INSTITUTE FOR DEVELOPMENT STUDIES

UNIVERSITY COLLEGE, NAIROBI

Discussion Paper No. 29

EXPENDITURE PATTERNS IN THE CENTRAL PROVINCE OF KENYA: A PRELIMINARY ANALYSIS

Benton F. Massell

September 1966

EXPENDITURE PATTERNS IN THE CENTRAL PROVINCE OF KENYA: A PRELIMINARY ANALYSIS

Benton F. Massell

This is part of a larger study being prepared on consumer behavior in Kenya. The present paper contains a preliminary analysis of expenditure patterns in three districts of the Central Province: Kiambu, Fort Hall and Nyeri. An attempt is made to isolate the principal determinants of inter-household variation in expenditure on each of fourteen groups of items. The study is confined to rural households, and is based on the Central Province Survey data, recently collected by the Kenya Government.

The Variables

The literature abounds with discussions of the analysis of budget survey data collected in the United States and Britain. However, little research has been done on expenditure patterns in Africa, and notably the East African countries. The three East African Governments have collected budget survey data. But they have thus far found insufficient time to perform more comprehensive statistical analyses of the data.[1]

One problem that arises in the present study, that does not arise in an analysis of consumer behavior in a developed country, concerns the distinction between cash expenditure and consumption from one's own production. In rural areas of Kenya, a large proportion of agricultural production consists of foodstuffs, and a large proportion of these food items is typically consumed on the farm. A distinction is frequently made between so-called subsistence income and cash income; and correspondingly, between subsistence consumption and cash expenditure. For some items, such as pulses, vegetables, roots and tubers, cash expenditure forms only a small part of total consumption.

---

[1] The Uganda Government are giving some thought to the estimation of income elasticities from data on 1500 Coffee farms in Buganda.

The Central Province Survey, upon which this study is based, includes data on both the subsistence and cash comparents of consumption. For expository convenience, we shall term the value of an item consumed from own production as "subsistence expenditure", and shall regard total expenditure as the sum of subsistence and cash expenditure. This study focuses on total expenditure of 14 groups of items: 8 food items, total food expenditure, 4 nonfood goods, and services. At a later stage, a larger number of groups will be dealt with. Moreover, some attempt will be made to deal with a more disaggregated classification of items.

Expenditure on each of the 14 groups is regarded as forming a set of dependent variables. Variation among households in each dependent variable is explained in terms of interhousehold variation in the independent variables. The principal independent variables dealt with are income and household size.

One is at first tempted to distinguish between subsistence and cash income, and include each separately in the analysis. However, this would yield misleading results. Subsistence income is by definition all consumed on food items. There is a real question whether subsistence income can be viewed as an "independent" variable. It is probably more correct to view subsistence income as jointly determined with subsistence expenditure on each food item. More precisely, it seems unlikely that a household regards subsistence income as given, and chooses to spend this income on different items. More likely, the level of subsistence income merely reflects the household's preference for food items that can most conveniently be grown on the farm. If this is so, then inclusion of subsistence income as an explanatory variable would give rise to biased (and inconsistent) estimates.

Thus, total household income (subsistence plus cash) is our first explanatory variable. One would expect a rise in income to result in increases in consumption of each of the items considered, although some more than others. One of the principal objectives of the study is to obtain a set of estimated income elasticities.

Another important explanatory variable is household size. This is measured as the number of adult equivalents residing in the household during the year. Children are weighted as one-half an adult. And an adjustment is made for absences during part of the year.

Of two households with the same income, one would expect the larger household to consume more of certain items (notably food), but possibly less of other items (such as those in the classed as "luxuries").

Still another explanatory variable relates to the source of income. One might conjecture that wage income would be spent differently from farm income. We have accordingly considered the ratio of nonfarm to total income as an additional variable to include in the model.

The Model

The most frequently used model is the log-log, written

$$E^*_{ij} = a_{1i} + a_{2i}Y^*_j + a_{3i}N^*_j + u_{ij} \qquad (1)$$

where E = expenditure, Y = income, N = household size, u a stochastic term, i denotes the commodity group and j the household, and an asterisk denotes a logarithm. Sometimes the variables E and Y are expressed on a per capita (or per adult equivalent) basis and the N term deleted from (1). Other terms can be added to (1) to introduce additional explanatory variables.

Equation (1) has the advantage of simplicity. The parameters a_2 and a_3 denote respectively the income and household size elasticities of expenditure. The model assumes that these elasticities are constants: i.e., that they do not vary with income or household size. The model also assumes that some positive amount is spent on each item, regardless how low the level of income is.

Problems of Estimation

There are two classes of problems with which an investigator may be confronted in obtaining unbiased (and consistent) estimates of the parameters of the model. The most obvious one relates to the quality of the data. If there are systematic errors in one or more variables, then these will be reflected in the results. As we began with a sample survey that was already collected, we have had only a limited opportunity to exercise quality control. All we can do is accept the data as are, eliminating observations containing obvious inconsistencies, and hope for the best. This problem is present in any collection of data, and there is little reason to expect this sample to be substantially worse than any other sample of households collected in a developing country.

A second class of problems arises even if there are no biases in the data. There are four problems in this class.

a. Simultaneous equation bias. If, as is frequently the case, total household expenditure (instead of income) is used as the independent variable, then there may be inconsistency in the estimates, due to simultaneous equation bias. For example, if a household makes an exceptionally large purchase during the period,say a durable good, then not only consumer durables but total expenditure will be higher as a result. This means that the "independent" variable (total expenditure) is functionally related to the "dependent" variable, and the assumptions underlying the use of single-equation methods of estimation are not satisfied. To get around this difficulty, one can either use another method of estimation or (as is done here) use income rather than total expenditure as the independent variable.

b. Deflation and Spurious Correlation. Sometimes, expenditure and income are both divided through by some measure of household size. This procedure may, introduce spurious correlation, depending on the distribution of the variables. It seems best not to deflate by

household size, and then this difficulty is avoided. Moreover, by treating income and household size separately, the model is less restrictive.

c. Heteroscedasticity. An assumption underlying least squares is that the variance in the dependent variable not be related to any of the independent variables; otherwise, the estimates, while consistent, will be inefficient. Heteroscedasticity is unlikely to be present in the log-log model used here.

d. Specification Bias: This source of bias arises if both expenditure and income are related to another variable, excluded from the model. This arises, in particular, if income is observed with error, or if some component of income is unobserved. It seems unlikely that specification bias can be eliminated in a budget study.[2] One can only hope that it is not too significant.

There were a few instances in which a household's expenditure was zero for one or more of the dependent variables. As the logarithm of zero is not defined, it was necessary to alter the data to enable estimates to be made. For variables in which this problem arose, a small positive constant (1 shilling) was added to each household's expenditure. This procedure is believed not to have an appreciable effect on the results, as the number of zero values was small for each of the 14 commodity groups (less than 3 percent).

Empirical Results

The data consist of households in the Nyeri, Fort Hall, and Kiambu districts. The intention was to divide Kiambu into two parts: a southern part, near Nairobi, and greatly influenced by its proximity to markets and employment opportunities; and a northern part. The northern part has not yet been processed, so that the present results are confined to southern Kiambu.

---

[2] One way to eliminate it is to use the method of "instrumental variables". However, the cure is in this case probably worse than the disease.

Regressions have been fitted to each district individually, with each household being regarded as an observation. The data was not grouped.

Table 1 presents the estimated coefficients of determination (R^2) for each of the 14 expenditure groups in each of the 3 districts, with income (Y) and household size (N) as the explanatory variables. The coefficients of determination vary from .17 to .84. Income and household size explain less than half of the expenditure variance for most items. The fit is highest in all 3 districts for total food expenditure, and relatively high for services, fats, and sugar. It is low in all three districts for pulses and vegetables.

The coefficients of determination are not much improved by adding the nonfarm income ratio (the ratio of nonfarm to total income), Z. The improvement is better for food than nonfood items, as one would expect, and is somewhat better for Kiambu and Fort Hall than for Nyeri, perhaps because nonfarm income plays a smaller role in Nyeri (or because households in Nyeri have less access to markets). For only a few items does the inclusion of Z increase the value of R^2 by at least .05: milk, sugar and pulses in Kiambu; and pulses, roots, and total food in Fort Hall.

Next, in Table 2, we examine the individual contribution of income, household size, and the nonfarm income ratio. A check indicates significance at the 5 percent level, and a double check significance at the 1 percent level, in both cases using a two-tailed test. Income is significant at the 1 percent level for all district-commodity combinations except pulses in Nyeri. Household size is significant in only a few cases, but notably total food in all three districts (1 percent level) and cereals in Fort Hall and Kiambu (also 1 percent level). The nonfarm income ratio is also significant in only a few cases, and fewer in Nyeri than the other two districts. It is significant at the 1 percent level in all three districts for total food expenditure.

Table 1

Proportion of Variance Explained By Income (Y), Household Size (N), and Ratio of Nonfarm to Total Income (Z)

	Nyeri		Fort Hall		Kiambu	
	Y, N	Y, N, Z	Y, N	Y, N, Z	Y, N	Y, N, Z
Cereals	.28	.29	.39	.40	.51	.54
Pulses	.41	.41	.27	.32	.28	.35
Roots and tubers	.55	.58	.31	.36	.26	.30
Sugar	.51	.51	.51	.55	.52	.65
Vegetables and fruit	.23	.26	.36	.38	.30	.30
Milk	.54	.55	.34	.38	.63	.68
Meat, fish, eggs	.39	.39	.63	.64	.67	.67
Total Food	.81	.85	.82	.89	.84	.87
Fats and oils	.56	.56	.43	.43	.71	.71
Tobacco and beverages	.50	.50	.44	.45	.17	.17
Fuel and light	.65	.65	.35	.36	.56	.58
Soap and related products	.38	.38	.48	.50	.57	.57
Clothing	.58	.58	.32	.32	.48	.50
Transport and other services	.53	.54	.47	.48	.64	.65

Table 2

Significance of Regression Coefficients

	Income (Y)			Household Size (N)			Ratio of Nonfarm to Total Income (Z)		
	Nyeri	Ft. Hall	Kiambu	Nyeri	Ft.Hall	Kiambu	Nyeri	Ft.Hall	Kiambu
Cereals	xx	xx	xx		xx	xx			
Pulses		xx	xx	xx				xx	xx
Roots and tubers	xx	xx	xx	xx			x	xx	x
Sugar	xx	xx	xx			xx		xx	
Vegetables and fruit	xx	xx	xx		x				
Milk	xx	xx	xx					xx	xx
Meat, fish, eggs	xx	xx	xx						
Total Food	xx	xx	xx	xx	xx	xx	xx	xx	xx
Fats and oils	xx	xx	xx						
Tobacco and beverages	xx	xx	xx						
Fuel and light	xx	xx	xx						
Soap and related products	xx	xx	xx		xx			x	
Clothing	xx	xx	xx	xx					x
Transport and other services	xx	xx	xx						

Notes: xx significant at .01 level

x significant at .05 level

Table 3 presents the estimated income elasticities, first net of household size and then net of both household size and the nonfarm income ratio. The values differ little between the two cases, as one would expect from the relatively small role played by Z.

One can calculate F ratios to test for the significance of the interdistrict differences in elasticities. Although this has not yet been done, inspection of the estimated elasticities, together with the standard errors, suggests that there might be significant interdistrict differences in the case of pulses, sugar, meat, and tobacco. The low income elasticity of tobacco and beverages in Kiambu may be a result of the fact that the figures include only cash expenditure, whereas in Nyeri and Fort Hall a major part of expenditure is from own production (not recorded in the data). The biggest item in this group probably beer. In Nyeri and Fort Hall, probably only the higher income families purchased beer for cash whereas in Kiambu, which is closer to Nairobi, the lower income households may also have purchased beer. This would make the income elasticity for cash expenditure on beer lower in Kiambu.

With respect to sugar and meat, Fort Hall has a much higher income elasticity than the other two areas. This may reflect the predominantly lower income of households in Fort Hall.

Pulses is more difficult to explain. The elasticity is appreciably lower in Nyeri than the other districts. This may reflect a basic difference in the consumption pattern in Nyeri, possibly because of differences in the ability to grow pulses. As Table 1 shows, income and household size are relatively unimportant in explaining the level of expenditure on pulses in Nyeri, unlike the other two areas.

Table 4 presents averages of the elasticities for the three districts. The averages are less meaningful, of course, for an item where the interdistrict variation is large: for example, tobacco. Nevertheless, they have been calculated for all items.

Table 3

Estimated Income Elasticities

	Nyeri		Fort Hall		Kiambu	
	(1)	(2)	(1)	(2)	(1)	(2)
Cereals	.46	.49	.25	.25	.37	.36
Pulses	.20	.21	.56	.57	.71	.72
Roots and tubers	.30	.32	.56	.60	.32	.31
Sugar	.66	.64	1.06	1.03	.43	.45
Vegetables and fruit	.56	.58	.71	.69	.48	.46
Milk	1.14	1.15	1.04	1.05	1.25	1.25
Meat, fish. eggs	.81	.80	1.35	1.27	.90	.82
Total Food	.56	.55	.68	.71	.63	.63
Fats and oils	.76	.75	1.00	.96	.86	.84
Tobacco and beverages	.91	.91	1.00	.98	.40	.47
Fuel and light	.82	.81	.80	.74	.85	.84
Soap and related products	.63	.62	.59	.56	.67	.66
Clothing	.78	.78	.91	.87	.97	.94
Transport and other services	1.27	1.26	1.44	1.39	1.28	1.26

(1) net of household size

(2) net of household size and nonfarmto-total income.

Table 4

Average Estimated Income Elasticities

	Elasticity	Ranking	Adjusted Elasticity	Adjusted Ranking
Cereals	.36	14		14
Pulses	.43	12	.64	8
Roots and tubers	.39	13		13
Sugar	.72	8	.54	12
Vegetables and fruits	.58	11		11
Milk	1.14	2		2
Meat, fish, eggs	1.02	3	.86	6
Total Food	.62	10		10
Fats and oils	.87	5		5
Tobacco and beverages	.77	7	.96	3
Fuel and light	.82	6		7
Soap and related products	.63	9		9
Clothing	.89	4		4
Transport and other services	1.33	1		1

The ranking of items is roughly what one would expect, with service at the top and the staple foods at the bottom. Milk and meat are highly income elastic, and clothing, fats, and fuel are relatively highly elastic.

The second column of Table 4 presents the averages of a few items with the extreme value in each case excluded: Fort Hall for meat and sugar: Nyeri for pulses; and Kiambu for tobacco. Here, meat drops from 3rd to 6th; tobacco rises from 7th to 3rd: sugar drops from 8th to 12th; and pulses rises from 12th to 8th. The positions of first two and last two commodity groups remain unchanged.

Table 5 presents the estimated elasticities with respect to household size. (N) as well as the coefficients associated with the nonfarm income ratio (Z). As the standard errors are in both cases substantially higher than those associated with the income elasticities, there is considerably more interdistrict variation in the figures. With respect to household size, these interdistrict differences appear to be significant only in the case of pulses. Here again, there appears to be a genuine difference in the role of pulses in the diet in Nyeri, compared with the other two districts. Just as the income elasticity in Nyeri was substantially lower, so is the household size elasticity significantly higher. Pulses are apparently regarded in Nyeri as a relatively inferior good, to be consumed by families with a low income per capita.

The interdistrict variation in the Z coefficients appears to be significant for 5 of the 14 expenditure groups: pulses, sugar, vegetables, soaps and clothing. As the standard errors are large (most coefficients are not significantly different from zero at the .05 level), the actual values of the coefficients do not have too much meaning. However, the Fort Hall coefficients appear on balance to be (algebraically) the largest and Nyeri the smallest. The coefficients are negative for all food items and positive for most nonfood items

Table 5

Household Size Elasticities and

Nonfarm Income Coefficients

	Household Size			Nonfarm Income Ratio		
	Nyeri	Ft.Hall	Kiambu	Nyeri	Ft. Hall	Kiambu
Cereals	.41	.54	.33	-.11	- .25	- .33
Pulses	1.03	.13	-.28	-.47	-1.05	-1.32
Roots and tubers	.49	.26	.18	-.62	-1.19	- .51
Sugar	.09	.05	.31	.02	-1.22	- .09
Vegetables and fruits	.14	.36	.16	-.90	- .72	- .17
Milk	.06	.28	.28	-.86	-1.64	-1.49
Meat, fish, eggs	-.10	-.09	.17	-.09	- .20	- .06
Total Food	.26	.16	.20	-.59	- .92	- .48
Fats and oils	-.04	.17	.16	.20	- .23	- .23
Tobacco and beverages	-.07	.13	.04	.20	- .66	.03
Fuel and light	-.02	.26	-.16	-.09	- .77	.47
Soap and related products	.06	.47	.07	.05	- .63	.01
Clothing	.57	.42	.26	.19	.42	.85
Transport and other services	.12	-.29	.17	.39	.07	.62

(except in Fort Hall). Thus, if one compares two households with a given income, the one with the larger proportion of this income derived from the farm will tend to spend more on food and less on the other items. This is undoubtedly simply a reflection of the underlying association between farm income and subsistence income.

It is interesting to note that in Fort Hall the Z coefficients are negative for all items except services and clothing, and are low for both of the latter. This suggests that much of nonfarm income in Fort Hall is spent on items not considered here (such as durables), or is saved.

Table 6 shows average household size elasticities. It is of some interest to compare an item's ranking in this Table with its ranking in Table 4, for income elasticities. An item whose ranking in the income elasticities exceeds its ranking in household size elasticities can be regarded as a "luxury." An increase in household size substitutes for expenditure on these items. Items that can be grouped in the luxury class are (in order of the difference in the rankings): services, meat, fats, fuel, tobacco, and milk.

At the other end of the scale are items whose ranking among the household size elasticities is higher than its ranking among income elasticities. Such goods are "necessities." An increase in household requires an increase in expenditure of these items. This group of items includes: cereals, roots and tubers, pulses, vegetables, and total food.

Clothing has a unique position, ranking 4th in income elasticities and 2nd for household size elasticities. Clothing is clearly a luxury in one sense, or nearly so. But larger families also require more clothing: so that it increases rapidly with an increase in either income or household size.

Table 6

Average Household Size Elasticities

	Household Size Elasticities	Ranking
Cereals	.43	1
Pulses	.29	4
Roots and tubers	.31	3
Sugar	.15	9
Vegetables and fruits	.22	5
Milk	.21	6
Meat, fish, eggs	-.01	14
Total Food	.21	6
Fats and oils	.10	10
Tobacco and beverages	.03	11
Fuel and light	.03	11
Soap and related products	.20	8
Clothing	.42	2
Transport and other services	00	13