RURAL URBAN STUDIES UNIT

UNIVERSITY OF NATAL DURBAN

THE HOUSEHOLD IN RURAL KWAZULU

FOUNDATIONS FOR A STATISTICAL MODEL

ALAN PETERS

CENTRE FOR SOCIAL AND DEVELOPMENT STODIES

WORKING PAPER NO. 5

THE HOUSEHOLD IN RURAL KWAZULU:

FOUNDATIONS FOR A STATISTICAL MODEL

by

ALAN PETERS

Rural Urban Studies Unit

University of Natal

Durban

1987

Rural Urban Studies Unit

The Rural Urban Studies Unit was founded in 1983 by the Human Sciences Research Council for the purpose of studying the dynamics of the links between the rural and urban areas of South Africa. It is situated at the University of Natal, Durban and works in close co-operation with the Development Studies Unit.

ISBN 0-86980-556-8

The Household in Rural KwaZulu

Foundations for a Statistical Model*

A considerable amount of academic attention has been focussed on South Africa's bantustans. Various results have ensued. Competing macro-level theories of the role and function of the bantustans within South Africas broader political economy have been developed, while micro-level anthropological and statistical studies have described the demographic, agricultural, and economic situation within the bantustans. Nevertheless, the macro-level work has tended to be deductive - usually seeing the socio-economic structure of households in the bantustans as a function of the migrant labour system - and the micro-level research has been too localized. Thus there is a need for investigations which are both rich in data and generalizable. This paper utilized an extensive KwaZulu data-base gathered by the Development Studies Unit at the University of Natal from the latter part of 1983, and still being updated. In the process of analysing the data, foundations for a general statistical model of the household in rural KwaZulu are laid.

---

* Financial assistance by the Rural Urban Studies Unit, established by the HSRC is hereby acknowledged. This paper was written during a short break from Rutgers University, where the author is partially funded by an H.S.R.C. Doctoral Merit Bursary. Views expressed and conclusions drawn are those of the author and should not be regarded as necessarily reflecting those of the H.S.R.C. The author would like to thank Jill Nattrass for her support and advice, Julian May with whom the original research was conducted, and Dave Perkins and Barbara Duffy for their help.

1. Methodological Background to the Research

Liberals have traditionally viewed the decline in the economic viability of the bantustans as a function both of the destructive consequences of the migrant labour system and of the failure of Africans to respond appropriately to their changed circumstances (Houghton 1973). The underdevelopment of African areas has thus been seen as closely connected to those specific features intrinsic to the African social structure which have inhibited innovation. Poverty and lack of development are thus represented as aspects of an original state for which the remedy is economic incorporation into the progressive capitalist sector. Such incorporation would lead to the decline of traditional values and the decline of the oscillating migrant labour.

In reaction to the above sort of account, structuralist Marxists argued that impoverishment in the bantustans was a direct result of capital accumulation in white South Africa (Wolpe 1972, Legassick 1974). They argued that the non-capitalists mode of production in the "reserves" was preserved as a direct result of the dominant capitalist sectors' need for a constant source of cheap labour. Migrant labour was cheap because the workers' families remained behind in the rural areas freeing mineowners and white farmers from the expense of paying a wage on which the whole family could subsist. However, since the early 1950's there has been a massive decline of per capita agricultural output of the bantustans (Simkins 1981). While there has been some debate among structuralists over the distinctive status of apartheid, their response to the fact of subsistence sector decline

has been to argue that apartheid involves brute repression as a means of securing cheap labour from the bantustans.

A crucial point needs emphasizing. The early dualist liberals saw household structure (its demography and economy) anthropologically, that is, as a function of "tradition", while the structuralists did just the reverse. They deduced the structure of the rural household from South Africa's broader political economy. Recent work has emphasized the need to focus on the rural household itself.

Crush (1984), amongst others, has argued that the structuralist account of migrant labour has ignored the forces within African society promoting participation in and resistence to wage labour. Along similar lines, Beinart (1982) claims that the migrant labour system and African reserves were not created merely to service the needs of capital. For instance, Pondo elders encouraged migrancy in order to maintain both their rural authority and their command over the young migrant's earnings. Lewis' (1984) study of the Mfengu of the eastern Cape indicates that the incorporation of households into the migrant labour system varied considerably according to the radically skewed access to productive resources, including family labour, that existed between household. Lewis argues that an understanding of the causes and consequences of migrant labour requires that far greater attention be paid to the relationships between households in the rural areas of the reserves (Lewis 1984:22). An essentially similar point is made by Murray (1981) and May (1985b:32) who conclude that theories which focus entirely on the

macro forces of the South African economy have little to offer research into the current perpetuation of migrant labour or bantustan immiseration. This criticism is directed not only at structuralists but all theorists who see a relatively simple and determinate relationship between the broader South Africa economy and those of the bantustans, especially those explanations of migration based upon interlocking "push and pull" factors (see Todaro 1976, Wilson, 1972).

A range of more sophisticated theories of black rural underdevelopment have been posited. For instance, Nattrass (1983, 1986) presents a three sector model (the modern sector, the commercial farming sector and the subsistence sector). The model is not purely economic : it also takes cognizance of the effect of government policies and politics on the present plight of rural blacks. For instance, Nattrass traces the consequences of the Land Act, population removals into the bantustans, and restrictions on African urbanization, on poverty in the bantustans. Nattrass' political-economic approach to rural underdevelopment has been elaborated by Brembridge (1986) and others, who use it to investigate such phenomena as environmental destruction, low productivity, poor nutrition and so on.

Other multi-sector models have been proposed. Hughes (1985), for example, offers a formalized four sector model of the rural African economy summerized in the equation:

$$N_R = N_M + N_T + N_I + N_S + N_L \quad (1)$$

Where N_R is the allocation of the rural population, N_M migrants to the

core economy, N_T Todaro unemployed, N_I intermediate sector activity, N_S agricultural sector activity, and N_L those economically inactive. Although the model has the advantage of formality it is excessively simplistic. It revolves around a push-pull theory of migrancy (the migrant is evidently a utility maximizer) and ignores the political and legal constraints placed on black urbanization and migration (those people who would migrate but are institutionally constrained from doing so are said to be part of N_T, the unemployed in Todaro's sense). Technically, Hughes' model is incorrect since it neglects wage employment within the bantustans and fails to differentiate 'frontier commuter' workers. Furthermore, the equation equalities upon which equation (1) is based imply, against evidence to the contrary (Wellings and Sutcliff 1984), substantial growth in informal sector employment (part of N_I) in the bantustans.

An alternative approach has been to study household and community structure in a particular regions. This has been done anthropologically - for instance Derman and Poultney's (1983) work on smallhold farming in the Makathini flats - or with greater reliance on the sociological survey technique. Ardington's (1984) study of the Nkandla district is one of the best example of the latter, and provides an intensive investigation of relationships between household demography (household size, education levels, age of household members), economy (migrancy, local work), agricultural activities and local political structures. Of course, a number of income and expenditure studies and agricultural surveys have been done in different areas of KwaZulu. However, there is a crucial problem with

all such work if generalizations are to be made for the whole of KwaZulu. Regions within KwaZulu differ markedly, with respect to average total household income, sources of income, agricultural involvement and demographic characteristics (May and Nattrass 1986).

In order to make some substantive generalizations on household structure in rural KwaZulu, a compromise must be reached between the formalized but substantively uninteresting model of Hughes, the rich but localized studies of Derman and Poultney, Ardington and others, and the theoretically deductive accounts of the structuralists and liberal dualists. Some authors have attempted to overcome this problem by comparing their findings to those from different regions (for instance Gandar and Bromberger 1984). Four papers (Nattrass, May and Peters 1985; May, Nattrass and Peters 1985; May and Nattrass 1986; Nattrass 1986) combine data from three different areas in KwaZulu into a single database, thus providing reasonably generalizable information on the rural household. However, their results are only marginally formalized or not formalized at all. This paper relies on the same database. Although a formal statistical model of the rural household in KwaZulu is not derived, various statistical analyses are performed, which, beside providing a set of empirical generalizations, suggest the appropriate statistical structure of a path analysis or similar model of the rural household in Kwazulu. Unfortunately the results presented here ignore political-economic considerations.

In this paper the relationships between household income, wage income and remittances, agricultural income, imputed subsistence income, other income, migrancy, wage employment, education, spatial isolation,

household size and dependency are investigated : the specific choice of statistical relationships studied was determined by the limitations of the database and the results of recent empirical research, in particular that of Ardington (1984), Gandar and Bromberger (1984), Derman and Poultney (1983), May (1985) and Nattrass (1986).

1.1 The Research Design

The data used in this paper are part of a much larger data-base on the economy and demography of households throughout KwaZulu. Unfortunately, the biggest obstacle to the immediate development of a statistical model of the rural KwaZulu household is the data-base which presently lacks detail in some crucial areas. Nevertheless in this paper, the foundations for such a model are laid through a rigorous analysis of empirical data. Further data extensions and refinements required by such a model are, in general terms, specified.

Figure 1 provides a diagram of the research design. Factor analysis was employed to clarify and summarize the original set of variables. The resultant factor loadings suggested certain patterns within the data, for instance, that there was a relationship between male migrancy and the household educational index. The patterns within and between the factors were investigated using Analysis of Variance and Chi Square tests. The relative importance of the variables in the determination of household income was analysed by means of regressions. Although spatial isolation indices did not feature in the factor analysis, analysis of these using the Kruskal-Wallis H

statistic was undertaken here in an attempt to measure intra-regional differentiation.

Figure 1

Basic Research Design

Step	Statistical Test	Performed On	Section
1.	1st-Order Factor Analysis	Variables	3
2.	ANOVA	Variables	4
	Chi Square	Variables	4
	Regression Analysis	Variables	5
	Kruskal-Wallis H	Variables	6
3.	2nd-Order Factor Analysis	Oblique Correlation Matrix	7
4.	t-Tests	1st Order Factor Scores	8

A broader scope was achieved by using a 2nd-order factor analysis. Three higher-order factors were generated. The relationship between the factors was investigated using t-tests on the 1st-order factor scores.

Overall, the two factor analyses reduced the mass of economic and demographic data on the rural households sampled to three basic 'patterning types'. The latter, while requiring further in-depth survey data for their formalization, suggest particular sets of relationships on which a statistical model of the rural KwaZulu household could be based.

Note that the factor analyses undertaken here are purely descriptive

of the sample data : inferences to the population are made using the ANOVA, Kruskal-Wallis, t and X^2 tests.

1.2 Objections to Statistical Models

Over and above the issue of costs, two general objections are often raised against statistical models. (1) Models cannot provide a causal account of the relationships under investigation. Thus the models have little substantive explanatory power. (2) They are ahistorical, with the result that they fail to take proper account of the social structures which permit and limit different types of social action. More generally, statistical models, even those which use time-series data, are likely to become quickly outdated since they are in essence no more than descriptions of correlations found at particular temporal intervals. (See Sayer 1984).

These criticisms are part of a much larger and well-rehearsed debate concerning positivist methodology in the social sciences, a debate which will not be restaged here (see, for instance, Adey and Frisby 1976). Nevertheless, some limited and altogether very practical answers can be given to allegations (1) and (2). (i) Provided the model sits within a wider framework of an acceptable theory, certain correlations and other statistical relationship may be adduced as evidence for postulated causal mechanisms. (ii) Although statistical models are indeed ahistorical (at least in the sense understood on the 'left'), time series analysis, sophisticated design of data-bases, data updates in areas most obviously undergoing change, combined with

some theoretical insight into the societal developments taking place, provide reasonable grounds on which to extrapolate some statistical models into the near future. For arguments sake, presume that a coherent model had been developed for the rural household in KwaZulu. Whether in a post-apartheid South Africa or not, certain empirical estimations will still have to be made. For instance, how would rural welfare be improved if migrant (presuming that migrant labour persists in some form for a time) wages were increased? Or, how would rural poverty be affected by an increase or decrease in old-age pensions? Or, to what extent would state investment in black subsistence agriculture improve output (presuming that subsistence agricultural is retained)? Used correctly, models can be useful, though limited, guides both to policy making and conceptual understanding.

2. The Sample

Some six hundred households (including data on 4600 individuals) were surveyed in the Mapumulo and Nqutu magisterial districts, and an area in the Inkanyezi magisterial district near the Mbongolwane Mission. These regions were divided into quadrats which differed in size roughly in accordance with population density. Systematic random samples were then taken in each quadrat. In Nqutu this procedure was modified somewhat. Representative clusters were taken in each quadrat and systematic random samples were then taken of the clusters. The latter method was necessitated by the presence of a number of large betterment and resettlement villages in the Nqutu region (Surplus Peoples Project 1983:vol.4).

There are important differences between the areas. Mbongolwane is the most isolated and consequently the furtherest from wage labour opportunities and the least integrated into the broader South African economy. On the other hand, Nqutu is relatively close to Vryheid and Dundee, and has a tarred road passing through the centre of the district connecting it to these two centres. Indeed, Nqutu acts in part as a 'commuter suburb' to Vryheid and Dundee with a large number of workers travelling back and forth on a daily basis. The Mapumulo district occupies an intermediate position. It is within a feasible commuting distance to Stanger, and a number of its workers travel to Durban for the working week and return home at weekends. With respect to cultivation, maize, vegetables and livestock are the only form of subsistence activity in Nqutu, whereas sugarcane and limited market-gardening occur in the other two area. Finally population density was highest in Nqutu (reflecting the population removals into the area) and lowest in Mbongolwane.

Insofar as the three regions are themselves representative of the social and economic diversity in rural KwaZulu, the sample may be considered representative of rural KwaZulu as a whole. There was however, no weighting of the sample in favour of a particular area as approximately two hundred surveys were conducted in all three districts. Thus if the sample is to be considered representative of rural KwaZulu, it must be additionally assumed that regional diversity in KwaZulu mirrors the proportions of the sample.

2.1 Demographic Profile

An extensive demographic analysis of each of the areas is given in Peters and May (1984), May and Peters (1984a) and May and Peters (1984b) respectively. Demographic comparisons between the regions are discussed in May and Nattrass (1986). This information will not be repeated here. Nonetheless, some basic demographic findings should be mentioned.

(1) The mean household size including migrants was 7.6 people. Thirty-five percent of the total or 47 percent of the non-migrant population was under 15 years of age. Absentees made up 22 percent of the total population.

(2) Illiteracy or functional illiteracy ranged between 67 percent of people in Mbongolwane, 47 percent in Mapumolo and 37 percent in Nqutu. Education profiles did not vary by sex.

(3) Well over four fifths of households had access to land for crops (over and above gardens) in Mapumulo and Mbongolwane, but only a third had similar access in Nqutu. Most landholders had a plot of less than a hectare. Land was unequally distributed. Those with less than one hectare (over two-thirds of landholders) had access to a mere quarter of all allocated land. Most of those who had land made some use of it, despite the drought of 1983. The drought probably resulted in a slight underestimation of farm income for Mapumulo.

3. The Initial Factor Analysis

3.1 Introduction

Attempts were made to model the rural household in KwaZulu using various regression techniques. These are discussed in Section 5.3. In general they failed to provide any new, interesting and rigorous insights into the economic and demographic structures of the rural household in KwaZulu. This indicated the need for more exploratory techniques. Factor analysis seemed the most appropriate, since this technique both provides a relatively parsimonious summary of a large number of differing variables and gives some idea of the underlying structure or pattern of relationships which may exist in the data.*

A number of factor analyses were tried in which variables included, the factor extraction techniques used, the number of factors in the final solution, and the type of rotation, were varied. The solution finally adopted offered the best compromise between data summary, interpretive simplicity and the formal criteria which the technique should satisfy (Rummel 1970:380).

3.2 Data

Separate anlyses were run on each of the individual areas, but with little success. Data from the three regions (Mapumulo, Mbongolwane

---

* For an introduction to and full discussion of factor techniques see Rummel (1970). Everitt and Dunn (1983:ch.11) is brief but good. Also see Afifi and Clarke (1984). For the algorithms used in this analysis see SPSS Reports (1985:68-84).

and Nqutu) were then combined. The data were left untransformed; missing data on any variable included in the analysis resulted in the exclusion of that case. A two-stage procedure was used to select variables to be included in the analysis. In the first instance, those variable which appeared to be of some theoretical interest and which were considered reasonably reliable were chosen. In the second instance, Pearson's product-moment coefficients were calculated for all pairs of chosen variables - those variables which were almost uncorrelated with most of the others were excluded from the analysis. In practice, a variable which had an 'r' of more than .1 with three or more other variables, was included. The variables finally included are listed in the factor loading matrices given below. The meaning and measurement of the variables are briefly discussed in Appendix A. The variables finally included were then transformed into a product-moment correlation matrix.

3.3 The Analysis

The initial statistics indicated that six of the factors had eigenvalues greater than one and therefore should be retained. Scree plots (see Appendix B) confirmed the choice of six factors. However, in the six factor solution four percent of the residuals between the reproduced correlation matrix and the original correlation matrix had a value greater than .05. Nevertheless, restricting or extending the member of factors allowed, decreased the interpretability of the factor loadings.

Common factor analysis with principal axis factoring (PAF in the

language of SPSS^{X}) were the technique and model used. Estimates of the final communality values are given in Table 1 below.

Table 1

Final Estimates of Communality with Six Factors

Variables	Communality	Variable	Communality
AMTSUGA	.33	ITEM90	.95
ECONINAC	.50	MLAWAY	.46
EDUC	.38	MLHEAD	.17
FARMINCA	.50	MLHOME	.45
FARMEXP	.31	NMIGRTS	.91
FLHEAD	.06	REG	.37
FLHOME	.14	SIZEHH	.93
		WEEKINCD	.74

In factor analysis, the variance of observed variables can be split into two parts. The first is the communality of the variable and indicates the amount of variance shared by at least one of the other variables in the set through the common factors. The second is the unique variance and relates to the variability in x_i not shared with other variable. In the six factor solution of this analysis, wages (ITEM90), household income (WEEKINCD), cash-crop income besides maize (AMTSUGA), and the number of non-migrants in the household (NMIGRTS), each had a high shared variance. However, the number of working females resident at home per household (FLHOME) and female headed households (FHEAD), each shared an insignificant amount of variance with the other variables. They are therefore unlikely to figure prominently in the analysis - indeed, their continued existence in the factor solution is unfortunate but was convenient given the original design of the study.

3.4 The Factor Loadings

The six factors only accounted for 48 percent of the total variance in the data. The unrotated first factor accounted for 16 percent and the second, 13 percent. The unrotated factor loading matrix is given in Appendix C. These loadings will not be discussed and are included for comparative purposes only. Various rotations of the factors were tried, both orthogonal and oblique. Finally a varimax rotation was adopted although the oblique rotation provided marginally superior factor loadings. The varimax rotation was preferred, at least at this

Table 2

Rotated Factor Loadings

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6
NMIGRTS	**.914**	.001	.020	.261	-.038	.047
SIZEHH	**.882**	.036	.026	.143	.356	.029
ECONOMAC	**.687**	-.020	-.097	-.114	-.090	-.050
FLHEAD	**.186**	-.055	-.006	.074	.103	-.065
ITEM90	-.084	**.963**	.058	.044	.070	-.035
WEEKINCD	.011	**.768**	.316	.084	.193	.050
FARMINCA	-.045	.088	**.703**	-.009	.013	-.010
AMTSUGA	-.013	.105	**.521**	.056	-.005	.206
MLHOME	-.048	.013	-.030	**.653**	-.112	-.106
MLHEAD	.113	.023	-.002	**.395**	.057	-.028
FLHOME	.135	.059	.135	**.291**	.106	.066
MLAWAY	.201	.140	.061	-.217	**.541**	-.006
EDUC	-.027	.080	-.036	.261	**.538**	.139
FARMEX	-.034	.133	.085	-.031	.040	**.530**
REG	.019	.299	-.089	.055	-.061	**-.512**

particular stage of the analysis, because of the added complexity introduced by the need to interpret correlations produced by the oblique rotation.

Table 2 above presents the varimax rotated factor loading matrix. Here the coefficients may be interpreted in two ways: (1) as weights to estimate variables from factors, or (2) as simple correlations between factors and variables. Factor 1 loads very heavily and positively on the number of non-migrants in the household, the size of the household and the number of people in the household who were economically inactive. FLHEAD merely adds noise to the analysis. The patterning found in the first factor can be supported by a simple review of the data. Larger households tended to have more non-migrants and more people who were economically inactive. The first factor will be called the "Size of Household Factor".

The second factor will be labelled the "Income of Household Factor". The factor loads strongly and positively on wages and weekly household income. The factor is however, problematic. Variables included in the factor analysis should be arithmetically independent of each other (Rummel 1970:213) since arithmetically related phenomenon may produce factors which are functions of arithmetical operations on the data and not functions of the empirical structure of the data itself. Total household income was calculated by the addition of household income from all sources and in all three areas wages formed the major source of income. Nevertheless the inclusion of both WEEKINCD and ITEM90 may at least be partly justified on the grounds that these two variable

provide (or should provide) information on important different aspects of the rural household, and thus are not mutually reducible.

Factor 3's heaviest loadings are on cash farm income and on cash income from agricultural products beside maize. The latter would, in both Mapumulo and Mbongolwane, be made up almost entirely of sugarcane production.* This factor will be called the "Cash Crop Factor".

Loadings on the final three factors are not as strong as on the first three. Furthermore, the factors explain considerably less of the variance in the variables than the first three factors, so they will be discussed more briefly. Factor 4 loads most significantly on male heads of household and the number of working males at home. It will be labelled the "Working Males at Home Factor". It should be noted that this factor has smaller but still significant loadings on the number of non-migrants in the household, on the number of working male migrants, and on the educational index of households.

Factor 5, the "Well Educated Households with Male Migrants Factor" loads almost equally and positively on working males who are not resident at home, and on the educational index of households. This factor also loads on the size of the household variable. The final factor, labelled the "Region and Agricultural Expenditure Factor" allows for no obvious interpretation. Note that this factor loads

---

* In both these regions there were reasonably successful sugarcane farmers. For a description of and debate over sugar farming in Mapumulo, see Cobbett (1984), Wiseman (1985) and Cobbett (1985).

positively on farming investment but **negatively** on region.* This appears to imply that levels of agricultural investment do not differ dramatically by region.

3.5 Relationships In Need of Investigation

Factor analysis is exploratory. In this particular case, it has suggested a certain set of economic and demographic patterns which might be used to model the rural household in KwaZulu. Moreover, the factor techniques used in this analysis do not provide a measure of the statistical significance of the patterns found. Thus the patterns must be tested against whatever other statistical evidence can be mustered.

The endogenous relationships between the variables on which a factor loads strongly need to be investigated. For instance, the connexion between the size of the household and the number of people not working in that household needs attention. Moreover the relationships between the factors, and between the variables that compose the

---

* The listing of variables in Appendix A indicates that REG was treated as an ordinally-scaled variable measuring relative distance from major metropolitan centres. In later analyses, REG was both excluded from the factor analysis, and included as two purely regional dummy variables. Both these strategies failed to alter the general factor loading structure in any theoretically important way. In the former case, five factors were extracted with FARMEXP joining FARMINCA and AMTSUGA in a single factor. This particular simplification was also achieved in the higher-order factor analysis (see Section 7.3) and therefore has no important implications which are not considered later in this paper. In the case where the dummy variables were included, six factors were extracted. Here the third factor loaded strongly on FARMEXP, AMTSUGA and FARMEXP. A separate factor loaded on the two dummy variable. The rest of the factor loadings matrix remained essentially similar.

different factors must be clarified. For example, to what extent does household size determine household income? Thus attention must be paid to three distinct types of relationship.

a) Endogenous relationship between the variables on which particular factors load.

b) Exogenous relationships between the variable composing different factors.

c) The relationships between the factors themselves.

Answers to questions raised by 'a' and 'b' above are provided by means of Chi-Square, Analysis of Variance, and regression techniques. A second order factor analysis is then used to clarify the connexion between the factors.

3.6 The Factor Scores and Their Use

Factor scores for each case were calculated using the regression method. The factor score co-efficient matrix (see Appendix D) provided the necessary weights to estimate the factor scores from the variables, much the same as 'Y' in the standard multiple regression equation may be estimated from the addition of the product of each of the respective 'B' weights and each of the respective variables. Thus all cases gained six new variables, FSULS1, FSULS2, FSULS3, FSULS4, FSULS5 and FSULS6 corresponding to the six factors. For example, case number 897 (chosen at random) in the Mbongolwane region had a factor

score of -.69 on the Size of Household Factor (FSULS1), - 3.8 on the Income of Household Factor (FSULS2), - .47 on the Cash Crop Factor (FSULS3), - .64 on the Working Males at Home Factor (FSULS4), - .02 on the Well Educated Households with Male Migrants Factor (FSULS5), and 1.88 on the Region and Agricultural Expenditure Factor (FSULS6). Thus the household recorded as case number 897 was probably smaller than the average, was not heavily involved in commercial farming and probably had neither a male head of household nor working males resident at home. If the household did have a male migrant, it had a low educational index score. The household had a higher than average income.* Obviously a case by case comparison is infeasible given the size of the sample. Nevertheless, the factor scores proved highly valuable in tests discussed in Section 8.

4. The Economic Structure of the Rural Household

The data were disaggregated by region. Analysis of Variance and Chi-Square tests were run in order to give some idea of the exogenous and endogenous relationships between the factors. The results are presented in Figures 2,3 and 4 below. As far as possible groupings were determined by "natural" breaks in the distribution of variables. Household income was normalized for the Analysis of Variance tests.

---

* This interpretation is borne out by the original variables. Household 897 had a weekly income of R266.38. It was smaller than average with four non-migrant members (six members including migrants). It was headed by a male but there were no males residents at home. Although the household had two working male migrants, the household had an extremely poor score (2) on the educational index. Farm income was minimal.

4.1 Agriculture

In Mapumulo the importance of non-stock farming is clearly evident. Those investing something in agriculture had significantly better household incomes. Likewise, those households involved in some maize or other crop production for the market, also had significantly higher household incomes. A further point should be noted in this regard. Livestock units possessed by the household did not increase with income. In conclusion, cash income from non-stock farming appears to bear an important relationship to household income.

The results of the Chi-Square tests complicated this picture somewhat. There was some evidence that households with any working males, be they migrants or residents, do better at non-maize cash crop production. However, the situation was ambiguous: at the more rigorous significance level of $p = 0.01$ none of the Chi-Square tests considered above were statistically significant. There was also an ambiguous relationship between the number of people economically inactive and non-maize cash crop production. This suggests, but does not confirm (because of the small sample of people actually engaged in cash crop production) that Mapumulo can be characterised by the following general model. Farming requires the injection of money from members of the household who are engaged in wage labour. Farm labour is provided by those members of the household who are 'economically inactive'. Further research focussing entirely on cash crop growers in Mapumulo would be needed to verify this model.

Figure 2
Mapumulo: Analysis of Relationships Suggested by Factor Solution

Dependent Variables / Independent Variables	TEST TYPE: ANOVA — Household Income	ANOVA — Assessment Of Assumptions	X^2 — No. of Male Working Migrants	X^2 — No. of Working Males At Home	X^2 — No. of Economically Inactive	X^2 — No. of Children at School	X^2 — Educational Index of Household
Expenditure of farming	S	A	T	T			
No. of children at school	T	W	T	T	T		
Educational index of households	S[1]	A	T	T	T	T	
Maize cash crop production	S	A	T	S	T		
Other cash crop production	S	A	S[2]	S[3]	S[4]		
Total Stock Units	T	A	T	T	T		
Cash income from farming	S	D	T	S			
Size of household	T	W	S	T	S		
Occupational index of households	S	A	S	T		T	S
No. of working male migrants	T	A					
No. of working males at home	S	A					
Male head of household	T	A			T		
No. of economically inactive	T	A	T	T			

n (ANOVA) = ±190
n (X^2) = ±110

Legend : 'T' Denotes that a test was conducted
'S' Denotes that respective 'F' and 't' results were significant at p = 0.05 unless otherwise stated.
'W' Denotes that the assumptions of analysis of variance were well met.
'A' Denotes that they were approximately met.
'D' Denotes deviation from the assumptions.

Notes : 1. The educational index was divided into three groups. The Scheffe procedure indicated that the extreme groups were significantly different, and the middle group was significantly different to the highest ranking group. Contrasts indicated the lowest group was significantly different to the top two groups combined.
2. at p = 0.07
3. Only significant before the Yeates Correction.
4. 50% of cells had an expected frequency less than 5.

Figure 3
Mbongolwane : Analysis of Relationships Suggested by Factor Solution

Dependent Variables / Independent Variables	ANOVA: Household Income	Assessment Of Assumptions	χ^2 TEST TYPE: No. of Male Working Migrants	χ^2 TEST TYPE: No. of Working Males At Home	χ^2 TEST TYPE: No. of Economically Inactive	χ^2 TEST TYPE: No. of Children at School	χ^2 TEST TYPE: Educational Index of Household
Expenditure of farming	S	W	T	T			
No. of children at school	T[1]	W	T	T	T		
Household Educational Index	T	A	T	T	T	S[7]	
Maize cash crop production	S[2]	W	T	T	T		
Other cash crop production	S[3]	A/D	T	T	T		
Total Stock Units	S[4]	W	T	T	S		
Cash income from farming	S	A/D	T	T			
Size of household	T	W	S[5]	T	S		
Occupational index of households	S	A	S[6]	T		T	S[8]
No. of working male migrants	T	D					
No. of working males at home	T	A					
Male head of household	T	W			T		
No. of economically inactive	T	W	T	T			

n (ANOVA)= ±200
n (χ^2)=±110

Legend : as in figure 2

Notes : 1. Number of children at school was divided into five categories. The two extreme categories proved significantly different, as did the lowest two categories contrasted with the highest two, and the lowest category contrasted with the other four combined.
2. At p = 0.07
3. At p = 0.06
4. The total standardized stocks unit were divided into three categories. The two extreme categories were significantly different. The smallest category was significantly different to the other two combined, and the lowest two were significantly to the highest one. On the Scheffe procedure, only the lowest and highest categories were significantly different at the p=0.05 level.
5. At p=0.06
6. 16 percent of all cells had an expected cell frequency of less than 5.
7. At p = 0.07
8. At p = 0.08

Figure 4
Nqutu: Analysis of Relationships Suggested by Factor Solution

Dependent Variables / Independent Variables	ANOVA: Household Income	Assessment Of Assumptions	χ^2 TEST TYPE: No. of Male Working Migrants	χ^2 TEST TYPE: No. of Working Males At Home	χ^2 TEST TYPE: No. of Economically Inactive	χ^2 TEST TYPE: No. of Children at School	χ^2 TEST TYPE: Educational Index of Household
Expenditure of farming	T	A	T	T			
No. of children at school	T	W	T	T	T		
Household Educational Index	T	A	S	T	T	S	
Maize cash crop production	/	D	/	T	/		
Other cash crop production	/	D	T	T	T		
Total Stock Units	T	A	T	T	T		
Cash income from farming	S		T	T[2]			
Size of household	T	A	S	S	S		
Occupational index of households	T[1]	A	S	S		T	S
No. of working male migrants	S	W					
No. of working males at home	T	W					
Male head of household	T	W			T		
No. of economically inactive	T	W	T	S			

n (ANOVA)=±200
n(χ^2)=±100

Legend : as in Figure 2.

Notes : 1. Significant at p = 0.09
2. Significant at p = 0.08 before Yeates correction.

The situation in Mbongolwane was more ambiguous, although the importance of livestock was apparent, possibly reflecting the isolated and more traditional nature of households in this region. Indeed, mean total stock units per household in Mbongolwane was 19, compared to 13 in Mapumulo and 15 in Nqutu. There was also a connexion between the size of the stock holding and the number of economically inactive people in the household. This should not be seen as merely reflecting the labour needs of herding. More isolated and traditional households are also likely to be less integrated into the labour economy and therefore are likely to have more workers who are economically inactive. The data for Nqutu presented no interpretable set of relationships.

Overall, no clear pattern is discernable. Even the rather trivial facts that Nqutu had less involvement in agriculture, Mapumulo had a few small-scale sugar farmers and in Mbongolwane the subsistence sector was more extensive, tell only part of the story. Commercial farms were anything but widespread in Mapumulo. On the other hand, there were some very successful cash crop farmers in Mbongolwane. Probably the most revealing data concern the percentage of total household income emanating from agriculture. In Mapumulo it was 3,7 percent, in Mbongolwane 7,2 percent and in Nqutu 1 percent. In Mapumulo only 15 percent of households received any cash income from agriculture, in Mbongolwane 19 percent and in Nqutu 6 percent. Agricultural involvement, both subsistence and cash crop, formed only a small part of the economic life of the households sampled. That is households relied for their survival on the wage economy.

4.2 Demography and Wage Employment

In Mapumulo, households with higher educational or occupational indices had a greater household income. These two indices were themselves statistically related, indicating that better educated households tended also to have better paying jobs. There were other relationships worth noting:

a) Those households which did have a working male at home, tended to have higher household income.

b) Bigger households appeared to have more male migrant workers and more household members who were economically inactive.

c) Better educated households did not have more migrants.

The situation in Ilbongolwane was more ambiguous (again reflecting the regions greater isolation from the broader South African economy). The important (though possibly insignificant) relationships coincide more or less with the Mapumulo case, except that (i) stock units took the place of cash crop production (ii) working males at home did not feature in the analysis, and (iii) there was a potentially interesting connexion between the educational index and the proportion of household children at school.

The situation in Nqutu is far clearer. Here, larger households had more working males, both migrant and resident, and they had more economically inactive household members. Better educated households had more male migrants and more children at school; they also did

better on the occupational index. Finally households with any male migrants had better incomes.

4.3 Comments on the Economic Structure of the Rural Household

The results discussed in the previous section suggest a general pattern which varies by region but nevertheless appears to apply more consistently in regions better integrated into the wage economy. Larger households had more migrant male workers and more members who were economically inactive. Better educated households had a greater proportion of members with better paying jobs (and in Nqutu had more male migrants). But two points should be borne in mind. Migrants were not necessarily better educated than other adult household members. Indeed almost a quarter (24 percent) of absentees from the three areas had no formal education and 45 percent had received insufficient to ensure the retention of literacy (less than 5 years). Only 6 percent had received a complete secondary schooling. Thus it would appear that while migrants tended to come from better educated households, the migrant labour system does not select out the most educated people of the rural population for migrancy. This conclusion requires further investigation. Other authors have provided conflicting results (Wattrass 1976, Wattrass 1986:6).

In Nqutu, better educated households had more children at school, but there was no evidence that wealthier households spent more on education. In the sample taken as a whole, the relationship between the household educational index and household income was ambivalent.

Migration levels for people of working age were considerably higher for men than women in all three areas (51 percent of males between the age of 15 and 64 and 15.3 percent of females). However, for both sexes, absenteeism was lower in Nqutu reflecting the region's advantageous location (see Table 3). Better transport and the availability of employment in Dundee and Vryheid allowed for 'frontier commuting'. That is, workers travelled 50 or more kilometres from their home in KwaZulu to their job in Natal on a daily or weekly basis. The potential for 'frontier commuting' was far more restricted in Mapumulo and Mbongolwane.

Table 3

Frontier Commuters and Weekly Migrants

(Adult Population, 1 - 64 years of age)

Characteristics	Mapumulo	Nqutu	Mbongolwane
(c) Frontier Commuters	17,1	32,7	19,1
(b) Weekly Migrants	18,7	9,4	5,2
Total (a & b)	35,8	42,1	24,3
Longer Term Migrants	64,2	7,9	75,7
Sample Size	304	287	325

Note : Taken from May & Nattrass (1986:12).

5. The Determinants of Household Income

5.1 Introduction

The relationship between household income and the other variables suggested in the factor analysis, has yet to be formalized. Regression analysis was applied to this task.

Regressions were run for the whole sample, and for each region individually. Within each region analyses were computed for households whose income fell in the top 20 percent of household incomes, in the top 40 percent, top 60 and top 80 percent. A complete discussion of the procedures adopted, statistical assumptions met, and results, obtained for each of these equations is out of the question given the limitations of space and the triviality of some of the findings. Moreover, the data currently available make it totally inappropriate to put much store by a precise statistical model of the economy of the rural household in KwaZulu. At most, the regressions can be seen as formalizing the patterns found in the factor analysis and as a general guide to understanding those variables affecting income in KwaZulu.

5.2 The Data

A basic set of variables* taken from Nattrass (1986) was used to

---

* Viz. FARMINCA, ITEM5, ITEM6, ITEM7, ITEM90, ITEM91, MLAWAY, PENSNER, SIZEHH, TSU, WKSUBS. The calculation of the variables used by Nattrass and those used here was, in some cases, slightly different, with the result that the regression equations are not identical. Substantively there is little dissimilarity.

generate the initial regression equations. This procedure resulted in some problems: (i) the regression models were improperly specified; (ii) since cases were deleted if they lacked information on any one of the variables in the variable set, more cases were excluded (possibly systematically) from the analysis than necessary. It was also clear that multivariate normality could be assumed only tenuously. Some of these problems were partially overcome by rerunning the regression analyses which looked most promising, taking greater care with model specification, removing unreliable variables and coaxing an approximately normal distribution from WEEKIND using logarithmic (to base ten) transformations. Greater emphasis was also placed on the analysis of outliers.

5.3 The Regression Analyses

The most robust of the equations generated was for the top 80 percent of households in the Mapumulo district. This regression will be discussed in some detail and will from the basis of the interpretation of the other regressions.

Four variables - ITEM90, ITEM91, FARMINCA and MLAWAY - were entered into the education. Since prediction is not one of the purposes presently at hand, only the beta weights 'B', part correlation co-efficients '$r_{i(j.k...n)}$' and partial correlation co-efficients '$r_{ij.k...n}$' will be given.

Table 4

Top Eighty Percent of Households - Mapumulo : Regression Coefficients

Variable	BETA	Part Correlation	Partial Correlation
ITEM90	.649	.642	.732
MLAWAY	.253	.252	.379
FARMINCA	.229	.226	.346
ITEM91	.117	.118	.188

n = 114

The three co-efficient types given above represent somewhat different measures of association. The three will usually (and in this case do) rank the variables in the same order of importance. Partial correlation is the measure of the amount of variation explained after the other variables have explained all they could. The beta weights predict how much change in the dependent variable is produced by a standardized change in one of the dependent variable when the others are controlled (Blalock 1972:453). The square of the part correlation gives the increase in R^2 when that variable is added to the equation.

All three measure indicate the dominant importance of income from wages and wage remittances. Indeed, the impact of wages, measured in 'Z' scores, is greater than all the other variables combined. In partial and part correlation terms, ITEM90 accounted for an approximate 50 percent ($.723^2$) increase in explained variation or 41 ($.642^2$) of the 61 R^2 percentage points. The number of working male migrants was the next most important determinant of total household income, cash income from farming the next, and income from non-agricultural informal sector activity the next. However, there was a

5 percent chance that the true slope of ITEM91 was '0' (that is, that there was no linear relationship between informal income and household income).

The regression equation had an adjusted R^2 of .61. Analysis of Variance and an analysis of the residuals indicated that the linearity, normality and homoskedasticity requirements were either well or approximately met. Judging from the correlation table, multicolinearity did not appear to be a problem. Residuals were plotted and regressed against theoretically interesting variables not in the equation : no linear relationships were found. There appeared to be no substantively significant interactive effects between the variables. An analysis of the outliers (three standard deviations above or below the predicted household income) indicated that these households had truly anomalous economic structures. For instance, one household was headed by a woman who practiced traditional medicine locally. Another received extraordinarily high (over R200.00 a month) pension payments. Removal of these cases improved the R^2 marginally. Wage income and wages remitted to the household were by far the most important determinants of household income for those top 80 percent of households in Mapumulo.

Regression analyses of the other income groups in Mapumulo confirmed this pattern. ITEM90 was strongly dominant in all the equations, with FARMINCA and MLAWAY also featuring in all. MLHOME was included in the equation (using the stepwise selection procedure) for only the top 40 percent and top 60 percent of households. ITEM91 did not enter into the equation for the top 20 percent and top 40 percent of households.