palm oil and rice are, on balance, demand-side substitutes in consumption. That high producers tend also to be high consumers of palm oil probably reflects the fact that the proportion of total palm oil consumption available at the relatively low farm sales prices is greater for large producers than for others as well as the fact that high levels of palm oil production contribute to high internal rates of marginal opportunity cost for the production of other foods and, in particular, high opportunity cost rates for the consumption of rice because of the time required for rice pounding.

These mechanisms only operate for households that expand palm oil production in response to a high price. Households that do not produce palm oil or do not respond to the production opportunity presumably reduce palm oil consumption and increase rice consumption if confronted by a relatively high price for palm oil. Undoubtedly both types of households exist. The data suggest that the dominant reactions are those of the first group (price-responsive producers of palm oil).

The second rice equation (1.2) indicates that rice consumption is negatively related to the price of groundnuts. We shall not discuss this result beyond pointing out that two of the mechanisms operating with respect to palm oil are also present with respect to groundnuts. (1) Groundnuts are grown primarily on small acreages tended by women. Any expansion in groundnut production in response to a high relative price for groundnuts will compete with rice pounding for women's time. (2) In addition, we know from the groundnut equations (5.1 and 5.2) that households producing large portions of their consumption of groundnuts consume more than do other households. Households that expand groundnut production because they face a high relative price of groundnuts may also expand home consumption and, as a consequence, consume less rice than other households. (For the sample as a whole, 81 percent of the groundnuts consumed are home-produced [Smith et al., 1980, Table 3.1, p. 28].)

Household Characteristics

The rice equations state that if households A and B have same number of children (CH) and young children (YCH) and face the same values for all other variables in the regressions, but B has one more infant (0-5 years of age) than A, then B will be expected to consume an added 70 kg of rice per year. If, instead, the difference were that B possessed one more young child (ages 6-10), the predicted consumption would be some 34 kg less for B than for A. This seems odd, for an infant certainly eats less rice than a child aged 6-10 years. But the presence of an infant in the household is likely to be associated with the presence of pregnant or lactating women or their female relatives. While the number of wives of the household head is held constant in these two equations, the number of pregnant or lactating women is not, and those female members of the household and any relatives who may have joined them to help with child care will be consuming rice that would otherwise not be eaten.

Both rice regressions state that of two households with the same number and ages of children, and otherwise alike, the household whose head has more wives consumes more rice. In these equations the total size of the household is not held constant, so part of the effect may be the result of having one additional person in the household. Still, if that were all there were to it the variable FAD (the number of female adults between the ages of 16 and 65) might have served as well, for an additional female adult would presumably eat about the same quantity of rice as an additional wife, unless the wife were pregnant or lactating. The equation does not tell us exactly what the mechanism is, but there are several possibilities:

1) In the first place, the number of children is specified in these regressions. An increase in the number of wives is an increase, on average, in the amount of labor available for feeding each child, an increase that the data tell us is more dependably effective than simply an increase in the number of female adults--even though, aside from the special role of the wife, these two groups might be regarded as essentially the same.

2) In Sierra Leone, as in much of West Africa, the wife has a special responsibility for seeing that her own children are properly fed. Often each wife prepares the food for herself and her children. Alternatively, the head wife may delegate the cooking for the whole household to a single wife on a given day, but in this case there is a tendency for other wives also to cook a little something for their own children or their husband. Moreover, each wife normally has a small plot of ground that she works for herself. Whatever she produces belongs to herself for her own use.

In the case of rice, in particular, pounding the grain is extremely time-consuming. The special commitment to caring for one's own children can well increase the amount of rice pounded and therefore consumed.

3) Nigerian informants suggest that when a new wife comes to a household she doesn't come alone. Our rice regressions do not hold the size of the house-

hold constant so the coefficient of the variable for the number of wives may also reflect the increase in household size that occurs when a wife brings a sister or other female relative with her.

This study does not allow us to isolate all the mechanisms involved, but it does identify relationships that deserve further study--perhaps to be carried out with the aid of an anthropologist. Given the frequency with which the number of wives is a useful explanatory variable in this set of equations, and given the role that rice pounding seems to play in the relationships between rice consumption and the prices of palm oil or groundnuts, it is evident that seeking ways to reduce the time and energy used by women for producing and preparing food might be one component of an effective approach to improving nutrition in rural Sierra Leone.

The regional and ethnic variables show us that, on the average, Northern households (almost all of them Limba and Temne households) consume more rice than an average household in the South or East, other things equal. However, Limba households do not consume as much as the Temne. Mende households, all located in the South or East, consume less rice, other things being equal. But other things are not equal because average expenditure levels and average household size vary among ethnic groups and among regions, as do relative prices and other relevant variables. Rice consumption per consumer equivalent, for instance, is highest in the South and about equal among the Mende and the Temne, when other variables are free to move. [Smith, et al., 1980, Tables 3.4.A and 3.4.B, p 44.]

In the second rice equation the positive effect on rice consumption of being located in Region Two is noticeably smaller than in (1.1). This may result in part from the presence of the variable SHOOPT, which measures the share of the value of total output plus labor sold out that is represented by the output of onions, peppers and tomatoes. Region Two includes a group of households that produce these vegetables commercially for sale in the Freetown market. As Equation (1.2) shows, each percentage point of this variable is associated with an extra 5.7 kg of rice consumption per year. For the small number of households engaged in this activity on a commercial basis, the effect on rice consumption can be appreciable. The maximum value of the variable in our sample was 51 percent, equivalent to 291 extra kg of rice consumption. For most households, of course, the effect is negligible. (The average share from onions, peppers and tomatoes was only 2.5 percent.)

Résumé

In sum, household rice consumption rises with expenditure (but less rapidly as expenditure levels rise), responds to relative prices (as the result of both supply and demand-side relationships), and is affected by the size and composition of the household as well as by ethnic group and by region. Policy actions that have an impact upon any of the relevant variables are likely to affect household consumption.

An Overall View

Household food consumption levels for almost every commodity are clearly associated with levels of total expenditure; they rise as expenditure rises. $(TEXP)^2$ appears in most regressions; when both TEXP and $(TEXP)^2$ appear, the consumption-expenditure relation is convex from above.

For some commodities, among them sorghum, groundnuts, broadbeans, peppers and chillies, salt, vegetables and fruits, the expenditure response is small, even though it is often quite significant in a statistical sense. (But note that each of the last four "commodities" has non-homogeneous components.)

For four foods (cassava, palm wine, alcoholic beverages and fruits) the data do not confirm the existence of an income relationship, even at the 10 percent level of significance. (In part this may reflect the fact that the data for cassava are not as reliable as those for most of our commodities, and that "fruit" is a conglomeration of quite different components.)

Commodity substitutions in response to differences in relative prices are quantitatively important for almost all foods. The exceptions are onions, palm wine and the two groups of foods, vegetables and alcoholic beverages.

The relative price of cassava is the price variable most often helpful in explaining the consumption of some other commodity. (It appears in at least one regression for each of 12 foods.) The relative prices of dried fish, groundnuts, rice and non-food goods are also useful in explaining the consumption of other foods. These price variables have negative coefficients more frequently than one would expect if these were pure demand regressions that describe the behavior of households buying all their food in the market. Most of these households produce large fractions of their own food, so prices affect household consumption through their effects on household production as well as through the effects that we conventionally think of within the consumption sphere. The data show negative coefficients for the price of rice in the regressions for cassava, fresh fish, salt and kola nut, as well as for the prices of palm oil and of groundnuts in the two rice regressions.

The cassava-rice coefficient is negative in the cassava equation and positive in the rice equation, but this difference could result from the fact that the coefficients include the income effects of changes in price. Rice represents 25 percent of total household expenditures in the sample, and cassava only 7.5 percent, so the reduction in well-being associated with a high relative price of rice is likely to force economies in the consumption of a number of foods, including cassava. This income effect may be an important factor in explaining the reduced consumption of each of the four foods, cassava, fresh fish, salt and kola nut, but interrelations on the production side may also be involved. The price of cassava, which takes a much smaller fraction of total expenditure, is likely to have a much smaller income effect on the consumption of other foods than is the price of rice.

Household size and composition are also important. Each size and age-sex variable appears in at least one regression for five or more of the foods; the number of infants is a useful variable for ten foods. In general these variables are more important for understanding the major foods than for some of the minor ones.

As we had expected, no single set of age-sex variables is optimal for use in a large number of equations. The dependency ratio (DEPR), a specific weighted combination of these variables, is serviceable for only two foods and statistically significant at the 10 percent level or better only for broadbeans. To be sure, had the other variables not been available as alternatives, the dependency ratio might have played a greater role.

The age of the household head and the number of wives he possesses prove to be effective variables at least once for each of seven or eight foods. The number of wives, incidentally, is often serviceable when the number of female adults is not, and vice versa; for palm oil and groundnut both variables are informative and statistically significant at the one percent level. We have already seen how such household characteristic variables as INF and WIF (like the price variables) measure the combined production and consumption effects of the particular age-sex group.

Knowing the ethnic affiliation of a household also helps explain its consumption pattern. Households in the Limba or Temne group behave differently from the remaining households (mostly Mende) in the case of seven individual foods. Limba households, for instance, consume less dried fish than do Mende households but more sorghum, cereals other than rice, Maggi cubes and palm wine. Households in the Temne group consume less rice than households in the Mende group, but more sorghum, cereals other than rice, groundnuts, broadbeans, Maggi cubes, and "salt and other condiments."

The regional variables were often deleted because of high collinearity with other variables, but Region 2 (the Northern Region) was a statistically significant classification (at the one percent level) for rice and groundnuts.

One concern in this study was to determine whether production characteristics and/or market orientation affect food consumption decisions. Clearly either or both may do so. Some production or market variable aids the explanation for 11 of the 14 single foods.

Market orientation, the percentage of the value of total output that is sold, improves the explanation for six foods or food groups, while the share of household consumption that is produced at home is a helpful variable in explaining the consumption of six foods, two-thirds of the total number for which it was available.[1]

Producing a large fraction of household consumption has a positive effect on the consumption of palm oil, groundnuts, onions, and peppers and chillies, and an adverse effect on the consumption of cassava and broadbeans. A high degree of market orientation has an adverse effect upon the consumption of cassava, sorghum, "other cereals" (all cereals except rice) and groundnuts[2], much as one might expect. Salt consumption is positively associated with market orientation, also as one would expect.

Of the variables representing the percentage of total product devoted to specific crops, SHOOPT, the share of onions, peppers and tomatoes, was the most useful, appearing in five food regressions and two for groups of foods. As one would expect, SHOOPT is positively associated with the consumption of onions, peppers and chillies, and vegetables, but also with the consumption of rice, cassava and fruits. It is negatively related to the consumption of dried fish.

---

[1] We did not use the share of consumption produced at home for palm wine, as 94 percent of consumption was home-produced, or for salt and Maggi cubes, where none was produced at home. Nor did we use this variable for fish or for groups of foods.

[2] The coefficient for groundnuts is not statistically significant at the ten percent level.

(Many of the households that produce large amounts of onions, peppers and tomatoes also produce large quantities of fresh fish.)

Two variables (SHOSS and SHLUR) were tested only for cassava and palm oil. SHOSS, the percentage of the value of output plus labor sold out that came from the list of specified sources, is statistically significant at the one percent level and positively associated with consumption in at least one regression for each of the two foods, while SHLUR, the share of labor devoted to upland rice, is significant at the same level for cassava, and also, as one might expect, is positive in its effect. SHOLSO, the contribution of labor sold out to the total value of output plus labor sold, appears only in regressions for groundnuts, cassava and fruits.

The six regressions for groups of foods are usually dominated by one or two of the individual foods that comprise them. In those cases, the regressions for the single foods are to be preferred because they describe the behavior of significant foods that are reasonably well defined rather than the average responses of some conglomerate of individual parts. "Other cereals" is an exception; it includes two rather important cereals in addition to sorghum (fundi and millet), plus benniseed and maize; the behavior of the group is quite different from that of sorghum alone.

Summary

Household expenditure levels, prices, and household size and composition affect the food consumption choices of rural households in Sierra Leone, as do location (Region) and ethnic group. No single variable is optimal for measuring household size and composition, for these factors affect different foods in different ways.

Consumption decisions for various foods are clearly affected by the household's production opportunities or decisions and by its orientation toward producing for the market or for home consumption. In general, producing a large fraction of household consumption is conducive to greater consumption of a particular food (but not always), but producing certain foods for the market is also conducive to greater consumption of those foods. Rice consumption, however, does not show a statistically significant relationship to the production or market orientation variables.

In the case of rice, as for these regressions in general, relationships operating on the production side of the household are important as well as those on the consumption side. Whether or not there are production or market orientation variables in the regressions, the behavior revealed by the data is behavior that responds to events on the production as well as the consumption side; the coefficients show the net effects of the entire set of relationships. Single-equation regressions are not well adapted to separating all the mechanisms at work, but they do identify the existence of important mechanisms and estimate the net quantitative importance of important variables affecting household food consumption levels.

CHAPTER V

SHARE REGRESSIONS: ENTIRE SAMPLE

The share regressions predict the share of total expenditure on a particular food. We have calculated them for six of the major foods and six groups of foods, primarily as exploratory work in preparation for systems estimation of the household-firm model. For that purpose we wished to know which demographic variables would be most useful, but in order to find out we had to fit regressions without the variables relating to production activities or market orientation. Those were deleted because they would not be used as independent variables in the systems estimation. (Production decisions are endogenous to the systems model--determined within the system.) Besides deleting the production and market orientation variables, we fitted these 12 regressions in the share form because doing so often reduces heteroskedasticity, yet in this case heteroskedasticity remained a problem. Still, the regressions yield unbiased point estimates of the parameters, although the estimates of their sampling variation could be improved if we were to use weighted regressions.

The share equation results sometimes differ appreciably from those of the quantity equations. In part this is because the production and market factor variables (and the regional variables) are not being used, but more is involved. In fitting the quantity regressions we minimized the sum of the squared deviations between predicted values and actual quantities consumed, while for the share equations we minimized the sum of the squared deviations between the predicted shares of total expenditure and the actual values.

Either procedure gives unbiased estimates if the disturbances have means of zero and are independent of the exogenous variables, but because errors are present the two procedures generally do not give the same estimates. Which one gives the better estimates depends upon whether heteroskedasticity is more of a problem with one form of the equation than with the other. If least squares estimation is to yield best unbiased estimates the variances of the disturbances must be the same for all observations (homoskedastic). If the errors in one form of the equation are homoskedastic, they will be heteroskedastic in the other form. We return to this point later.

The variables used in the i^{th} food regression were selected by the same computer procedure followed for the quantity regressions. The variables available were: TEXP and $(TEXP)^2$; the prices of rice, cassava, palm oil, groundnut, dried fish and non-food, plus some prices from the remainder of the 12 foods or food categories under study; and the household characteristics variables (except that regional variables were omitted and male and female adults were combined into the variable AD). Production and market factor variables were excluded from the available set.

In running the quantity regressions a few variables had to be deleted because of multicollinearity within the available set, but with the regional variables missing no other variables had to be deleted from the available sets for the share equations.

From all possible combinations of variables in the available set we chose the equation for which $\bar{R}^2$ was the greatest.[1] For the share equations, maximizing $\bar{R}^2$ generally provided more income and price variables than we would have had if we had minimized C_p. Using the $\bar{R}^2$ instead of the C_p criterion tends to increase the number of variables in the regression. The second Alcoholic Beverages regression was included because maximizing $\bar{R}^2$ gave no price response information.

Table 5 (see page 70) contains the share equation regressions. In reading it we must remember that the column headings now have different interpretations than in Table 2. As we know, the share equation (5) is simply the quantity equation (2), multiplied through by p_i/y, where p_i is the price of the dependent variable in the i^{th} quantity equation and y is the total expenditure of the household. The dependent variable in the share equation for the i^{th} food is the ratio of household expenditure on that food to total household expenditure $(\frac{q_i p_i}{TEXP})$. The linear expenditure term in (2), $b_1(y/p_i)$, has now become the constant term and the former quadratic term has become $b_2(y/p_i)$. The TEXP column in Table 5 refers to y/p_i; TEXP is now measured

[1] If TEXP was not a variable in the equation, it was added, as long as its addition had little effect on the coefficients of the remaining variables.

in kilograms of the i^{th} food.[1] The price columns now refer to the p_j/y, rather than p_j/p_i, $i \neq j$. The term in p_i, p_i/y, appears in the share equation just as does any other price variable; it no longer has been transformed into a constant term. The denominator of these price variables is constant across equations, while in the quantity equations it was not. (The denominator of the TEXP variable, however, still varies among equations.) The household characteristics variables are now $h_k(p_i/y)$ rather than simply h_k. Household size, for instance, is the number of persons per unit of total expenditure, where total expenditure is measured in kilograms of the i^{th} food.[2]

The share equations usually contain more price variables and fewer household variables than their quantity equation counterparts. The absence of production and market orientation variables probably contributes to this, in addition to the fact that using maximum $\bar{R}^2$ as a choice criterion generally raises the number of variables in the equation.

TEXP is less often statistically significant in the share equations than its counterpart, $(TEXP)^2$, in the quantity equations, but this simply means that demonstrable departures from a constant share of expenditure are less often noted than departures from a constant linear relation between quantity and TEXP.

The share equation for rice shows that high prices of palm oil, groundnuts and "other cereals" are associated with reduced expenditure shares for rice. This supports results obtained from the two quantity equations for rice.[3] With the expenditure term (y/p_i) constant, a reduced share means a reduced quantity of rice consumed.[4]

---

[1]Table 3 gave the mean values of y/p_i for the consuming households in the sample.

[2]That is, $h_k \div (y/p_i)$.

[3]POC was not included in either quantity regression.

[4]When the price of the j^{th} food changes, both the share and the quantity of the i^{th} food change in the same direction, but if the price of the i^{th} food changes, the share and quantity of the i^{th} food may move in opposite directions. The share equation coefficient of p_i/y can be positive even though q_i falls when p_i/y rises, if the own-price demand for q_i is inelastic.

TABLE 5

THE SHARE REGRESSIONS

Commodity and Mean Share[a] of TEXP	$\overline{R}^2$	C_p	Constant Term	Independent Variable: Expenditure TEXP	Prices: PRB	PCA	PPO	PDF	PNF	PGN	POC	Other
(1) Rice .250	.317	9.79	.311 (5.96)	-1.71E-4 (-1.74)	-456.5 (-3.66)	743.1 (2.94)	-160.6 (-2.53)	196.8 (2.89)	427.2 (2.99)	-338.8 (-3.99)	-121.6 (-5.22)	
(2) Cassava .027	.479	7.21	.031 (2.35)	-.487E-6 (-1.14)	64.7 (1.85)	-501.1 (-4.66)			-76.8 (-3.58)	109.4 (5.77)	28.9 (6.33)	
(3) Palm Oil[b] .075	.153	0.52	.123 (8.61)	-.839E-5 (-0.76)	200.9 (4.34)	211.7 (1.53)		84.7 (2.53)	-204.4 (-4.30)	117.7 (2.81)		
(4) Groundnut .026	.203	11.81	.027 (2.27)		-168.9 (-2.19)	153.2 (1.45)	-106.2 (-4.55)	-184.9 (-3.40)	368.6 (3.73)	-751.2 (-3.47)	c	PFF 214.3 (3.45)

NOTE: The following price variables were made available for each of the share equations: rice, cassava, palm oil, groundnut, dried fish and non-food, plus the price of the food to which the dependent variable refers. The remainder of the available set consisted of any variable available for the quantity regressions for the same commodity, except for region and the production and market factors. In running the quantity regressions a few variables had to be deleted because of multicollinearity within the available set, but with the regional variables missing, no other variables had to be deleted from the available sets for the share equations.

The variables selected for any given regression were those that maximized $\overline{R}^2$, unless TEXP was not included in that set. In the later case TEXP was added, but only if doing so had little effect on the coefficients of the remaining variables. The second Alcoholic Beverages regression was included because maximizing $\overline{R}^2$ gave no price response information.

The dependent variable is the ratio of household expenditure on the i^{th} food to total household expenditure $(\frac{q_i p_i}{TEXP})$.

Each independent variable has been deflated according to the following schema:

Variable	TEXP	p_i	Household characteristic
Deflator	p_i	TEXP	$TEXP/p_i$,

where p_i is the price of the commodity represented by the dependent variable and p_i the price of any other commodity.

Unless stated otherwise, each regression was fitted to consuming households only.

The t-ratios are in parentheses.

An entry like 1.71E-4 is to be read as 1.71×10^{-4}.

[a]The mean expenditure share was calculated over all households.

[b]Fitted to all the households in the sample. Non-consuming households were so few (ranging from none for rice to six for vegetables) as to warrant considering them as potential consumers, responding in the same fashion as the others to changes in the variables affecting consumption.

[c]Not included in the available set.

TABLE 5--Continued

Commodity[a]	Independent Variables									
	Household Characteristics									
	SIZE	INF	YCH	CH	AD	DEPR	WIV	AGEHD	LIMB	TEMN
(1) Rice .250		43.5 (2.25)	-34.4 (-2.27)					1.5 (1.74)		-114.0 (-1.89)
(2) Cassava .027			-41.8 (-1.42)				37.9 (1.15)		364.5 (3.45)	331.0 (4.04)
(3) Palm oil[b] .075						-6.1 (-1.77)				22.7 (2.04)
(4) Groundnut .026		8.4 (1.04)			7.5 (1.21)			0.9 (1.65)	249.7 (4.36)	154.7 (3.09)

TABLE 5--Continued

Commodity and Mean Share[a] of TEXP	$\overline{R}^2$	C_p	Constant Term	Independent Variables									
				Expenditure TEXP	Prices								
					PRB	PCA	PPO	PDF	PNF	PGN	POC	Other	
(5) Fish, fresh[d] .041	.511	10.15	.066 (4.18)	-.286E-5 (-0.97)	330.9 (3.26)		-25.2 (-1.39)	419.6 (6.10)	-705.7 (-5.87)	1468.8 (6.71)	c	PFF -286.2 (-4.75)	
(6) Fish, dried[b,d] .063	.346	-0.78	.051 (5.05)	-.284E-5 (-0.91)		-421.0 (-5.06)	19.9 (1.60)		68.0 (4.01)	-86.5 (-4.09)	c	PFF	
(7) Other cereals .045	.090	0.27	.136 (5.13)	-.123E-4 (-2.05)					-88.2 (-2.78)	88.2 (1.42)			
(8) Other legumes .014	.576	-1.61	.032 (2.70)	-.282E-5 (-1.31)			-10.7 (-2.15)				c	POL	
(9) Vegetables[b] .016	.230	-2.64	.005 (0.68)	.370E-5 (2.15)	-22.1 (-2.00)	162.7 (3.18)		-28.0 (-3.91)	28.4 (3.36)	-11.0 (-1.23)	c	PVG -21.2 (-4.01)	
(10) Fruits .001	.081	-1.85	.13E-2 (0.97)			-84.7 (-3.10)	-15.6 (-2.68)			7.3 (1.63)	c	PVG 8.8 (2.57)	PFT -29.7 (-2.54)
(11) Salt and other condiments[b] .023	.696	16.88	.014 (4.57)		-61.7 (-3.03)	107.6 (1.59)	-110.5 (-5.36)	109.8 (4.89)	-236.3 (-5.05)	702.5 (5.82)	c	PFF -201.0 (-6.04)	PCN 63.7 (6.01)
(12) Beverages, alcoholic .024 (1.1)	.225	-7.64	.016 (1.34)	.766E-5 (1.72)						c			
(1.2)	.221	-0.41	.005 (1.28)	.238E-4 (2.34)	-1139.8 (-2.29)	-1209.4 (-1.94)	143.5 (1.95)	186.5 (2.06)		c		PAB 220.0 (2.27)	

[d]The households in EA 13 were included when calculating the fish regressions.

TABLE 5--Continued

Commodity[a]	Independent Variables									
	Household Characteristics									
	SIZE	INF	YCH	CH	AD	DEPR	WIV	AGEHD	LIMB	TEMN
(5) Fish, fresh[d] .041	45.2 (2.80)	-43.2 (-2.81)	-30.6 (-1.83)	-54.6 (-4.06)	-58.4 (-3.05)	-17.8 (-1.56)				
(6) Fish, dried[b, d] .063	-3.01 (-1.34)								-59.1 (-3.59)	
(7) Other cereals .045						10.5 (1.78)			294.3 (1.27)	29.0 (1.42)
(8) Other legumes .014		5.6 (1.41)		12.8 (2.20)					63.7 (4.78)	
(9) Vegetables[b] .016										
(10) Fruits .001							-4.9 (-1.13)			
(11) Salt and other condimentsb .023	-1.2 (-1.86)	.92 (1.11)	1.2 (1.89)	1.7 (2.18)	1.0 (1.59)		-0.7 (-1.15)	0.1 (3.13)	4.2 (1.63)	3.3 (1.74)
(12) Beverages, alcoholic .024									259.2 (3.46)	
									1956.6 (2.59)	114.0 (1.53)

In the palm oil equation the signs of the price variable coefficients agree in the share and quantity equations, wherever the variable appears in both; in the cassava equation the signs agree with the exception of the coefficient for the price of rice. The negative sign of the coefficient of YCH in the quantity equations for rice and cassava is also confirmed by the share equations.

In the share equation regressions, the fish equations apply to a different set of households than were used for the quantity equation. The share equations include households in Enumeration Area 13, where fishing is an important commercial activity. Those households were included because we planned to use information from the share equation in designing the model to be used for the systems estimation. In systems estimation we cannot eliminate the EA 13 households from the fish equation without eliminating them from all equations, which we do not want to do.

In the share equation regression for fresh fish, household size and composition variables are very important. SIZE and the age-sex variables are statistically significant at levels ranging from 10 percent to less than one percent; DEPR is not quite significant at the 10 percent level. Given any set of values of the age and sex variables, SIZE can increase only when a person over 65 is added to the household. Thus both SIZE and DEPR will increase (but not by the same amount). The effect on fresh fish consumption will be the net effect of both influences, opposite in direction. If SIZE is given, the dependency ratio can change only if one or more of the age-sex variables also changes.

The dried fish regression contains no own-price variable. (The regression coefficient is not statistically significant at the 10 percent level). But the own-price quantity elasticity can still be calculated, for the price of dried fish enters the regression as a divisor in the term in TEXP and as a factor in each of the household variables.

For our last commodity group, alcoholic beverages, we present two regressions. The second has nearly as good a fit as does the first (measured by $\bar{R}^2$), but provides much more information concerning the relationship of the consumption of alcoholic beverages to the prices of other commodities. Neither the quantity equation for alcoholic beverages nor for palm wine gave information of that sort. High prices of rice and cassava are associated with reduced expenditure on alcoholic beverages; high prices of palm oil and dried fish are positively associated with such expenditure.

Although the share equation form sometimes eliminates heteroskedasticity, that was not so for these data. While the share equations are exactly equivalent to weighted regressions derived from the quantity equations by weighting each observation by p_i/y, this is not the correct weighting to use with these data in order to eliminate the heteroskedasticity problem.[1] Still, for some commodities the share equations provide information about cross-price relationships that is not available from the quantity regressions.

---

[1] The weighting best suited to eliminating heteroskedasticity from the share equations appears to be multiplication by the inverse of the predicted share of expenditure.

CHAPTER VI

QUANTITY REGRESSIONS BY GROUPS OF HOUSEHOLDS

Underlying all the regression analysis to this point has been the assumption that households have a common utility function. They make different consumption choices only because the independent variables (and the random component) in the function assume different values. But such a common utility function may not exist. There may be distinct groups of households that behave differently when confronted with the same values for the independent variables. Geographic location (or its ecological characteristics), ethnic composition, income level, attitudes toward production for the market, the type of rice culture practiced, household size and composition, or other socioeconomic characteristics of the household may identify groups whose behavior differs because their preference functions differ, not simply because they face different sets of opportunities and/or market prices.

Analysis by Groups of Households

Groups may behave differently because the utility functions to which they respond imply different forms for their demand functions or--a much less fundamental matter--because the values of the parameters differ even though the form of the demand function does not. In this case all households may be regarded as responding to utility functions of the same general class. We have not dealt with the first case, in which each group of households responds to its own form of demand function, but we have considered the second.

In this second case it is possible, in principle, to define a comprehensive demand function in which the arguments include variables identifying the different types of households. The demand function specific to a particular group of households would then be defined as the comprehensive demand function with the variables defining particular groups assuming values appropriate to those groups. This is what has been done, in part, in the regressions presented so far. Given the equation form being estimated,[1] the regression coefficient relating the dependent variable, q_i,

[1] The variables enter additively.

to one of the price variables, p_j, describes an average relationship for the whole sample, but the prediction for q_i differs among household groups because each of the variables describing a characteristic of a given group of households acts as a shift variable, raising or lowering the prediction of q_i in accordance with the demand function specific to that group of households. In effect, this coefficient shifts the constant term for all households possessing that characteristic.

The demand equation could also have been written to allow the slope coefficients to be adjusted in accordance with the characteristics of the various groups, but we did not do this. What we did do is allow the slope coefficients to vary by dividing the households into separate groups and fitting the regressions independently for each group. We present those results in this chapter. In terms of the specific equations we are using, each regression coefficient (slope coefficient) between q_i and p_j (or any other independent variable) assumes a value determined only by the group of households to which the regression is fitted. The slope coefficients are no longer constrained to be the same for all groups in the total sample. In addition, of course, the constant terms of the grouped regressions can vary among groups.

Given the linear form of the equations we are using, fitting the regressions to separate groups of households allows both slopes and constant terms to differ among groups; fitting the same form to the whole sample permits shifting the whole regression up or down (in effect, adjusting the constant term) to take account of differences among the preference functions of different groups. Clearly, fitting to groups of households separately gives more flexibility and responsiveness to the differences among groups, but there are disadvantages to this procedure.

Dividing the sample into groups reduces the number of observations to be used in fitting one regression. If the total sample is small this can seriously reduce the number of degrees of freedom and thus the number of independent variables that can be used. Even if this does not become a problem, having fewer observations tends to reduce the amount of variation in the independent variables and thus to lower the amount of information they contain and the statistical significance of the regression coefficients. (Yet an increase in the homogeneity of the group may increase the significance level of some of the coefficients.) As the price variation in our sample is geographical, grouping by region or by ethnic group (also distributed geographically) can reduce the amount of price variation by one half.

One consequence of these effects is an increase in multicollinearity and therefore in the standard errors, reducing the level of the t-statistics. Multicollinearity can become so high that a regression cannot be computed unless one or more variables is dropped. The ethnic grouping, which would otherwise have been a highly desirable grouping to employ, had to be dropped for this reason, while grouping by region was possible only for the rice and cassava equations.

Determining appropriate groupings is also a problem. Ethnicity and region certainly should be considered. (They are represented by binary variables in the whole-sample regressions, so no information is lost by treating all members of the group as alike.) But region is not as clear-cut a category as it might seem. Households on opposite sides of a regional boundary are probably more like each other than households in opposite corners of the same region.

Grouping by production patterns or market orientation is more difficult. Production patterns in Sierra Leone are extremely varied, usually involving a number of activities and shading gradually from one type to another. Spencer and Byerlee [1977, pp. 14, 53] identified 27 different farm types. Of the principal categories of activities involved, only cocoa and/or coffee production appeared at zero level in more than four farm types. Any grouping has to be based on percentages of effort or revenue from particular activities, thus converting a continuous variable to a class in which all households are viewed as alike despite their differences with respect to the very basis of classification.

If a group is to contain enough households to make fitting a regression possible it normally has to be defined in terms of only one or two characteristics, so it may still be less homogenous than is desirable. Furthermore, the best set of characteristics for one purpose may not be ideal for another. Grouping by income should certainly be considered, but a low money income does not necessarily mean low purchasing power, and low purchasing power for rice does not necessarily mean low purchasing power for cassava. Furthermore, grouping by income ignores relevant information from households that are excluded from the group but have much in common with those retained in the group.

Whatever the difficulties, if clear differences in behavior exist among households in different groups, we must identify those groups if possible and obtain the best measures we can of their behavior. The fundamental question is not whether the data should be analyzed by groups but what procedure is best suited for taking into account all factors that affect household behavior. In this instance we experiment with grouping as one possibility.

We have examined five of the most important foods--rice, cassava, palm oil, fresh fish and dried fish, dividing the whole sample in each case into four alternative groupings, by expenditure, by region, and by two variables representing farming practice: that is, the percentage of value output obtained from our list of specified activities (SHOSS) and the percentage of labor devoted to the production of upland rice (SHLUR).

For each grouping of households the procedure was to fit a regression using the variables included in the comparable quantity regression for the full sample. Then, for any given food, the deviations of the estimated from the actual quantities consumed by each household were calculated for each of the group regressions and the regression from the full sample. These residuals were squared, summed, and used in the Chow test to determine whether the regression coefficients in any regression based upon groups of households differed significantly from their counterparts in the other regression based upon that grouping. To be more precise, we tested the hypothesis that each regression coefficient (including the constant term) was equal in value to each of the comparable regression coefficients obtained from the other subgroups established by that grouping, or, in other words, the hypothesis that the overall regression was well specified--that there was not variation in coefficients across groups. The appropriate test statistic is the F-statistic.

The test is severe because it requires that the equality postulated among the different values of the same coefficient obtained from different regressions must hold for each and for all of the coefficients (including the constant term) in the regressions. The test can be failed either because one of the thirteen or fourteen coefficients in one equation lies outside the range of chance variation or because some set of coefficients as a group differs from its counterparts by more than is considered consistent with the hypothesis of equality. The rejection of the hypothesis for the latter reason may occur even when no single coefficient departs enough in value from other comparable coefficients to be regarded as refuting the hypothesis for that particular coefficient.

Groupings that Make a Difference

The specific regressions for which we calculated alternatives by household groups were the first equations for rice, cassava and palm oil in Table 2, plus the equations for fresh and dried fish. (The fish equations do not include the ten households in Enumeration Area 13.)

When grouped into three classes by TEXP, the hypothesis that the regression coefficients were equal fails for two of the five commodities: palm oil and cassava. The F-ratio for palm oil is 6.23 with 28 and 91 degrees of freedom; for cassava it is 2.18 with 30 and 69 degrees of freedom. For these cases the hypothesis of equal coefficients has to be rejected at the five percent level of significance. Grouping reveals different behavior patterns among these households. For the other three foods, classifying the households by TEXP causes no statistically significant difference in the coefficients.

Dividing the households into two groups according to the percentage of value product from activities on the specified list reveals only one food (palm oil) for which the coefficients of the new equations differ significantly among themselves. The F-ratio for palm oil is 4.06 with 14 and 105 degrees of freedom.

When divided into two groups by the percentage of labor devoted to the production of upland rice, the regressions do not differ significantly at the 5 percent level for any of the foods except cassava. For that the F-statistic is 2.11 for 15 and 84 degrees of freedom.

Regional grouping of the households creates such multicollinearity that it is impossible to calculate group regressions that have the same variables as the regressions for the whole sample, except for rice and cassava. Only for rice does the grouping alter the coefficients significantly. (The F-statistic is 4.27 with 10 and 115 degrees of freedom.)

Although we calculated regression equations for the four different groupings of households for each of the five important commodities, only for rice, cassava and palm oil does the statistical evidence justify rejecting the hypothesis that the coefficients are the same for each of the groups. Of course, the test applied only to the specific form of function used here. Had some other functional form been employed, the differences might have been either more or less significant.

The Regressions for Groups of Households

Table 6 (see page 84) presents the regressions for rice, cassava and palm oil, using the grouping for each commodity that provides the highest values of $\bar{R}^2$ for the sample as a whole.[1] Grouping by the share of output value obtained from activities in the selected list had a significant effect on the palm oil regression at the five percent level and grouping by the percentage of labor used for upland rice was nearly significant at that level for cassava, but we do not present these equations because grouping by TEXP provides better predicting equations (higher values of $\bar{R}^2$) for each commodity.

The added flexibility provided by grouping yields appreciably higher levels of $\bar{R}^2$ than were obtained from the regressions for the whole sample. (The whole-sample $\bar{R}^2$ values were .599, .481 and .529, for rice, palm oil and cassava, respectively.) However, the values of $\bar{R}^2$ for regressions for individual groups may be quite small and many t-statistics for individual regression coefficients are insignificant.

Let us examine the effect of regional grouping on our estimates of rice consumption. The regional variable in the full-sample regression[2] is no longer present. In the full-sample regression it adjusted the constant term upward for any household in Region 2; it has no role in the group regressions, for each regional regression has its own constant term. In testing for the equality of regression coefficients among the grouped regressions the constant term was not included. The test applied only to the slope coefficients of the two regressions.

In both regions rice consumption is positively associated with total expenditure through the lower two-thirds or more of the expenditure range, but the rate of increase diminishes with TEXP (Table 7). The rising portion of the consumption-expenditure curve[3] is much steeper in the North, but its curvature is much greater and the slope becomes negative at a lower

---

[1]Where $R^2 = 1 - \dfrac{\text{unexplained sum of squares}}{\text{total sum of squares}}$. We take the sum of the squared deviations of predicted from actual values of the dependent variable as the unexplained sum of squares. The predicted value for each household is the value predicted by the regression for the group to which the household belongs.

[2]Regression (1.1) in Table 2.

[3]The figures in Table 7 are the <u>slopes</u> of that curve.

level of TEXP. In both samples the relationship is statistically significant at the five percent level or better. The South, with 89 households, clearly dominates the regression for the ungrouped data.

According to Table 7, households in the North have a surprisingly high marginal propensity to consume rice at low income levels. Still, this finding is consistent with the rice consumption behavior per consumer equivalent revealed for that region by our 1980 tabular analysis [Smith et al., Table 3.4, p. 44].

The regional rice equations reveal that although the sign of the coefficient of the relative price of palm oil was negative in the whole-sample quantity equation, this situation holds only for households in the South (Regions One and Three), no doubt because the output of palm products per household in the North (Region 2) is much smaller than in the South [Spencer and Byerlee, 1977, Table 8.1, p 53]. For better understanding of the relationship between rice consumption and palm oil production, grouping by volume of palm oil produced might be better than the regional grouping.

The positive relationship between cassava price and rice consumption that was discovered for the whole sample in Regression (1.1) appears to have been based primarily upon the strong and statistically significant relationship that exists in Region 2. In the South the relationship, though still positive, is neither strong nor significant.

The positive sign for the coefficient of the number of wives of the household head holds only in Region Two. In the South, ethnic variables are of negligible importance; ethnic variation among the households in our sample was almost nonexistent in that regional group.

Grouping households by money expenditure levels reveals statistically significant differences at the five percent level with respect to the consumption of palm oil and cassava (though not with respect to rice). The expenditure groups used were (below 350 Leones), 350 but under 700 Leones, and 700 Leones and over. The mean TEXP values for these expenditure groups were 237, 513 and 1074 Leones, respectively.[1] In U.S. dollars the mean expenditure per capita in the highest expenditure group was $136 per year. For palm oil each expenditure class has about the same number of households (Table 6).

---

[1] In 1974-75 one Leone equalled U.S. $1.10 [Spencer and Byerlee, 1977, p. 24].

TABLE 6

QUANTITY REGRESSIONS FOR GROUPS OF HOUSEHOLDS

Commodity and Household Group	Mean Consumption of Consuming Households (Kilograms)	Number of Consuming Households	$\bar{R}^2$	Constant Term	Independent Variable								
					Expenditure		Prices						
					TEXP	$(TEXP)^2$	PRB	PCA	PPO	PDF	PNF	PGN	POC
(1) Rice-whole sample	589		.682[a]										
Region													
1 and 3	500	89	.541	-200.5	.188	-.128E-4		163.0	-79.7	243.9			
				(0.06)	(3.65)	(-2.12)		(0.01)	(-0.07)	(0.13)			
2	750	49	.796	-3571.7	.777	-.786E-4		3374.1	267.4	2483.4			
				(-2.30)	(4.96)	(-3.96)		(2.50)	(1.67)	(2.60)			
(2) Cassava-whole sample	394		.653[a]										
Expenditure group (Leones)													
Less than 350	225	33	.430	23.4	.018	.997E-6	-62.0			-25.8		70.8	14.7
				(0.05)	(0.12)	(0.10)	(-1.67)			(-0.58)		(1.88)	(1.61)
350 or less than 700	503	38	.701	232.3	-.384)	.137E-4	7.6			-134.6		154.5	56.3
				(0.37)	(-3.49)	(4.20)	(0.13)			(-3.45)		(3.85)	(5.35)
700 or more	427	43	.657	-16.9	-.034	.305E-6	-148.6			-94.0		225.3	42.7
				(-0.05)	(-1.03)	(0.61)	(-2.85)			(-2.03)		(5.95)	(3.88)
(3) Palm oil-whole sample	85		.646[a]										
Expenditure group (Leones)													
Less than 350	28	41	-.010	-30.8	.462	-5.08E-4				-16.8	-19.8	0.9	
				(-0.44)	(1.74)	(-1.52)				(-0.42)	(-0.32)	(0.02)	
350 or less than 700	64	46	.352	79.9	-.131	1.77E-4				9.7	-124.2	52.3	
				(0.42)	(-0.29)	(0.76)				(0.12)	(-1.56)	(0.44)	
700 or more	157	46	.640	-224.2	.143	-.32E-4				337.2	-520.0	916.7	
				(-1.17)	(1.47)	(-1.77)				(3.19)	(-4.06)	(3.66)	

NOTE: In this table the absence of a regression coefficient means that the corresponding variable was not available for the regression.

[a]This is the $\bar{R}^2$ for the combined set of grouped equations.

[b]This variable is omitted from this regression because Regions 1 and 3 have no households in the Limba group.

TABLE 6--Continued

Commodity and Household Group	Independent Variable															
	Household Characteristics										Production and Market Factors					
	SIZE	INF	YCH	CH	FAD	DEPR	WIV	AGEHD	LIMB	TEMN	SHOCC	SHOLSO	SHOSS	SHLUR	MKTOR	SHCPH
(1) Rice-whole sample																
Region																
1 and 3		73.7	-11.3	139.2			-1.1		b	-9.6						
		(2.17)	(-0.33)	(3.52)			(-0.03)			(-0.07)						
2		18.1	-24.0	-24.2			29.4		-534.7	-253.3						
		(0.66)	(-0.90)	(-0.71)			(0.75)		(-2.62)	(-1.19)						
(2) Cassava-whole sample																
Expenditure group (Leones)																
Less than 350	-13.5		-102.4			134.0						-2.83	1.58	1.40	-2.65	-1.79
	(-0.41)		(-1.11)			(1.48)						(-0.49)	(0.31)	(0.50)	(-0.41)	(-1.08)
350 or less than 700	118.1		178.8			-480.5						-19.84	16.37	14.30	-18.95	5.14
	(2.81)		(1.73)			(-2.74)						(-1.55)	(2.81)	(4.85)	(-2.73)	(1.99)
700 or more	33.4		-117.5			118.3						-10.64	4.77	7.73	-6.89	-4.11
	(1.61)		(-2.07)			(1.17)						(-1.31)	(1.02)	(2.08)	(-1.36)	(-1.87)
(3) Palm oil-whole sample																
Expenditure group (Leones)																
Less than 350	-5.8	5.7			4.3		7.3	-.2			.01		.15			.06
	(-1.61)	(0.70)			(0.37)		(0.67)	(-0.34)			(0.03)		(0.59)			(0.52)
350 or less than 700	2.5	-9.5			-5.2		-18.5	1.1			-.66		.72			.64
	(0.38)	(-0.89)			(-0.30)		(-1.29)	(1.36)			(-0.41)		(0.69)			(2.33)
700 or more	-18.2	72.8			81.4		-51.2	2.6			-2.77		1.81			.69
	(-1.87)	(3.68)			(3.78)		(-2.24)	(2.31)			(-1.38)		(1.95)			(1.16)

TABLE 7

MARGINAL PROPENSITY TO CONSUME RICE:
ADDED RICE CONSUMPTION (IN KILOGRAMS)
ASSOCIATED WITH AN ADDED KILOGRAM OF TEXP
(MEASURED IN POWER TO PURCHASE RICE)

TEXP (Kilograms of Rice)	Region		Entire Sample
	South (Regions 1 and 3)	North Region 2	
1058	0.16	0.61	.22
2120[a]	.13	0.44	.18
2680[b]	.12	.36	.16
4943	.06	0.00	.09
6860[c]	.01	-0.30	.02
7344	0.00	d	.00[e]
9576	-0.06	d	-.08
Number of Households	89	49	138

[a]Mean TEXP in North.

[b]Mean TEXP in South; the mean for the entire sample is 2481.

[c]Maximum TEXP in North.

[d]Outside the range of the sample in the North.

[e]Exactly zero at TEXP = 7358.

For palm oil the signs and magnitudes of most regression coefficients differ appreciably by expenditure groups. Relatively few regression coefficients are statistically significant except in the highest expenditure group. That group seems to dominate the results from the whole-sample regression; at least, the signs of the coefficients for the highest income group are the same as those for the sample as a whole.

But for four variables the direction of response is the same in each expenditure group (although the magnitude may differ greatly). Palm oil consumption is positively related to the price of groundnuts and negatively related to non-food prices, while both the share of value output coming from "specified sources"[1] and the share of consumption that is produced at home are positively associated with palm oil consumption, as is the case for the whole-sample regression. For these four variables, as for the other non-expenditure variables, the magnitude of response increases as does the level of the expenditure group.

At and below the mean of each expenditure group there is a small but positive relationship between the level of TEXP and household consumption of palm oil, but in the low and high expenditure groups the relationship becomes negative before the maximum income level for the group is reached (Table 8). In the middle income group, for which the TEXP coefficients are not statistically significant, the marginal propensity to consume palm oil increases with TEXP.

The full-sample consumption-expenditure curve is convex from above, rising slowly to its maximum at a TEXP level of 2988 kg of palm oil--in effect, a smoothed version of the three consumption-expenditure curves found in the data by expenditure groups: convex upward curves in the low and high expenditure groups, with a concave upward curve filling the middle range.

As previously stated, the data on cassava consumption are much less reliable than our figures for the other major foods. Nonetheless, the regressions do provide information of interest. In the low expenditure group (when households are grouped by expenditure levels), the t-values of most of the individual regression coefficients are very small, but they tend to be considerably larger in the two higher groups, perhaps because the greater range of expenditure levels in the two higher groups brings

---

[1]Which includes palm products.

TABLE 8

MARGINAL PROPENSITY TO CONSUME PALM OIL:
ADDED PALM OIL CONSUMPTION (IN KILOGRAMS)
ASSOCIATED WITH AN ADDED KILOGRAM OF TEXP
(MEASURED IN POWER TO PURCHASE PALM OIL)

TEXP (Kilograms) of Palm Oil)	Expenditure Group (Leones)			
	Low (Less than 350)	Middle (350 but under 700)	High (700 or More)	Entire Sample
376[a]	.08			.17
693[b]	-.24	.11		.15
874[c]		.18		.14
1732[d]			.03	.08
2988			-.05	.00
5061[e]			-.18	-.14
Number of Households	41	46	46	133

NOTE: There is no entry if the TEXP level lies outside the range of the sample for a given expenditure group.

[a]Mean TEXP, low group.

[b]Maximum TEXP, low group.

[c]Mean TEXP, middle group.

[d]Mean TEXP, high group.

[e]Maximum TEXP, high group.

with it more variation in the levels of the independent variables. In each expenditure group the t-values tend to be higher for price variables than for other variables.

For three out of the four price coefficients the grouped equations confirm the signs obtained in the regression for the full sample. The negative sign of the dried fish coefficient persists in each group, but the negative coefficient for the price of rice in the full-sample regression only appears in high and low expenditure groups. (The coefficient for the middle group is positive, but only about one-eighth of its standard error.)

Households in the low and high expenditure groups seem dominant in determining the whole-sample signs of the coefficients for YCH (young children) and DEPR (the dependency ratio).

The expenditure relationship for the sample as a whole is very weak; the data by groups reveal a stronger, statistically significant relationship for households with expenditures between 350 and 700 Leones per year and identify that relationship as negative but concave upward for households with total expenditures in real terms (in kilograms of cassava) of 14,014 kg or less. Above that figure consumption rises with TEXP, although this seems a most unlikely turnaround (Table 9). In the low expenditure group consumption rises slightly through the range of expenditures within the group (but the expnditure coefficients are only about one-tenth of their standard errors). In the highest expenditure group the relationship is weak, not statistically significant, and negative over the whole range of expenditure levels within the group.

Although the consumption-expenditure curve for each of the higher expenditure groups eventually turns upward, for the sample as a whole there is a minute downward drift for TEXP $\geq$ 5700 kg. Below that figure the relationship is positive, very weak and statistically insignificant. In short, the data do not provide convincing support for the hypothesis that cassava consumption is inversely related to income, but neither do they refute it. We can only conclude with confidence that the expenditure-consumption relationship is very weak and of extremely little consequence.

The negative relationship detected by the whole-sample regression between cassava consumption and either market orientation or SHOLSO (the share of value output from labor sold out) is found to exist in each expenditure group, but is much stronger in the two higher groups. However, production activities on the specified list (SHOSS) appear to be positively

TABLE 9

MARGINAL PROPENSITY TO CONSUME CASSAVA:
ADDED CASSAVA CONSUMPTION (IN KILOGRAMS) ASSOCIATED
WITH AN ADDED KILOGRAM OF TEXP (MEASURED IN
POWER TO PURCHASE CASSAVA)

TEXP (Kilograms of Cassava	Expenditure Group (Leones)			Entire Sample
	Low (Less than 350)	Middle (350 but under 700)	High (700 or more)	
6048[a]	.03	-.22	-.030	.000
11980[b]	.04	-.06	-.027	-.002
20946[c]		+.19	-.021	-.004
52475[d]			-.002	-.013
Number of Households	33	38	43	114

NOTE: There is no entry if the TEXP level lies outside the range of the sample for a given expenditure group.

[a]Mean TEXP, low group.

[b]Mean TEXP, middle group.

[c]Mean TEXP, high group; maximum TEXP for the middle group is 23,300 kg.

[d]Maximum TEXP, high group.

related to cassava consumption, especially in the two higher income groups. These two measures of attitudes toward production for the market yield opposite results. As has been widely asserted, producing upland rice is favorable to cassava consumption--in all expenditure groups. Consuming cassava from one's own production, however, increases consumption levels only in the middle expenditure groups.

The Choice among Regressions

Given the regressions presented in this report, it is appropriate to ask which of them are likely to be the better predictors. Yet both the whole-sample and the group regressions can still be improved, and with such improvement the nature of the choices to be made might change.

If the whole-sample quantity regressions are to be used as bases for policy decisions, their heteroskedasticity should be removed or reduced by fitting weighted regressions, weighting by the inverse of the predicted consumption. This would reduce the influence of large consumers.

Likewise, we could modify the regressions by expenditure groups by imposing the constraint that the regressions for two adjacent groups give equal predicted consumption figures for a household on the border between those groups. This probably would moderate some of the more extreme responses that we now have (for instance, the responses for palm oil and cassava in the middle expenditure group). Presumably this would make the results by groups more like the whole-sample results.

However that may be, we shall indicate which of the regressions presented here we believe to be the better predictors. For palm oil, cassava and rice the regressions by groups give better predictions for the sample as a whole than the single regression for the whole sample. (The $\bar{R}^2$ values are appreciably higher from the regressions by groups.) With one exception, however, each regression by groups is based upon only 1/4 to 1/3 as many observations as the regression for the whole sample; thus it runs a greater risk of being non-representative, has fewer degrees of freedom, and is more affected by multicollinearity; likewise there is also likely to be less variability in the observed values of the independent variables, and, in the case of grouping by TEXP, some loss of information by the exclusion of households that are often quite like those within the group. Furthermore, the process of splitting the sample into subgroups may result in error terms that are not normally dsitributed within each subgroup even though they were normally distributed for the sample as a whole. Consequently, as the expected value of the error term within any group is no longer zero, biases may be present in the coefficient estimates for subgroups that do not exist in the estimates for the sample as a whole.

On balance, under the circumstances, the whole sample regressions seem more reliable than the regressions by groups, although the latter can be useful in identifying variables to which the response appears to differ greatly among groups. Thus they indicate aspects of the relationships that deserve more careful study when time and the data permit.

The marked differences among the expenditure coefficients for the groups within each set of commodity regressions raise doubts about their dependability, given the small number of cases on which the results depend. Likewise the number of statistically insignificant regression coefficients is large and in several of the group regressions more regression coefficients are statistically insignificant than not.[1] Lastly, while in the palm oil and cassava equations there are reasonably high values for $\bar{R}^2$ for at least one group (see Table 6), the values are also quite low for one or more groups (as low as -.010 in one case). This is, for one or two of the groups the reliability of prediction is low. It is not clear that we would gain much by increasing the value of $\bar{R}^2$ for the whole sample through using the group regressions if in doing so we obtained very poor estimates for some groups. A conservative approach suggests using the whole-sample regressions, but remembering that behavior within certain groups may depart appreciably from the average.

Still better prediction equations could be developed for the expenditure groups, but only at the expense of losing some of the information we were most interested in obtaining from this study. If it is indeed true that the households in different groups behave differently, and that the statistical evidence obtained thus far is not simply the result of the functional form being used, one could improve the estimates for the individual groups

---

[1]Of course, statistically insignificant regression coefficients do not necessarily rule out using a regression for prediction. Where multicollinearity is great the reliability of parameter estimates may be reduced, even though the regression as a whole is reliable for prediction. (Where certain independent variables are highly correlated the regression may represent the total effect correctly even though the total influence is incorrectly divided among the independent variables.)

by dropping the requirement that for a given commodity the same independent variables be used for each group of households. Given the frequency of low t-statistics in the regressions by groups, higher values of $\bar{R}^2$ for some or all of the groups should be attainable by using only the set of independent variables that maximizes $\bar{R}^2$ in each case.

Still further improvements might be obtained by using a different functional form for each group. In that case, the effect of using the best equation for each group would probably be to increase the $\bar{R}^2$ for the complete set of estimates for the sample, but not necessarily. For instance, the best estimates for the sample as a whole might still require the assumption that the regression coefficients be the same across groups for at least some variables.

The possible advantages of improving the group estimates by dropping variables must be balanced against the loss of information that would thereby occur. The palm oil regression for low-expenditure households (Table 6) might turn out to have only SIZE and TEXP as explanatory variables. Given the size of the sample, we must turn to the full-sample regression to obtain any information about responses to prices or to any other variable other than SIZE and TEXP, even though we know that the full-sample regression represents the sample as a whole rather than any group in paricular.

Desirable as it may be to have estimates specific to the low-expenditure group, we must question the worth of such detail, given the loss of information about important policy variables that would result. To be sure, we have a special interest in low-income households, but in rural Sierra Leone almost all households are low-income households in a broader sense. The mean annual expenditure of the 138 households in the sample is 660 U.S. dollars. At the sample average of 6.55 persons per household this amounts to only $101 per capita per year. The median annual expenditure per capita, assuming the same household size, is $83, while the mean per capita expenditure in the highest expenditure group is only $136.

Summary

Consumption behavior differs among expenditure groups with respect to palm oil and cassava but not with respect to rice.[1] Regional grouping,

---

[1]Given the particular functional form we are using.

however, reveals statistically significant differences in rice consumption. Grouping by region, expenditure or two variables associated with farming practice does not, however, reveal statistically significant differences in behavior with respect to either dried or fresh fish.

Unfortunately, a number of the regressions obtained for groups of households do not inspire confidence, given the small sample size for most groups and the high degree of multicollinearity in the data. Conservative practice suggests that the whole-sample regressions be used, but that the user remember that if more data were available grouping might yield still further improvements in the estimates.

CHAPTER VII

ELASTICITIES

Judging the strength of a relationship between two variables by the size of a regression coefficient is impossible unless one knows the units in which they are expressed. Even when the units are known such judgments can be difficult and time-consuming. The problem is eliminated when all changes in variables are expressed in percentage terms, as is done when an elasticity is computed.

The elasticity of Y with respect to X is the percentage change in Y in response to a given percentage change in X (calculated for an infinitesimally small change in X). Using this measure, elasticities can be compared directly for commodities expressed in different units, for real incomes measured in the power to purchase different commodities, or for any set of variables whatever.

We present here only the expenditure and price elasticities--(the elasticities of quantity available for household consumption with respect to expenditure or price). The formulae, as given in Chapter III, are as follows:

Expenditure[1]

$$\frac{\partial q_i}{\partial y}\frac{y}{q_i} = [b_1 + 2b_2(y/p_i)]\frac{y}{p_i q_i} . \qquad (8)$$

Own-Price

$$\frac{\partial q_i}{\partial p_i}\frac{p_i}{q_i} = -1 + [a_i - b_2(y/p_i)^2 + g]/q_i, \text{ where } g = \sum_k c_k h_k + \sum_m d_m v_m + \sum_n e_n r_n .^1 \qquad (6)$$

Cross-Price

$$\frac{\partial q_i}{\partial p_j}\frac{p_j}{q_i} = \frac{a_j}{p_i}\frac{p_j}{q_i} = \frac{a_j}{q_i}\frac{p_j}{p_i} . \qquad (7)$$

[1]Not the elasticity of expenditure, but the elasticity of the quantity available for consumption with respect to expenditure.

In each case, quantities are measured in kilograms, income is in Leones, and prices are in Leones per kilogram.

Because these elasticities vary with price and expenditure levels, and (in the case of the own-price elasticity) with the levels of other variables in the regression, we present sample values calculated at the mean levels of the variables for the households in each expenditure group, using all sample households in that expenditure group in order to have the same set of households for each commodity. (Both consuming and non-consuming households are included when calculating these mean values.) Price and expenditure values are measured in Leones. The mean quantity value, q_i, is the mean of the predicted values of q_i at the mean levels of the independent variables for the expenditure group.[1]

Tables 10, 11 and 12 give price and expenditure elasticities for the three types of regressions we have presented in Tables 2, 5 and 6. In each case, if more than one version of the regression was presented for a particular food, elasticities were computed for only the first.

It is clear from these values that expenditure and price elasticities play important roles in the allocation of foods (and therefore nutrients) among households. Strong expenditure responses (almost invariably positive, except for cassava) occur often for rice, palm oil, fish, vegetables, and alcoholic beverages, not to mention Maggi cubes and kola nuts. Own-price elasticities are frequently large (usually negative) for rice, cassava, groundnuts and dried fish, as well as for Maggi cubes and kola nuts. Cassava, palm oil, groundnuts, fish and vegetables (and Maggi cubes and kola nuts) often have large cross-price elasticities with respect to the prices of a number of other commodities. The commodities most often giving rise to large cross-elasticities are dried fish, non-food, rice, groundnuts, palm oil and cassava.

The reasonably high values of the own-price elasticities for various staple goods (the share elasticities tend to be a good deal higher than the quantity elasticities) reinforce the views of Mellor and Timmer that price can be a powerful short-run allocator of food intake. Mellor concentrates on income effects, which we find clearly important. However, not all the price effect is through the effect of price on real income.

---

[1]The mean values of TEXP for the three expenditure groups are 237, 513 and 1074 Leones, respectively.

TABLE 10

EXPENDITURE ELASTICITIES

Commodity	Expenditure Level								
	Low			Medium			High		
	From Quantity Regression		From Share Equation	From Quantity Regression		From Share Equation	From Quantity Regression		From Share Equation
	Whole Sample	Household Group		Whole Sample	Household Group		Whole Sample	Household Group	
Rice									
Region									
1 and 3	-2.86	1.01		1.15	.63		.68	.49	
2	.32	.75		.37	.39		.33	.27	
Whole sample	.87		1.03	.75		.96	.48		.67
Sorghum	.12			.12			.28		
Cassava	.00	.87	1.47	-.04	-16.92	.75	-.16	-1.25	.89
Palm oil	3.92	1.01	1.96	1.76	2.89	1.62	.80	.33	1.44
Groundnut	-.19		.77	.11		.70	.24		.93
Broadbean	.07			.14			.18		
Fish, fresh	1.09		.97	.88		1.05	1.36		.79
Fish, dried	.51		.56	1.18		.78	1.92		.68
Onions	.72			.63			.60		
Peppers and chillies	.18			.48			.75		
Salt	.26			.35			.33		
Maggi cubes	.70			1.77			1.05		
Kola nut	.72			2.03			2.59		
Other cereals	.31		.85	.26		.71	.24		1.09
Other legumes	1.24		.63	1.65		.70	1.35		.58
Vegetables	.25		.94	.59		1.16	.91		1.27
Fruits	.39		-.03	1.29		-.06	.85		-.12
Salt and other condiments	.53		.44	.70		.77	.63		.83
Beverages, alcoholic	1.20		.16	.85		.41	1.03		1.44

TABLE 11

OWN-PRICE ELASTICITIES

Commodity	Expenditure Level								
	Low			Medium			High		
	From Quantity Regression		From Share Equation	From Quantity Regression		From Share Equation	From Quantity Regression		From Share Equation
	Whole Sample	Household Group		Whole Sample	Household Group		Whole Sample	Household Group	
Rice									
Region									
1 and 3	2.97	-1.51		-1.41	-.97		-.79	-.68	
2	-.33	-5.61		-.45	-2.34		-.38	-2.02	
Whole sample	-.90		-2.53	-.92		-1.58	-.56		-1.12
Sorghum	.83			.42			.43		
Cassava	-.70	-1.28	-4.22	-.86	4.48	-2.17	-.47	-.32	-1.87
Palm oil	4.47	-.23	-.91	.43	-.93	-.94	.21	-.10	-.83
Groundnut	22.5		-19.36	-2.93		-8.99	-2.28		-6.67
Broadbean	-.05			-.12			-.16		
Fish, fresh	.89		-8.65	-.14		-5.03	.57		-2.37
Fish, dried	-.72		-1.31	-1.06		-1.22	-2.19		-.94
Onions	-.72			-.63			-.60		
Peppers and chillies	6.99			4.48			1.73		
Salt	-1.15			-1.04			-.74		
Maggi cubes	-1.39			-2.09			-1.19		
Kola nut	-.66			-1.49			-1.98		
Other cereals	.74		.07	.17		-.33	.06		-.69
Other legumes	3.87		-.00	1.96		-.36	1.79		-.30
Vegetables	-.25		-1.26	-.59		-1.40	-.91		-1.53
Fruits	.36		-.74	.27		-.67	.10		-.72
Salt and other condiments	-.05		11.82	-.33		8.08	-.37		3.50
Beverages, alcoholic	-1.20		-.16	-.85		-.41	-1.03		-1.44

TABLE 12

CROSS-PRICE ELASTICITIES

Commodity	Price	Low: From Quantity Regression: Whole Sample	Low: From Quantity Regression: Household Group	Low: From Share Equation	Medium: From Quantity Regression: Whole Sample	Medium: From Quantity Regression: Household Group	Medium: From Share Equation	High: From Quantity Regression: Whole Sample	High: From Quantity Regression: Household Group	High: From Share Equation
Rice										
Region										
1 and 3	Palm oil	3.74	-1.34		-.79	-.44		-.35	-.26	
	Fish, dried	-2.85	1.66		.78	.70		.33	.39	
	Cassava	-1.01	.18		.27	.08		.14	.05	
2	Palm oil	-.42	.87		-.26	.28		-.17	.26	
	Fish, dried	.32	3.26		.25	1.37		.16	1.18	
	Cassava	.11	.74		.09	.30		.07	.32	
Whole sample	Palm oil	-1.13		-1.62	-.52		-.75	-.25		-.35
	Fish, dried	.86		.80	.51		.48	.23		.21
	Cassava	.31		.50	.18		.29	.10		.16
	Other cereals			-.46			-.89			.37
	Groundnuts			-1.36			-.69			-.34
	Non-food			4.12			2.18			1.14
Sorghum	Rice	.99			.41			.56		
	Cassava	-.42			-.22			-.31		
	Non-food	-1.52			-.73			-.96		
Cassava	Rice	-3.01	-1.99	3.29	-1.18	.58	1.03	-1.22	-2.04	1.07
	Fish, dried	-1.62	-.86		-.82	-18.80		-.71	-1.44	
	Other cereals	1.57	.94	3.50	1.25	17.38	1.87	.99	1.83	1.49
	Groundnuts	3.76	2.31	6.75	1.64	13.29	1.97	1.56	3.23	1.89
	Non-food			-11.42			-3.45			-3.46
Palm oil	Rice			3.41			1.17			.60
	Fish, dried	2.65	-.211	1.49	.79	.09	.66	.29	.97	.28
	Cassava			.62			.27			.14
	Groundnuts	4.94	.01	2.03	1.27	.42	.77	.51	2.47	.37
	Non-food	-15.99	-.59	-8.50	-4.25	-2.46	-3.35	-1.81	-3.66	-1.67

TABLE 12--Continued

Commodity	Price	Expenditure Level								
		Low			Medium			High		
		From Quantity Regression		From Share Equation	From Quantity Regression		From Share Equation	From Quantity Regression		From Share Equation
		Whole Sample	Household Group		Whole Sample	Household Group		Whole Sample	Household Group	
Groundnut	Rice			-5.09			-1.92			-1.48
	Palm oil			-8.21			-3.11			-2.11
	fish, dried	13.74		-5.78	-2.05		-2.81	-1.14		-1.81
	Fish, fresh			9.45			3.87			1.96
	Cassava	-5.69		.80	.83		.38	.57		.30
	Non-food	-63.97		27.24	8.50		11.80	5.49		8.84
Broadbean	Fish, dried	-.01			-.02			-.02		
Fish, fresh	Rice	-7.34		5.46	-2.84		2.64	-4.67		1.50
	Palm oil	5.46		-1.07	2.12		-.52	3.07		-.26
	Fish, dried	1.84		7.18	.92		4.49	1.26		2.13
	Cassava	-1.93			-.94			-1.58		
	Groundnuts			24.66			13.26			6.96
	Non-food			-28.56			-15.89			-8.76
Fish, dried	Rice	1.49			1.42			1.34		
	Palm oil			.66			.43			.25
	Fish, fresh	-1.69			-1.76			-1.09		
	Cassava	-1.53		-.94	-1.85		-.76	-1.78		-.52
	Groundnuts	1.94		-1.14	2.07		-.81	1.79		-.50
	Non-food			2.17			1.59			1.04
Peppers and chillies	Cassava	-9.06			-6.15			-2.86		
	Onions	1.89			1.19			.38		
Salt	Rice	-1.11			-.71			-.59		
	Fish, fresh	.44			.31			.17		
	Cassava	-.67			-.54			-.46		
	Non-food	2.23			1.65			1.31		
Maggi cubes	Rice	2.31			2.72			1.40		
	Cassava	-1.62			-2.40			-1.26		

TABLE 12--Continued

Commodity	Price	Expenditure Level								
		Low			Medium			High		
		From Quanity Regression		From Share Equation	From Quantity Regression		From Share Equation	From Quantity Regression		From Share Equation
		Whole Sample	Household Group		Whole Sample	Household Group		Whole Sample	Household Group	
Kola nut	Rice	-2.43			-3.07			-2.50		
	Palm oil	3.69			4.68			3.35		
	Fish, dried	-1.32			-2.15			-1.46		
Other cereals	Cassava	.78			.36			.53		
	Groundnuts	1.88			.76			1.02		
	Non-food	-3.70			-1.54			-2.18		
Other legumes	Palm oil			-.63			-.34			-.28
	Groundnuts	-8.55			-6.19			-5.57		
	Non-food	3.44			2.58			2.43		
Vegetables	Rice			-2.32			-.78			-.24
	Fish, dried			-3.06			-1.32			-.34
	Cassava			2.95			1.25			.39
	Groundnuts			-1.18			-.44			-.12
	Non-food			7.33			2.82			.83
Fruits	Palm oil			.91			.84			.81
	Cassava	-2.35		.33	-4.02		.39	-1.57		.43
	Groundnuts			-.17			-.17			-.17
	Vegetables	1.59		-.31	2.47		-.33	.62		-.23
Salt and other condiments	Rice			-2.04			-1.49			-.93
	Palm oil			-9.36			-6.86			-3.78
	Fish, fresh			-9.71			-7.70			-3.15
	Fish, dried			3.76			3.54			1.85
	Cassava	-.48		.61	-.37		.56	-.26		.36
	Groundnuts			23.61			19.15			11.05
	Non-food			-19.14			-16.06			-9.73

For instance: the own-price elasticity for rice, as estimated from the households in Region 1 and 3 (the South and East), is -.97 at the medium expenditure level, and the corresponding expenditure elasticity is .63. The mean share of expenditure devoted to rice by the middle-expenditure group of households was 24.6 percent[1], so a one percent rise in the price of rice is approximately equivalent to a fall of .246 percent in the purchasing power of household expenditure. The income effect of such a fall in purchasing power is to reduce rice consumption by approximately .16 percent.[2] Of the total own-price elasticity of -.97, the remainder, -.81, is a substitution effect. Clearly there are substitution (and production) effects, the former of which are ignored by Mellor [1978] but not by Timmer [1978]. In the case of rice the substitution effect reinforces the income effect. No commodity other than rice and "non-food" represents more than 7 1/2 percent of total expenditure on the average, so the income effect will normally be an even smaller proportion of the total price effect.

Another feature of these results is the dramatic change in elasticities that often occurs as expenditure levels change: for rice, cassava and palm oil expenditure elasticities decline as expenditure rises; for dried fish, kola nuts and vegetables they increase. Declines in the absolute values of own-price elasticities occur for a number of foods, including rice, fresh fish, peppers and chillies, salt, and "other cereals," while marked declines in the absolute values of cross-elasticities take place for rice, palm oil and groundnuts. In part this is because budget shares for most foods tend to decline at higher expenditure levels, thus reducing the income effect component of the price elasticity, but in some cases declining expenditure elasticities at higher expenditure levels also play a part.

Given the frequency of falling elasticities and large cross-elasticities for major foods, we must conclude that the allocation effects of price and income changes are particularly important for low-income households. Responses to prices and income (as indicated by these elasticities) may significantly affect the nutrition of these households.

---

[1] The share for low-expenditure households was 24 percent; for high-expenditure households 23 percent.

[2] (.00246) x (.63) = .00155.

We note by examining Tables 10-12 in more detail that elasticity estimates often vary widely between the quantity and share regressions (elasticities from the share regressions often being much higher in absolute value) and between the regressions for groups of households and for the entire sample. To be sure, we have only presented regressions for groups of households where grouping affected the results significantly. If we compare only the values for grouped households and those for the sample as a whole, we believe the whole-sample elasticities to be more dependable, even though they are less affected by behavior differences among groups.

The choice between share-equation and quantity-equation elasticities[1], both calculated from the entire sample, raises different considerations. Differences between the estimates are caused in part by the exclusion of production and market orientation variables from the share equations, in part by the use of different constant terms in the two equations, and in part by the different equation forms themselves. In principle, the form to be used should depend on the statistical properties of the error terms. (The form for which the disturbances are homoskedastic will give the best estimates.) In our case we have heteroskedastic disturbances with either form of the regression. Study of the disturbances suggests that the heteroskedasticity could be substantially reduced with either form of the equation by using a weighted regression. For the quantity regressions, weighting by the inverse of the predicted quantity appears to be appropriate, while for the share equations, weighting by the inverse of the predicted expenditure share is appropriate. However, we did not calculate the weighted regressions for either form. As between the unweighted forms of the quantity and share regressions, we believe that the quantity regressions are the better predictors. Using the share form did not eliminate heteroskedasticity; the share regressions do not contain the production and market orientation variables; and the elasticity values implicit in the share regressions are less plausible.

Tables 10-12 contain three sets of elasticities for rice that are identified as "Whole Sample" elasticities. Each is derived from the whole-sample regression, (1.1) in Table 2. The elasticities for Regions 1 and 3 (the South and East) were obtained by setting the regional variable, REG 2,

---

[1]In each case, the elasticity of quantity in kilograms with respect to price in Leones per kilogram.

equal to zero in the whole-sample regression, and those for the North by setting REG 2 = 1. Aside from the value chosen for REG 2, the elasticity at each expenditure level was evaluated at the mean value of the independent variables for all households in the corresponding expenditure group (regardless of region).

To calculate the whole-sample elasticity from the whole-sample regression, REG 2 was set equal to its mean value for all households in the expenditure group under examination; all other right-hand-side variables were taken at the same levels as stated above.

The elasticities in the column headed "Household Group" are estimated by using regressions based only on data from the households in the group specified (in the case of rice, the regional groupings). Except for rice, all household groupings are by expenditure.

The rice expenditure elasticities in Table 10 form a stable and consistent pattern, with the exception of the elasticity at the low expenditure level calculated for Regions 1 and 3 from the whole-sample regression.[1] The Region 2 elasticities by expenditure level make it clear that the estimates of marginal propensity to consume for the Northern Region (given in Table 7) are not as surprising as they had seemed when evaluated out of context.

The expenditure elasticities for cassava are greatly affected by the type of regression used. The whole-sample quantity elasticities are plausible (very close to zero and negative at the higher expenditure levels), but the share regressions and regressions by expenditure groups give very different results. Completely implausible is the elasticity of -17 based on middle-expenditure households. This is associated with the convex upward segment of the consumption-expenditure curve for cassava, remarked upon in Chapter VI (Table 9).

Palm oil expenditure elasticities are positive and generally high, except at high expenditure levels. The elasticities for fresh and dried fish generally establish reasonable ranges for these quantities, the

---

[1]That elasticity (-2.86) is negative because the predicted value for mean household consumption at the low expenditure level is negative. If actual consumption values for these households were used, the elasticity would by positive. The expenditure elasticity, $\frac{\partial q}{\partial y}(\frac{y}{q})$, may also be written $\frac{\partial q}{\partial y} \div \frac{q}{y}$, or marginal propensity to consume (mpc) ÷ average propensity to consume (apc). The marginal propensity to consume rice as TEXP rises is positive at this level.

quantity and the share equation results supporting each other quite well.[1]

Rice consumers are highly responsive to the price of rice (Table 11). According to the preponderance of evidence, the own-price elasticities of rice consumption are negative, large (in absolute value) and decline sharply in the higher expenditure groups, although the whole-sample results from the quantity equations contradict this conclusion sharply for Region 2 and yield a rather non-committal average value (still only slightly below unity in the low- and medium-expenditure groups) for all regions together. The whole-sample regression also yields an improbable positive value for the low-expenditure group in Regions 1 and 3 (the South and East).[2] When households are grouped by region, the North is much more price-responsive than the South and East.

The cross-price elasticities for rice with respect to the prices of palm oil, dried fish, groundnuts and non-food are large for low-expenditure households, falling rapidly in absolute value for the higher expenditure groups.[3] The prices of cassava and other cereals have relatively little effect on rice consumption. In the South and East, where the value of the output of palm products is much larger than it is for most Northern households, the cross-elasticity with respect to the price of palm oil is negative--high palm oil prices are associated with reduced rice

---

[1]The ten households in Enumeration Area 13 were omitted when fitting the share regressions for fish.

[2]This own-price elasticity of +2.97 is positive because the whole-sample prediction for mean household consumption in Regions 1 and 3 is negative at this expenditure level. Were actual consumption values used, the elasticity would be negative. (The own-price elasticity, $\frac{\partial q}{\partial p}(\frac{p}{q})$, equals the marginal response rate, $\frac{\partial q}{\partial p}$, divided by the average, q/p. The marginal response to the price of rice is negative.)

[3]The reversal of signs that occurs between the whole-sample elasticities for the South (Regions 1 and 3) and those for the North or for the whole sample is likewise the result of the negative consumption value predicted for the South and East by the whole sample regression at this expenditure level. Note that at the low expenditure level, fitting the regression to only the Region 1 and 3 households gives signs opposite to those obtained for these regions from the whole-sample regression.

consumption.[1] The same negative relationship holds for the sample as a whole, but when only Northern households are used to fit the regression the sign is reversed (and the magnitude of the relationship much reduced).

These elasticities, we remember, describe the combined effect of reactions on the supply side and the demand side. A rise in the price of palm oil is a rise in the sale price of palm oil produced as well as a rise in the price paid when the oil is purchased as food from the market. As we have already suggested, the negative cross-elasticity probably reflects a substitution of palm oil for rice that is associated with high levels of domestic palm oil production. The negative elasticity with respect to the price of groundnuts apparently has a similar explanation.

The positive own-price elasticities scattered through Table 11 undoubtedly represent cases in which the supply side response dominates. Consider peppers and chillies in Equation 10 in Table 2, where consumption is positively related to the level of production.

Groundnut own-price elasticities are very high, but the share and quantity regressions differ in sign at low expenditure levels. (The share equations do not contain production and market factor variables.) Cross-price elasticities between groundnut consumption and the prices of other foods are also large, and often negative.

The consumption of alcoholic beverages is also responsive to their prices, with elasticities ranging around -1. No cross-price elasticities appear in Table 12 for this commodity because there were no price variables in the alcoholic beverages regression. For the same reason we have no cross-price elasticities for onions. For palm wine we have no elasticities of any sort, because the palm wine regression contained neither expenditure nor price variables.

Expenditure and price elasticities are important determinants of household food consumption--in the case of several major foods, more important for low-income households than for others. Elasticities often differ markedly among income groups and across regions. The elasticities based on

---

[1] The whole-sample regression yields a large positive coefficient for low-expenditure households in the South and East. This is quite out of line with the remainder of the evidence.

quantity generally apear to be better for prediction than those derived from the share regressions.

Based on the grouped data, expenditure elasticities for rice in the South and East range from +1.01 at low expenditure levels to +.49; in the North they are appreciably lower, ranging from +.75 to +.27. Own-price elasticities in the South and East are quite high at low expenditure levels (-1.51), ranging downward to -.68 at the mean of the high-expenditure group.[1] In the North rice consumption is much more responsive to price than in the South.

Whether we're looking at one or more of the regional groups or the sample as a whole, the prices of palm oil, dried fish, groundnuts and non-food have strong effects upon rice consumption at low expenditure levels. In the South and East, where palm oil production is large, high palm oil prices are associated with reduced rice consumption levels, probably because the increased availability of palm oil from home production is associated with lower consumption of rice.

Even where households produce large quantities of their own foods, market price alternatives play important roles in shaping consumption patterns. Both purchasing and production alternatives matter; the elasticities presented here summarize the net effects of choices made on both the consumption and the production sides of the market.

---

[1]Even the elasticities for the sample as a whole represent low-income households. The mean per capita expenditure in the highest expenditure group is only 136 U.S. dollars per year.

CHAPTER VIII

CONCLUSION

This report concerns low-income rural households producing major portions of their own food. The mean annual expenditure of the 138 households in the sample was $660 (U.S.). At the sample average of 6.55 persons per household this amounts to only $101 per capita per year; the mean per capita expenditure in the highest expenditure group is only $136. The elasticities and predicting equations that we present are specific to these low-income households.

Household size and composition affect consumption choices in measurable ways, but no simple pattern emerges and there appears to be no single satisfactory way to adjust for these factors, whether by using a dependency ratio or by using consumption per capita or per consumer equivalent as the dependent variable. In general, the presence of infants is associated with higher total household consumption levels, and the presence of young children (aged 6-10 years) with lower ones (given the levels of the other variables defining household composition). The presence of infants is also associated with reduced household consumption of kola nut and palm wine.

Nutritionists and others often assert that when households shift from producing their own food to producing for sale, the quality of the diet decreases. The data provide partial support for this proposition for households at a constant level of total expenditure. Production and market orientation variables have no demonstrable effect on the consumption of rice, but households that produce large fractions of their own consumption do consume more palm oil and groundnuts than others (but less cassava and broadbeans). A high degree of market orientation reduces the consumption of cassava,[1] sorghum, and "other cereals" (all cereals except rice). However, palm oil is produced for sale as well as for consumption and the market-oriented production of onions, peppers and chillies is associated with high consumption of these three foods. The share of labor devoted to upland rice, usually grown as a mixed crop, is positively associated with cassava consumption.

These results take no account of the effect of cash crop production

[1]But another measure of production for the market, SHOSS (the share of value output coming from a specified list of activities), is positively associated with cassava consumption.

on income. In an earlier study in which income levels were not held constant [Smith et al., 1980, pp. 57, 60, 61], we found that producing a large portion of the quantity consumed was associated with increased consumption (per consumer equivalent) of cassava, palm oil and groundnuts. But for rice, the most important crop, the evidence was mixed [ibid., p. 46]. In that report market orientation was adversely related to the consumption of cereals other than rice, cassava and alcoholic beverages [ibid., pp. 54, 56, 65].

Economists usually take the position that cash crop production raises incomes and thus leads to better diets. Certainly in rural Sierra Leone there are positive expenditure elasticities for rice, palm oil, fish, vegetables and alcoholic beverages; for rice and palm oil these generally fall as expenditure levels rise. For rice the expenditure elasticities appear to vary also by region. Based on the data for households in the South and East, the elasticities range from +1.01 to +.49 at the three expenditure levels for which they were calculated; the data for Northern households reveal elasticities ranging from +.75 to +.27. Whether these are large enough to justify ignoring the possible adverse effects of cash crop production is another question.

Some would argue that habit and physical environment are the primary determinants of food consumption by households producing mainly for their own use. Certainly food preferences, climate and soil are major determinants, but the data show clearly that rural households in Sierra Leone adapt their consumption practices to the prices they confront. Price elasticities (both own-price and cross-price) are often large--often largest at low expenditure levels. However, the prices that affect these households are both sales prices and the prices paid for food purchases from the market. These single-equation regressions and the elasticities derived from them summarize the total effects of both production and consumption responses, so the signs are not always what one would expect if he were thinking of demand regressions affected only by influences operating on the consumption side of the household's activities.

Rice consumption at low expenditure levels is highly responsive to the prices of palm oil, dried fish, groundnuts and non-food goods, but is little affected by the prices of cassava or of other cereals. The influence of a production response on the elasticity of rice consumption with respect to the price of palm oil is seen in the negative sign of the cross-elasticity coefficient for households in the South and East, where the output of palm products is much larger than in the North. Greater production of palm oil is associated with greater consumption of palm oil and less of rice.

In short, income and price variables play significant roles in influencing food consumption among rural households in Sierra Leone. Their effects must be taken into account in any prediction of the nutritional effects of economic policies.

REFERENCES

Furnival, G.M., and Wilson, R.W., Jr. 1974. "Regression by Leaps and Bounds." Technometrics 16: 499-512.

Gaver, Kenneth M., and Geisel, Martin S. 1974. "Discriminating among Alternative Models: Bayesian and Non-Bayesian Methods." In Frontiers in Econometrics. Edited by Paul Zarembka. New York, N.Y.: Academic Press.

Hamilton, Donald B. 1977. National Training Center Survey Report. Progress Report No. 2, The Interpretation Report. Freetown, Sierra Leone: Government at Sierra Leone/IDA Second Education Project.

Hocking. R.R. 1976. "The Analysis and Selection of Variables in Linear Regression." Biometrics 32: 1-49.

Kolasa, Kathryn M. 1979. "The Nutritional Situation in Sierra Leone." Rural Development Working Paper No. 2. East Lansing, Michigan: Department of Agricultural Economics, Michigan State University.

Mellor, John. 1978. "Food Price Policy and Income Distribution in Low-Income Countries." Economic Development and Cultural Change 27: 1-26.

Smith, Victor E.; Lynch, Sarah; Whelan, William; Strauss, John; and Baker, Doyle. 1979. "Household Food Consumption in Rural Sierra Leone." Rural Development Working Paper No. 7. East Lansing, Michigan: Department of Agricultural Economics, Michigan State University.

Smith, Victor E.; Strauss, John; Schmidt, Peter; and Whelan, William. 1980. "Non-Price Factors Affecting Household Food Consumption in Sierra Leone." Rural Development Working Paper No. 12. East Lansing, Michigan: Department of Agricultural Economics, Michigan State University.

Spencer, Dunstan S.C., and Byerlee, Derek. 1977. "Small Farms in West Africa: A Descriptive Analysis of Employment, Incomes and Productivity in Sierra Leone." African Rural Economy Working Paper No. 19. East Lansing, Michigan: Michigan State University.

Spencer, Dunstan S.C.; Byerlee, Derek; and Franzel, Steven. 1979. "Annual Costs, Returns, and Seasonal Labor Requirements for Selected Farm and Nonfarm Enterprises in Rural Sierra Leone." African Rural Economy Working Paper No. 27. East Lansing, Michigan: Department of Agricultural Economics, Michigan State University.

Timmer, C. Peter. 1978. "The Impact of Price Policy on Protein-Calorie Intake." (Mimeographed.)

MSU RURAL DEVELOPMENT PAPERS

RDP No. 1 Akhter Hameed Khan, "Ten Decades of Rural Development: Lessons from India," 1978.

RDP No. 2 Lane E. Holdcroft, "The Rise and Fall of Community Development in Developing countries, 1950-1965: A Critical Analysis and an Annotated Bibliography," 1978.

RDP No. 3* James E. Kocher and Beverly Fleisher, "A Bibliography on Rural Development in Tanzania," 1979.

RDP No. 4 Enyinna Chuta and Carl Liedholm, "Rural Non-Farm Employment: A Review of the State of the Art," 1979.

RDP No. 5†‡ David W. Norman, "The Farming Systems Approach: Relevancy for the Small Farmer," 1980.

RDP No. 6† Elon H. Gilbert, David W. Norman and Fred E. Winch, "Farming Systems Research: A Critical Appraisal," 1980.

MSU RURAL DEVELOPMENT WORKING PAPERS

RDWP No. 1* Benedict Stavis, "Turning Point in China's Agricultural Policy," 1979.

RDWP No. 2* Kathryn M. Kolasa, "The Nutritional Situation in Sierra Leone," 1979.

RDWP No. 3 Benedict Stavis, "Agricultural Extension for Small Farmers," 1979.

RDWP No. 4* Steve Haggblade, Jacques Defay, and Bob Pitman, "Small Manufacturing and Repair Enterprises in Haiti: Survey Results," 1979.

RDWP No. 5 Peter Riley and Michael T. Weber, "Food and Agricultural Marketing in Developing Countries: An Annotated Bibliography of Doctoral Research in the Social Sciences, 1969-79," 1979.

RDWP No. 6 Harold M. Riley and Michael T. Weber, "Marketing in Developing Countries," 1979.

RDWP No. 7* Victor E. Smith, Sarah Lynch, William Whelan, John Strauss and Doyle Baker, "Household Food Consumption in Rural Sierra Leone," 1979.

RDWP No. 8 Omar Davies, Yacob Fisseha and Claremont Kirton, "The Small-Scale Non-Farm Sector in Jamaica: Initial Survey Results," 1980.

RDWP No. 9 John Hatch, "A Record-Keeping System for Rural Households," 1980.

RDWP No. 10 Sarah G. Lynch, "An Analysis of Interview Frequency and Reference Period in Rural Consumption Expenditure Surveys: A Case Study from Sierra Leone," 1980.

RDWP No. 11 John Holtzman, John Staatz, and Michael T. Weber, "An Analysis of the Livestock Production and Marketing Subsystem in the Northwest Province of Cameroon," 1980.

RDWP No. 12 Victor E. Smith, John Strauss, Peter Schmidt and William Whelan, "Non-price Factors Affecting Household Food Consumption in Sierra Leone," 1980.

RDWP No. 13 Victor E. Smith, John Strauss and Peter Schmidt, "Single-Equation Estimation of Food Consumption Choices in Rural Sierra Leone," 1981.

RDWP No. 14 John Strauss, Victor Smith and Peter Schmidt, "Determinants of Food Consumption in Rural Sierra Leone: Application of the Quadratic Expenditure System to the Consumption-Leisure Component of a Household-Firm Model," 1981.

RDWP No. 15 Middleton Wilson, "Some Problems in Operating a Loan Program for Craft and Emerging Small-Scale Non-Farm Enterprises in Jamaica," 1981.

Single copies of the MSU Rural Development Papers and MSU Rural Developing Working Papers may be obtained free by writing to: MSU Rural Development Program, Department of Agricultural Economics, 206 International Center, Michigan State University, East Lansing, Michigan 48824; U.S.A.

*Out of Print.
†Available in French.
‡Available in Spanish.