Poverty and Social Policy Division
Technical Department
Africa Region
The World Bank

September 20, 1992

CONSTRUCTION OF ANALYTIC VARIABLES AND DATA SETS USING DATA FROM THE COTE D'IVOIRE LIVING STANDARDS SURVEY 1985-88

Concepts, Methodology and Documentation

Gi-Taik Oh
Meera Venkataraman

This paper is an output of the research project "Poverty and Social Dimensions of Structual Adjustment in Cote d'Ivoire, 1985 - 88 -- A Policy-Oriented Analysis" (RPO 675-26). The authors would like to thank Christiaan Grootaert for his close guidance at every stage of this document's evolution. Thanks are also due to Lionel Demery and Ravi Kanbur for the discussions which led to the development of the concepts outlined in this paper. Further, acknowledgements are owed to Andrew McKay and Martin Johnson who developed and constructed many of the underlying Income and Expenditure data sets discussed in this paper.

CONTENTS

I. INTRODUCTION

This paper describes the concepts and estimation involved in the creation of all major analytic data sets and variables utilized in the analysis of the Cote d'Ivoire Living Standards Survey (CILSS) data undertaken for a research project entitled "Poverty and the Social Dimensions of Structural Adjustment in Cote d'Ivoire, 1985-1988 - A Policy-Oriented Analysis (RPO 675-26). The major analytical outputs resulting from this study are listed in the References at the end of this paper.

The research project utilized survey data collected for the Cote d'Ivoire Living Standards Survey (CILSS), a random sample survey of 1600 households undertaken annually over a period of four years (1985-1988). The CILSS was executed by Cote d'Ivoire's Direction de la Statistique under the auspices (in 1985 and 1986) of the World Bank's Living Standards Measurement Study (LSMS). Detailed information about the CILSS is found in Grootaert (1986) and Ainsworth and Munoz (1986).

This documentation is primarily intended for data analysts who will be working with the data sets and analysis files created for this research project. These user files form an important part of the research project's output. The user data file package would typically include: Income and Expenditure Aggregate and Sub-Aggregate data files, Corrective Weights, Regional and Temporal Price Index data and, also, Panel data sets. The process by which the major analytic variables and data sets were created, is outlined and described in this document. Thus, readers of the analytical papers resulting from this study who may want further information and details on these aspects, would also find this document useful.

Discussion of the five major topics included in this paper corresponds to the contents of the Data Package described above - (i) Income and Expenditure Aggregates (ii) Data Cleaning and Preparation (iii) Corrective Weights (iv) Price Index construction and (v) Panel data sets. Knowledge of the hierarchical nature of the Income and Expenditure data sets (Chapter 1) is necessary before issues such as treatment of outliers and imputation of

missing values can be described in Chapter 2, on Data Cleaning and Preparation.

BACKGROUND NOTES

All data sets available in the research project's package are stored in SAS (Statistical Analysis System) format. This background is intended to provide the reader with requisite technical details, in particular, the file naming conventions utilized to facilitate identification of file contents.

Since the CILSS spans four years of survey data the data files we worked with were named such that the year of the survey could be easily identified from the name of the SAS data set. As those who are familiar with SAS will know, a SAS data set is identified by three components: the filename, filetype and the filemode. The filemode is system specific and inappropriate for discussion in this context, but a short description of our efforts in using appropriate naming conventions for the filename and filetype, should facilitate the user's access to the needed data set. The filetype for each file reflects the year of the survey data represented in that file. In particular:

Filetype	Survey Year
CI85	1985
CILS2	1986
CI87	1987
CI88	1988

The filename of the data set was chosen to reflect the type of data stored, such as ALLWAITS for a data set containing sampling weights. As an example, a data set containing weights for the year 1985 would be named ALLWAITS CI85. Each of the following Chapters provides specific information on pertinent data set names.

Most analyses in the Project were performed on a regional basis. Consequently, a variable representing values for Region is to be found in most higher-level data sets. The five regions used were Abidjan, Other Cities, East Forest, West Forest and Savannah.

II. HOUSEHOLD INCOME AND EXPENDITURE AGGREGATES

Standard measurement of household welfare derived from household survey data relies on monetary measures of household incomes and expenditures. The conceptual justification for this is explained by Grootaert (1983) and Glewwe (1987), among others. While some analysts utilize specific components of household expenditure such as **expenditure on food** or **wage income** as the indicator of welfare, most analysts require the identification of much broader measures of total household incomes and expenditures. The complex process by which these totals are derived is described in detail by Johnson, McKay and Round (1990).

At the first stage, variables from the questionnaire are used to build up estimates of income/expenditure from/on specific items - such as, annual income from each of the household's nonfarm enterprises. Next, these values are summed up over all relevant cases, whether they be individual household members, specific crops grown by the household or different nonfarm enterprises (as in our example), to yield a household level estimate. At this level, the summed values are known as sub-aggregates. It should be noted that sub-aggregates may also be derived by aggregating variables that are either conceptually similar or that complement each other (for instance, cash and in-kind incomes from the secondary jobs are aggregated together to yield the sub-aggregate level variable S7TOT, wage income from secondary jobs). When the sub-aggregates have been derived either by summation or aggregation, they are grouped together in a manner such that their constituent components are conceptually similar and then aggregated across all sub-aggregate categories within each group, to yield the so-called aggregate values - i.e. estimates of household income/expenditure from/on a relatively broad category; for instance, income from all non-farm enterprises operated by the household.

The hierarchical nature of the variable level, sub-aggregate and aggregate level data sets is explained by Johnson, McKay and Round (1990) as follows: "The clear and flexible approach is achieved by constructing a hierarchical data set relating to the household

accounting, with at least three levels to the hierarchy. At the lowest level (the variable level) would be those survey responses of direct relevance to the household accounts, each expressed on an annual basis for consistency. The next level (the sub-aggregate level) would comprise aggregations of a small number of these annualized variables which are very similar in nature. At the highest level (the aggregate level) these sub-aggregates would be aggregated into a small number of totals each of which are conceptually distinct." (p.10)

While this three-tiered hierarchy is the conceptual framework for this exercise and indeed most variables, subaggregates and aggregates adhere to this pattern, there are, nevertheless, a few exceptions to the rule. In some cases, the sub-aggregate category may be considered broad enough to stand on its own and in these cases, the aggregate level variable, although named differently, will have the exact same values as the sub-aggregate level variable. Expenditure on purchased food, for instance, is not transformed from the sub-aggregate to the aggregate level but is named differently at each level. In the case of one aggregate level variable - wage income, the data set contains information on wage income for *each* household member, rather than wage income for the *entire* household, thus clearly deviating from the stated pattern.

Starting from selected variables from the CILSS data sets corresponding to each of the sections in the CILSS questionnaire, this conceptual framework was applied to the creation of hierarchical data sets at the variable, sub-aggregate and aggregate level. The data set filenames are documented in the following pages. Following this, we document the creation of the Income and Expenditure variables giving precise definitions and information which will allow the user to trace the computation of Total Income and Expenditure variables back to their aggregate level components, further back to the sub-aggregate and variable level components and ultimately, to variables from the questionnaire. This will permit the reader to identify the precise origin and evolution of the Total Income and Expenditure variables that are found in the final user files.

Documentation on File Names

Questionnaire Level Variables

These are stored in data set files corresponding to each section of the questionnaire; F1A CI85, for example, contains 1985 data corresponding to Chapter 1A of the questionnaire. The naming conventions vary slightly, F1A in 1985, F01A in 1986 and SEC1A in 1987 and 1988, but the underlying principle and variable names remain the same.

Individual Item Level Variables

These are stored in the BIC (Income) and BEC (Expenditure) files. These files contain selected income and expenditure variables that have been standardised and cleared of outlying values (explained in Chapter II).

Sub-Aggregate Level Variables

These are stored in the IEC files (IEC1-IEC43). They contain higher level variables that have undergone the necessary summation and/or aggregation procedures described earlier.

Aggregate Level Variables

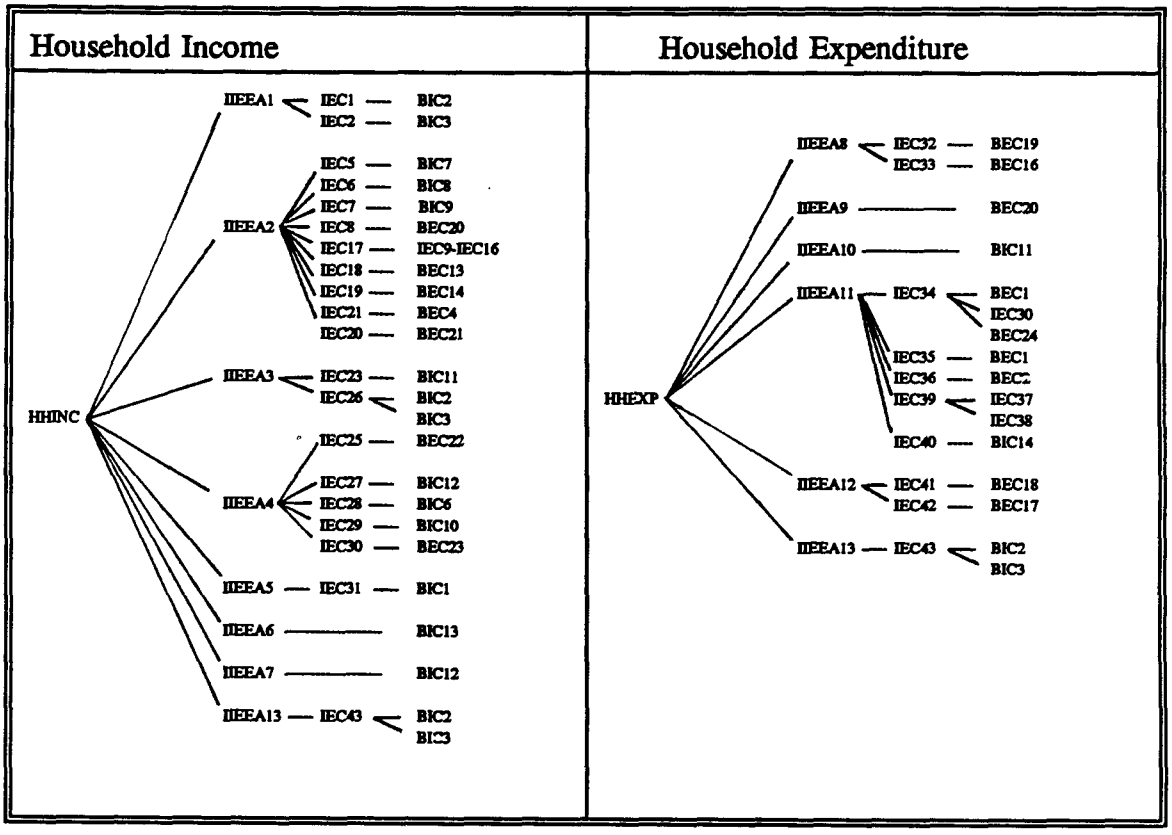
These estimates of the broadest income and expenditure categories are stored in the IIEEA files (IIEEA1-IIEEA13). The first 7 files, IIEEA1-IIEEA7 contain Income variables and the rest, IIEEA8 - IIEEA13 contain Expenditure related variables. The IIEEA files contain data that have been cleared of outliers at the aggregate level. (Chapter II contains much more on this topic). The same aggregate-level data, not cleared of outliers, is found in files IIEEC1 - IIEEC13. The same aggregate level variables found in the IIEEA files have different names in the IIEEC files. Figure 2.1 of the next chapter provides both sets of names.

Total Income and Expenditure Variables

At this level, the broad aggregates from the Income related IIEEA files have been brought together into one file, HHINC (filename), representing all Income variables used in the calculation of Total Income. The broad aggregates have then been aggregated further to yield Total Income for each household. Analogous operations were executed with the Expenditure related files (IIEEA8-IIEEA13) to yield a file, HHEXP, containing Total Expenditure for each household in the survey.

Figure 1.1 shows the precise nature of the hierarchical relationships between the data sets.

Figure 1.1



VARIABLE DEFINITIONS

As mentioned earlier, this section documents the creation of the Income and Expenditure variables. Component variables (each labelled with a unique numeric identification) used to build up the higher-level variable in question, are specified and defined. Following this, each of the component variables are defined in turn. The process of defining component (aggregate level down to sub-aggregate and further down to variable level) variables continues until the evolution of each of the higher-level variable is traced back to the BIC-BEC level files. The last piece of documentation in this section (found on pp. 19-24) documenting the derivation of each of the BIC-BEC level variables from the CILSS questionnaire, allows the analyst to further pursue the trail to the questionnaire, if so desired.

This documentation covers only those variables which were selected, from among alternative estimates of the same variable, to be included in the definitions of Total Income and Expenditure used in this research project. For example, of two alternative estimates on food expenditure (FOODEXP1 based on a 2 week recall period and FOODEXP2 based on a one year recall period), only the estimate used in our research project would be described here. For information on alternate estimates, refer to Johnson, McKay and Round (1990).

DESCRIPTION OF INCOME & EXPENDITURE VARIABLES

1. Total Household Income (HHINC)

HHINC = WGMY7 + FARMY - FARMDEP + NFYAVG - NFDEP + RENTY
+ SCHOLY3 + REMITY + OTHERY

Household income is built from the following broad aggregates:

1.1 wage income (WGMY7)
1.2 farm income (FARMY)
1.3 farm equipment depreciation (FARMDEP)
1.4 non-farm income (NFYAVG)
1.5 non-farm capital asset depreciation (NFDEP)
1.6 rental income (RENTY)
1.7 income from scholarships (SCHOLY3)
1.8 income from remittances (REMITY)
1.9 other income (OTHERY)

(All these variables are found in file HHINC).

1.1) Wage Income (WGMY7)

The first component variable of HHINC. It is found in the Total Income file, HHINC. As an aggregate level variable, it is found in the file IIEEA1.

WGMY7 = M7TOT + S7TOT

WGMY7 is defined as 'total employment income for each individual'. It is based on self-reported income over a 7-day reference period. It is built from the following components:

1.1.1) Total wage income from primary job of last 7 days (M7TOT)
1.1.2) Total wage income from secondary job of last 7 days (S7TOT)

1.1.1) Wage Income from Primary Job (M7TOT)

The first component variable of WGMY7, representing total wage income from primary job of the past 7 days. It is built from the following components.

M7TOT = M7CASH+ M7BONK + M7K + M7H + M7CL + M7TR + M7OTH

1.1.1.1 M7CASH = Employment income in cash, main job, past 7 days
1.1.1.2 M7BONK = Bonus income not already included in M7CASH, main job, past 7 days
1.1.1.3 M7K = Employment income in food, main job, past 7 days
1.1.1.4 M7H = Employment income paid as subsidised housing, main job, past 7 days
1.1.1.5 M7CL = Employment income as subsidised clothing, main job, past 7 days
1.1.1.6 M7TR = Employment income as subsidised transport, main job, past 7 days
1.1.1.7 M7OTH = Employment income paid in other forms, main job, past 7 days
(all variables at the level of each applicable household member).

Each of these variables is found in data set BIC2. Refer to the documentation at the end of this Chapter to further trace the origins to the CILSS questionnaire level variables.

1.1.2) Total Wage Income from Secondary Job (S7TOT)

S7TOT is the second component variable of WGMY7, representing wage income from the secondary job of the past 7 days. It is built from the following variables.

S7TOT = S7CASH + S7K

1.1.2.1 S7CASH = Employment income in cash, secondary job, past 7 days
1.1.2.2 S7K = Employment income in kind, secondary job past 7 days
(all variables at individual level).

Both variables are found in file BIC3.

1.2) Farm Income (FARMY)

FARMY is the second component variable of HHINC. It stands for 'net household agricultural income' and is found in both in data set HHINC and in IIEEA2.

FARMY = CRPSINC + TRCRPINC + ANIMINC + NOSHHOME
- TOTCRPXP - INPTRANS - INPREAR - SHAREXP
(from files IEC5, IEC6, IEC7, IEC8, IEC17, IEC18, IEC19, IEC21 respectively)

1.2.1) CRPSINC = revenue from sale of crops. It is derived by summing up INCCRPS (found in file BIC7) over all crops. INCCRPS = revenue from sale of crops (at level of each individual crop).

1.2.2) TRCRPINC = revenue from the sale of transformed crop products. Derived by aggregating INCTRCRP (BIC8) over all products.
INCTRCRP = revenue from sale of transformed crop products (at level of each individual product).

1.2.3) ANIMINC = revenue from the sale of animal products. Derived by aggregating INCANIM (BIC9) over all products. INCANIM = revenue from sale of animal products (at level of each individual product.

1.2.4) NOSHHOME = consumption of home production. Derived by aggregating HOMENOSH (BEC20) over all commodities. HOMENOSH = consumption of home production (at level of each individual item).

1.2.5) TOTCRPXP = total expenditure on inputs for growing crops. It is explained in full on the next page.

1.2.6) INPTRANS = expenditure on inputs for transformed food products and is derived by aggregating TRANSINP (BEC13) over all products.
TRANSINP = expenditure on inputs for transformed crop products (at level of each individual product).

1.2.7) INPREAR = expenditure on inputs for livestock rearing and is derived by aggregating REARINP (BEC14) over all products.
REARINP = expenditure on inputs for livestock rearing (at level of each individual item.

1.2.8) SHAREXP = expenditure on renting land (at household level) and is derived from SHARE (BEC4) which is 'expenditure on renting land (at household level).'

1.2.5) Total expenditure on inputs for growing crops (TOTCRPXP)

This is the fifth component variable of FARMY.

TOTCRPXP = SEEDSEXP + FERTEXP + MANUREXP + INSEXP
+ TRANSEXP + SACKEXP + STOREXP + LABEXP + EXPOTH
(from files IEC9, IEC10, IEC11, IEC12, IEC13, IEC14, IEC15, IEC16, IEC16 respectively)

1.2.5.1) SEEDSEXP = (household level) expenditure on seeds. Derived by aggregating SEEDS (BEC5) over all crops. SEEDS = expenditure on seeds (at level of each individual crop).

1.2.5.2) FERTEXP = (household level) expenditure on fertiliser. Derived by aggregating FERT (BEC6) over all crops. FERT = expenditure on fertilizer (at level of each individual crop).

1.2.5.3) MANUREXP = expenditure on manure. Derived by aggregating MANURE (BEC7) over all crops. MANURE = expenditure on organic manure (at level of each individual crop).

1.2.5.4) INSEXP = expenditure on insecticides and herbicides. Derived by aggregating INSECT (BEC8) over all crops. INSECT = expenditure on insecticides/herbicides (at level of each individual crop).

1.2.5.5) TRANSEXP = expenditure on inputs for crops - transport. Derived by aggregating TRANSPT (BEC9) over all crops. TRANSPT = expenditure on transporting crops (at level of each individual crop).

1.2.5.6) SACKEXP = expenditure on inputs for crops - sacks etc. Derived by aggregating SACKS (BEC10) over all crops. SACKS = expenditure on sacks/twine/containers (at level of each individual crop).

1.2.5.7) STOREXP = expenditure on storing crops and is derived by aggregating STORAGE (BEC11) over all crops. STORAGE = expenditure on storage (at level of each individual crop).

1.2.5.8) LABEXP = expenditure on inputs for growing crops - labour'. LABEXP is taken from XPLAB (BEC12) which is 'expenditure on labour (at household level).'

1.2.5.9) EXPOTH = expenditure on other inputs for growing crops. It is taken from OTHINPEX (BEC12) which is 'expenditure on other crop inputs (at household level).'

1.3) Farm Equipment Depreciation (FARMDEP)

FARMDEP = depreciation of farm equipment. It is found in the HHINC and IIEEA2 datasets. It is taken from DEPNEQ (IEC20), which is derived by aggregating EQDEPN (BEC21) over all categories. EQDEPN is 'depreciation of farm equipment (at level of each individual item).'

1.4) Average Non-Farm Income (NFYAVG)

This is the fourth component variable of HHINC and is found in the HHINC and IIEEA3 datasets. NFYAVG is derived by taking the average of NFYDIR and NFSEY2. Both NFYDIR and NFSEY2 are aggregate level variables found in file IIEEA3?. As aggregate level variables, they are not to be considered component variables of NFYAVG. NFYAVG reflects the average value of two different estimates [NFYDIR (a) and NFY5 (b)] of Household Non-Farm Income.

NFYAVG = (NFYDIR + NFY5) / 2

1.4a) NFYDIR = net non-farm self-employment income.

NFYDIR = PROFITNF + NFDOMINC (both from IEC23)

1.4a.1) PROFITNF = money profit from non-farm enterprises (at the household level). It is derived by aggregating the sum of two variables, (PRNFDOM + PRNFUND, both from BIC11) over all firms run by the household. PRNFDOM is 'profit of non-farm enterprises used within household' and PRNFUND is 'retained profit of non-farm enterprises (at level of each enterprise).'

1.4a.2) NFDOMINC = value of output of nonfarm enterprises consumed domestically (at the household level). This is derived by aggregating INCNFDOM (BIC11) over all enterprises. INCNFDOM is 'value of output of non-farm enterprises consumed domestically (at level of each enterprise).'

1.4b) NFY5 = net non-farm self-employment income. It is found in IIEEA3.

NFY5 = SENONF (IEC26) + NFDOMINC (IEC23)

1.4b.1) SENONF = non-farm self-employment income. It is derived by aggregating the sum of [SE7CASH (BIC2) + SSE7CASH (BIC3)] for non-agricultural activities only, over all members of the household.' SE7CASH = self-employment income in cash, main job, past 7 days (all variables at individual level). SSE7CASH = self employment income in cash, secondary job past 7 days (all variables at individual

level).

1.4b.2) NFDOMINC - Refer to variable 1.4a.2, same definition.

1.5) Non-Farm Capital Asset Depreciation (NFDEP)

NFDEP = non-farm capital assets depreciation. Found in HHINC and IIEEA3. NFDEPN is taken from DEPNASS (IEC25). DEPNASS is the aggregation of ASSDEPN (BEC22) over all enterprises and asset categories. ASSDEPN = depreciation of non-farm capital assets (at level of each individual item).

1.6) Rental Income (RENTY)

RENTY stands for 'rent and imputed rent'. Available in dataset HHINC and IIEEA?.

It is formulated as follows:

IF MISCINC1 GT (LNDINC + LEASINC) THEN RENTY = MISCINC1 + RENTIMP
IF MISCINC1 EQ (LNDINC + LEASINC) THEN RENTY = MISCINC1 + RENTIMP
IF (LNDINC + LEASINC) GT MISCINC1 THEN RENTY = LNDINC + LEASINC + RENTIMP

(MISCINC1 from IEC27, LNDINC from IEC28, LEASINC from IEC29, RENTIMP from IEC30)

1.6.1) MISCINC1 = income from leasing land/equipment/etc and is taken from INCMISC (BIC12, category 413 only) which is 'miscellaneous income (at level of each individual category).'

1.6.2) LNDINC = income from renting out land etc and is from INCLND (BIC6) which is for 'income for renting out land etc (at household level).'

1.6.3) LEASINC = income from leasing farm equipment and is derived by aggregating INCLEAS (BIC10) over all categories. INCLEAS is 'revenue from leasing farm equipment (at level of each individual item).'

1.6.4) RENTIMP = imputed rent of owner occupied dwellings and is taken from IMPRNT (BEC23) which is 'imputed rent of owner-occupied dwellings (at household level).'

1.7) Scholarships (SCHOLY3)

SCHOLY3 = income from educational scholarships/grants and is found in File IIEEA5. It is taken from the sub-aggregate level variable SCHOL1 (IEC31). SCHOL1 is a household-level variable derived by aggregating SCHOL (BIC1) over all individuals. SCHOL means 'value of educational scholarship (at individual level).'

1.8) Remittance Income (REMITY)

REMITY stands for 'income from remittances' and is derived by aggregating INCREM (BIC13) over all remittances to arrive at a household level figure. INCREM is 'income from remittances (at level of each individual remittance).'

1.9) Other Income (OTHERY)

OTHERY stands for 'other income' and is found in HHINC and IIEEA7. It is derived by aggregating INCMISC (BIC12) over the following categories: 401, 402, 403, 404, 405, 406, 408, 412, 417. INCMISC means 'miscellaneous income (at level of each individual category).'

2. Total Household Expenditure (HHEXP)

Household expenditure consists of the following aggregates:

2.1 average food expenses (FOODEAVG)
2.2 farm product home consumption (FARMHC)
2.3 non-food home consumption (NFHC)
2.4 other expenditures (OTHERE)
2.5 paid remittances (REMITE)
2.6 wage income in kind (WGKY7).

Household expenditure (HHEXP) is formulated as follows:

HHEXP=FOODEAVG +FARMHC + NFHC + OTHERE + REMITE +WGKY7.

All variables are found in file HHEXP as well as in specific IIEEA files.

2.1) Expenditure on Purchased Food (FOODEAVG)

FOODEAVG is the average of two different estimates of food expenditure. One estimate is based on the respondent's recall of expenditures over the last 14 days (FOODE2W) and the second estimate (FOODEAN) is based on a one year recall period. Both variables are found in IIEEA8.

FOODAVG = (FOODE2W + FOODEAN) / 2

2.1a) FOODE2W = food expenses - first estimate, based on 2 week recall period.

FOODE2W = FOODEXP1 (IEC32) + EXPDAY (IEC33, category 101 only)

2.1a.1) FOODEXP1 is derived by aggregating NOSH1 (BEC19) over all categories. NOSH1 is 'food expenses - 1st estimate (at level of each individual item).'

2.1a.2) EXPDAY - EXPDAY is from DAYEXP (BEC16), category 101 only (expenditure on food consumed outside the home). DAYEXP is 'daily expenses (at level of each individual item).'

2.1b) FOODEAN = 'food expenses - second estimate'.

FOODEAN = FOODEXP2 (IEC32) + EXPDAY (IEC33, category 101 only)

2.1b.1) FOODEXP2 is derived by aggregating NOSH2 (BEC19) over all categories. NOSH2 is 'food expenses - 2nd estimate (at level of each individual item).'

2.1b.2) EXPDAY is defined the same way as in 2.1a.2

2.3) Consumption of Home Produced Food (FARMHC)

The third component variable of HHEXP. FARMHC stands for 'consumption of home produced food' and is found in File IIEEA9. It is derived by aggregating HOMENOSH (BEC20) over all categories. HOMENOSH is 'consumption of home production (at level of each individual item).'

2.4) Consumption of Home Produced Non-Food Items (NFHC)

NFHC stands for 'consumption of home-produced non-food items' and is found in File IIEEA10. It is derived by summing INCNFDOM (BIC11) over all enterprises. INCNFDOM is 'value of output of non-farm enterprises consumed domestically (at level of each enterprise).'

2.5) Other Expenditure (OTHERE)

OTHERE stands for 'other expenditure' and is found in File IIEEA11.

OTHERE = EXPHOUSE + HHUTILS + EXPEDUC + EXPYRDAY
+ VALUSE
(from files IEC34, IEC35, IEC36, IEC39, IEC40 respectively)

2.5.1) EXPHOUSE - expenditure on housing
2.5.2) HHUTILS - expenditure on housing utilities
2.5.3) EXPEDUC - expenditure on education
2.5.4) EXPYRDAY - daily and annual expenses
2.5.5) VALUSE - Use value of durable goods

All variables are explained in detail on the next page.

2.5.1) Expenditure on Housing (EXPHOUSE)

EXPHOUSE = RENTM + RENTGS (both from BEC1) + RENTIMP (IEC30) + IMPRNTPA (BEC24)

2.5.1.1 RENTM = Rent paid in cash (at the household level)
2.5.1.2 RENTGS = Rent paid in goods and services (at household level)
2.5.1.3 RENTIMP = IMPRNT (BEC23)
2.5.1.4 IMPRNT = Imputed rent of owner-occupied dwellings (at household level)
2.5.1.5 IMPRNTPA = Imputed value of rent paid by relatives (at household level)

2.5.2) Expenditure on household utilities (HHUTILS)

HHUTILS = WATB + ELECB (both from BEC1)

2.5.2.1 WATB = Annual water bill (at household level)
2.5.2.2 ELECB = Electric bill (at household level)

2.5.3) Expenditure on education (EDUCEXP)

Derived by summing up EDUCEXP (BEC2) over all individuals. EDUCEXP is 'expenditure on education (at individual level).'

2.5.4) Daily and Annual Expenses (EXPYRDAY)

EXPYRDAY = EXPDAY (IEC37) + EXPYEAR (IEC38)

2.5.4.1) EXPDAY = daily expenses. It is derived by aggregating DAYEXP (BEC16) over all categories except 101 and DAYEXP is 'daily expenses (at level of each individual item).'

2.5.4.2) EXPYEAR = yearly expenses. It is derived by aggregating YREXP2 (BEC17) over all categories except 122, 126, 127, 139, 143, and 145 and YREXP2 is 'annual expenses - 2nd estimate (both at level of each individual item).'

2.5.5) Use Value of Durable Goods (VALUSE)

VALUSE is derived by aggregating USEVAL (BIC14) over all goods and USEVAL is 'use value of durable goods (at level of each individual good).'

2.6) Remittances Paid Out (REMITE)

REMITE stands for 'expenditure on remittances paid out' and is found in File IIEEA12.

REMITE = EXPREMS (IEC41) + LOSTCASH (IEC42)

2.6.1) EXPREMS - 'expenditure on remittances' is aggregated from REMITS (BEC18) over all remittances and REMITS is 'expenditure on remittances (at level of each individual remittance).'

2.6.2) LOSTCASH - 'money lost' comes from YREXP2 (BEC17), category 145 only. YREXP2 is 'annual expenses - 2nd estimate (both at level of each individual item).'

2.7) Wage Income In Kind (WGKY7 - IIEEA13)

WGKY7 means 'imputed consumption of employment income received in kind - 1st estimate' and is found in File IIEEA13. It is derived by aggregating TOTKIND1 (IEC43) over all individuals in the household.

TOTKIND1 = M7KTOT + S7KTOT
M7KTOT = M7K + M7H + M7CL + M7TR + M7OTH (all from BIC2)
S7KTOT = S7K (all from BIC3)

The variables M7KTOT and S7KTOT, have been defined in the Income side of the Household Accounts, refer to Variables 1.1.1 and 1.1.2

Documentation on derivation of BIC-BEC level variables
from the CILSS questionnaire variables

Disaggregated Household Income Variables (expressed on an annual basis)

Note : The documentation in brackets refers to the source question in the CILSS questionnaire, e.g. 5B10 refers to Question 10 from Section 5B of the CILSS questionnaire.

BIC1

SCHOL = Value of educational scholarship - from CILSS questionnaire (3A2)
(at individual level)

BIC2

M7CASH = Employment income in cash, main job, past 7 days - (5B10, 5B21)
SE7CASH = Self-employment income in cash, main job, past 7 days - (5B10)
M7BONC = Bonus income already included in M7CASH, main job, past 7 days - (5B21)
M7BONK = Bonus income not already included in M7CASH, main job, past 7 days - (5B21)
M7K = Employment income in food, main job, past 7 days - (5B23)
M7H = Employment income paid as subsidised housing, main job, past 7 days - (5B25)
M7CL = Employment income as subsidised clothing, main job, past 7 days - (5B27)
M7TR = Employment income as subsidised transport, main job, past 7 days - (5B29)
M7OTH = Employment income paid in other forms, main job, past 7 days - (5B31)
(all variables at individual level)

BIC3

S7CASH = Employment income in cash, secondary job, past 7 days - (5C9)
SSE7CASH = Self employment income in cash, secondary job past 7 days - (5B10, 5C9)
S7K = Employment income in kind, secondary job past 7 days - (5C16)
(all variables at individual level)

BIC4

M12CASH = Employment income in cash, main job past 12 months - (5E11 - 5E21)
SE12CASH = Self-Employment income in cash, main job past 12 months - (5E11)
M12BONC = Bonus income already included in M12CASH, main job, past 12 months - (5E21)
M12BONK = Bonus income not already included in M12CASH, main job, past 12 months - (5E21)
M12K = Employment income in food, main job, past 12 months - (5E23)
M12H = Employment income paid as subsidised housing, main job, past 12 months - (5E25)
M12CL = Employment income as subsidised clothing, main job, past 12 months - (5E27)
M12TR = Employment income as subsidised transport, main job, past 12 months - (5E29)
M12OTH = Employment income paid in other forms, main job, past 12 months - (5E31)
(all variables at individual level)

BIC5

S12CASH = Employment income in cash, secondary job past 12 mths - (5G10)
SSE12CAS = Self-Employment income in cash, secondary job past 12 mths - (5G10)
S12K = Employment income in kind, secondary job past 12 mths - (5G15)
(all variables at individual level)

BIC6

INCLND = Income for renting out land etc - (9A14)
(at household level)

BIC7

INCCRPS = Revenue from sale of crops - (9B4, 9B5)
(at level of each individual crop)

BIC8

INCTRCRP = Revenue from sale of transformed crop products - (9E2, 9E4, 9E5)
(at level of each individual product)

BIC9

INCANIM = Revenue from sale of animal products - (9G3)
(at level of each individual product)

BIC10

INCLEAS = Revenue from leasing farm equipment - (9K)
(at level of each individual item)

BIC11

INCNFC = Revenue in cash from non-farm enterprises - (10A19)
INCNFK = Revenue in goods/services from non-farm enterprises - (10A21)
INCNFDOM = Value of output of non-farm enterprises consumed domestically - (10 A23)
INCNFTOT = Total revenue of non-farm enterprises in cash and in kind - (10A25)
PRNFDOM = Profit of non-farm enterprises used within household - (10A28)
PRNFUND = "Retained" profit of non-farm enterprises - (10A30)
(at level of each enterprise)

[Note that INCNFC and INCNFK are calculated only for businesses which were operational in the two week period between the interviews, and INCNFTOT is estimated only for enterprises which were not operational in this two week period.]

BIC12

INCMISC = Miscellaneous income - (14A2)
(at level of each individual category)

BIC13

INCREM = Income from remittances - (14B6)
(at level of each individual remittance)

BIC14

USEVAL = Use value of durable goods - (11C4)
(at level of each individual good)

Disaggregated Household Expenditure Variables (expressed on an annual basis)

BEC1

RENTM = Rent paid in cash - from questionnaire (2B4)
RENTGS = Rent paid in goods and services - (2B16)
WATB = Annual water bill - (2B14, 2B15)
ELECB = Electric bill - (2B26, 2B27)
(at household level)

BEC2

EDUCEXP = Expenditure on education - (3A18)
(at individual level)

BEC4

SHARE = Expenditure on renting land - (9A20)
(at household level)

BEC5

SEEDS = Expenditure on seeds - (9D3, 9D9)
(at level of each individual crop)

BEC6

FERT = Expenditure on fertilizer - (9D1)
(at level of each individual crop)

BEC7

MANURE = Expenditure on organic manure - (9D15)
(at level of each individual crop)

BEC8

INSECT = Expenditure on insecticides/herbicides - (9D20)
(at level of each individual crop)

BEC9

TRANSPT = Expenditure on transporting crops - (9D26)
(at level of each individual crop)

BEC10

SACKS = Expenditure on sacks/twine/containers - (9D30)
(at level of each individual crop)

BEC11

STORAGE = Expenditure on storage - (9D36)
(at level of each individual crop)

BEC12

XPLAB = Expenditure on labour - (9D39)
OTHINPEX = Expenditure on other crop inputs - (9D41)
(at household level)

BEC13

TRANSINP = Expenditure on inputs for transformed crop products - (9E7)
(at level of each individual product)

BEC14

REARINP = Expenditure on inputs for livestock rearing - (9I2)
(at level of each individual item)

BEC15

NFINP = Expenditure on (current) inputs of non-farm enterprises - (10B2, 10B3)
(at level of each individual enterprise and each category)

BEC16

DAYEXP = Daily expenses - (11A2)
(at level of each individual item)

BEC17

YREXP1 = Annual expenses - 1st estimate - (11B3, 11B4)
YREXP2 = Annual expenses - 2nd estimate - (11B3, 11B4)
(both at level of each individual item)

BEC18

REMITS = Expenditure on remittances - (11D6)
(at level of each individual remittance)

BEC19

NOSH1 = Food expenses - 1st estimate - (12A3, 12A4, 12A5, 12A6)
NOSH2 = Food expenses - 2nd estimate - (12A3, 12A4, 12A5, 12A6)
(at level of each individual item)

BEC20

HOMENOSH = Consumption of home production - (12B3, 12B4, 12B5)
(at level of each individual item)

BEC21

EQDEPN = Depreciation of farm equipment - (9K4)
(at level of each individual item)

BEC22

ASSDEPN = Depreciation of non-farm capital assets - (10C2)
(at level of each individual item)

BEC23

IMPRNT = Imputed rent of owner-occupied dwellings - (2B1)
(at household level)

BEC24

IMPRNTPA = Imputed value of rent paid by relatives - (2B1)
(at household level)

III. DATA CLEANING AND PREPARATION

Raw survey data requires a number of stages of initial cleaning and preparation before any transformation and aggregation can take place. This includes routine cleaning, standardisation of values, clearing data of outliers or extreme values, and imputation of missing values.

[While this discussion logically should precede the earlier discussion on Income and Expenditure Aggregates in Chapter 1, in practice it becomes necessary to maintain this arrangement since some sections of the following write-up presume knowledge of the hierarchical relationship between the various levels (questionnaire level, variable level, sub-aggregate and aggregate level) of data.]

Initial cleaning

(i) Invalid data values in the survey data files had to be corrected or removed.

(ii) Duplicate records which were present in some files had to be eliminated.

(iii) Records for individuals who are not listed as household members in the Roster Chapter (Chapter 1A) needed to be deleted.

(iv) In a small number of instances, records which are present in other data sets for households which do not exist in the household roster file were deleted.

Standardisation

The responses given in the questionnaire generally relate to a range of different reference periods; for example, earnings may be quoted per week, per month, per year etc. Before any aggregation can take place it was necessary to express these responses on a standardised annual basis.

Identification of outliers

Once the relevant variables from the surveys have been standardised, they may be compared across households to search for any outliers or extreme values of the variable in question. There are, of course, multiple and varied approaches for the identification and treatment of outliers, some of which are time consuming processes, often requiring subjective judgements. Given the large number of primary and intermediate variables utilized in our study, it seemed expedient to use an automated procedure.

Outliers were defined as those values lying more than five standard deviations from the mean value of the variable in question. This statistical approach or treatment was performed for variables with more than thirty observations. For variables with more than thirty observations, the statistical procedure becomes questionable since statistics derived from a small sample (less than 30 observations) are generally considered suspect. For such variables, outliers were identified by visual inspection of the data. More will be said on the visual inspection method later in this chapter.

Outlying values were identified and treated at two stages; first, at the stage of the standardised questionnaire-level variables and secondly, at the aggregate level. At the variable level, values were identified as outliers if they lay more than five standard deviations from the national mean of the distribution. For variables at the aggregate level, however, the process was refined to include the regional dimension; an outlier was defined as a value lying more than five standard deviations of the corresponding regional mean. (The standard deviation was also calculated on a regional basis). In order to provide for a sufficient number of observations, the regional breakdown was restricted to Abidjan, Other Cities and Rural Areas rather than the standard five regions used in most analyses of CILSS data.

Replacement of Outliers

The statistical procedure employed for variables with more than thirty observations first identified outliers as explained above, then excluded those outlying values to obtain a new mean value of the variable and then replaced those outliers with this new mean. The exercise was performed twice, to account for the possibility that some outliers may not have been identified in the first iteration as the mean and standard deviation derived from the first iteration might have been severely affected by the extreme values. For the second iteration, the mean and standard deviation resulting from the new distribution were utilized. Refer to Table 2.1 for documentation on the numbers of outliers identified and replaced.

Further on the treatment of outliers, the following specifics are worth noting :

- In the case of variables which have commodity or other codes associated with them (such as food expenditure or revenue from the sale or crops), the outlier identification and replacement process was applied at the individual commodity levels.

- Per-capita values were used to account for the systematic variation of most household expenditures with the size of the household in question. After outliers have been identified and replaced at the per-capita level, all values of the variable are then multiplied by the corresponding household size to yield household level values.

Figure 2.1 NUMBER OF OUTLIERS IDENTIFIED AND REPLACED, USING THE STATISTICAL METHOD

(For Aggregate-level variables)

New Variable Names (IIEEA files) / Old Variable Names (IIEEC files)		1985			1986			1987			1988		
		no. of outliers	no. of observations	% of outliers	no. of outliers	no. of obs	% of outliers	no. of outliers	no. of obs	% of outliers	no. of outliers	no. of obs	% of outlier
WGMY7	(TOTALMP1)	4	551	% 0.73	4	540	% 1.11	8	589	% 1.36	5	520	% 0.96
WGMY12	(TOTALMP2)	5	670	0.75	9	634	1.42	6	650	0.92	7	576	1.22
FARMY	(HHAGINC1)	8	995	0.80	7	998	0.70	4	142	2.82	2	138	1.45
FARMY5	(HHAGINC2)	7	1020	0.69	6	1014	0.59	7	846	0.83	8	918	0.87
FARMDEP	(HHAGDEPN)	4	224	1.79	2	212	0.94	7	850	0.82	9	925	0.97
NFY	(NFSEY1)	4	233	1.72	5	314	1.59	5	283	1.77	3	351	0.85
NFYDIR	(NFSEY2)				5	529	0.94	8	464	1.72	4	489	0.82
NFY5	(NFSEY3)	10	512	1.95	9	469	1.92	7	474	1.48	3	497	0.60
NFDEP	(NFDEPN)	10	422	2.37	9	409	2.20	6	355	1.69	10	394	2.54
RENTY	(IMPRT)	14	1119	1.25	8	1123	0.89	10	1004	1.00	18	1004	1.79
SCHOLY3	(SCHOL1)	7	203	3.45	0	175	0.00	1	119	0.84	0	112	0.00
SCHOLY	(SCHOL2)	0	140	0.00	2	102	1.96	0	63	0.00	0	36	0.00
REMITY	(REMINC)	4	366	1.09	11	457	2.19	4	346	1.16	3	406	0.74
OTHERY	(OTHRINC)	6	327	1.83	5	294	1.70	3	277	1.08	4	256	1.56
FOODE2W	(FOODEXP1)	8	1566	0.51	9	1592	0.57	10	1581	0.63	10	1596	0.63
FOODEAN	(FOODEXP2)	11	1569	0.71	12	1598	0.75	11	1595	0.69	12	1599	0.75
FARMHC	(HOMECON1)	4	1001	0.40	3	1001	0.30	3	838	0.38	1	912	0.11
NFHC	(HOMECON2)	7	241	2.90	2	243	0.82	2	180	1.11	2	189	1.06
OTHERE	(OTHCONXP) [1]	14	1588	0.82	11	1600	0.69	10	1600	0.63	15	1600	0.94
REMITE	(XPREMIT)	10	901	1.11	9	870	1.03	8	669	1.20	12	616	1.94
WGKY7	(TOTEMPK1)	2	270	0.74	2	258	0.78	3	283	1.06	5	238	2.10
WGKY12	(TOTEMPK2)	4	315	1.27	3	287	1.05	2	306	0.65	5	259	1.93

NOTE: 1) The IIEEA variable, OTHERE, contains an important correction to the IIEEC variable OTHCONXP. OTHERE does not include the sub-aggregate level variable EXPDAY as this variable has already been included in the definition of FOODEXP1 and FOODEXP2.

The Visual Identification of Outliers

Following the completion of the statistical identification of outliers, the data was further checked for outlying values by visual inspection. The visual identification of outliers began by plotting the per-capita values of each aggregate level variable against per capita household income, if the aggregate was a component of Total Income, or against per capita household expenditure, if the aggregate level variable was a component of Total Expenditure. Any observation lying unequivocally away from the clustered mass of observations was identified as a potential outlier. The household or individual level characteristics associated with this (possibly) outlying value, (such as the household head's education, the head's main job and its type of industry, and above all related income and expenditure variables), were identified so as to assess the plausibility of the observation.

Comparison of variable values across similar households provided further information with which to make the subjective decisions required. For example, when a potential outlier for the variable FARMDEP (Depreciation of Farm Equipment) was discovered, the value was compared with that of another household with similar farming income (FARMY) characteristics. The suspected case was then confirmed to be an outlier by noting that the FARMDEP value continued to seem unreasonable, despite the availability of the type of information which could explain the abnormality of the observation's value. Table 2.2 documents the number of cases in which suspected outliers were imputed with replacement values. Replacement values (at the per-capita level) were derived by obtaining the per-capita regional mean of the variable in question.

Figure 2.2 Documentation on the number of outliers identified through the visual inspection method

YEAR	Number of suspected outliers / Number of variables with suspected outliers	Number of suspected outliers imputed with replacement values and as a percent of all households surveyed each year
1985	54 / 11	13 (0.82 %)
1986	22 / 11	3 (0.19)
1987	24 / 8	6 (0.38)
1988	40 / 13	11 (0.69)

Treatment of missing observations

The data sets from the field, at the questionnaire level, contain missing values. Such missing values may arise from non-response, i.e. respondent is unable or unwilling to answer a particular question (a genuine "Missing Value"); or because a question is not-applicable and was skipped in the questionnaire ("Not-Applicable"). In this research project, the genuine missing data values were imputed with the value of the regional mean of the variable in question. In principle, this imputation was necessary because such missing values represent positive and possibly important components of total income or expenditure and as such, their values should certainly not be assumed to be zero.

IV. CORRECTIVE WEIGHTS

Further information on the distribution of the sampled population sometimes becomes available only after field interviews have been conducted. For example, comparison of CILSS results with Cote d'Ivoire's 1988 Population Census, showed that the distribution of population by region derived from the CILSS did not match the Census figures. These problems were traced to sampling defects in the CILSS, thus making it necessary to rectify the flawed distribution with corrective weights.

Three different sets of corrective weights were used in the analysis of the CILSS data. Before discussing these in detail, however, it is important to note that the survey sampling plan and listing operations changed after the first two years, i.e. in 1987. We are left, therefore, with two conceptually distinct blocks of data; 1985, 1986 and one-half (the 1986-87 panel component) of 1987 which is called Block 1 data in this paper, and, the data from the remainder of households sampled in 1987 which along with the 1988 data is referred to here as Block 2 data. Each of the corrective weights is applied to one or the other of the two blocks but never uniformly across all the data encompassing all four years.

Weights were needed to correct for three factors: i) the over-sampling of rich households in Abidjan for Block 1 data ii) the over-sampling of urban households in Block 2 data and iii) the probable over-sampling of larger sized households in Block 1 data. The rationale and methodology for each of these weights is described in detail in this section. Data Processing documentation is available at the end of this Chapter in Appendix 3.1. For details on the application of the weights, see Grootaert (1992) and Demery and Grootaert (1992).

The Abidjan Weights

A selection error at the Primary Sampling Unit (PSU) stage for Block 1 data in Abidjan resulted in the over-sampling of richer households in that city. Since the error occurred at the level of the PSU i.e. in the selection of Clusters, the corrective weights are applied at the level of the Cluster. [3]

Cluster	Household Numbers	1985 Weights	1986 and Panel 1987 Weights
1	31,61	0.895	0.840
1	All households except 31 and 61	0.895	0.894
2 to 5	All households	0.899	0.894
6	All households	0.899	0.894
7,8	All households	0.895	0.840
9-11	All households	0.329	0.371
12	All households	0.899	0.894
13-17	All households	4.182	4.498
18-21	All households	0.895	0.840
22-100	All households	0.899	0.894

[3] The weights were computed by C. Scott, statistical consultant, following a review of Block 1 procedures.

The Regional Weights

As mentioned earlier, the sampling plan for Block 2 data was updated on the basis of an electoral survey that took place in 1986. Results of a new Population Census conducted in 1988 revealed, however, that the 1987-88 CILSS had over-sampled the population in urban areas. Given the Population Census' distribution of population by region, it was possible to construct corrective regional weights for Block 2 data, as follows.

	Abidjan	Other Cities	Rural
Percent of Population in each region, derived from the Population Census results	18	21	61
Distribution of Population, by region, CILSS 1988	19	27	54
Distribution of Population, by region, for second-half of CILSS 1987	11	16	23

Thus, the regional weights for 1988 are :

Abidjan : 18/19, Other Cities : 21/27, Rural : 61/54

For the second-half of 1987 data,

Abidjan : 9/11, Other Cities : 10.5/16 Rural : 30.5/23

Since the selection error corresponding to these weights occurred at the level of region, the regional weights were applied to households based on the household's region of residence.

The Household Size Weights

Since the design of these weights is completely dependent on the way in which the sampling plan and operations changed in 1987, it is important to first clarify the relevant details. To save time during the initial household listing operation which took place in 1984 in preparation for the 1985 survey, only 64 households per cluster were listed. Results of this listing operation were utilized for CILSS 1985 and 1986 until an electoral census conducted in 1986 provided an updated count of the population. The 1986 Electoral Census results were used as the basis for a new sample frame for the 1987 and the 1988 CILSS surveys. The 1987 survey, however, still partially depended upon the old sampling frame and listing procedures because one-half of the households interviewed in 1986 had to be retained in 1987 as panel households. The net effect for the 1987 sampling frame was that 50 clusters in the frame were derived from the 1984 listing procedure, while the rest of the clusters were derived from the new frame. By 1988 the old sampling frame had been completely phased out; the panel households in 1988 were those selected from the new frame for 1987, and the replacement clusters were also obtained from the new (1987) frame.

During discussions which took place in the CILSS Data Workshop at the World bank, in July 1991 (see Venkataraman, 1991, for more on this), it became clear that the puzzling and dramatic decrease in mean household size from 7.96 in 1985, 7.66 in 1986, 6.79 in 1987 to 6.15 in 1988, (these figures are corrected for the over-sampling of rich households in Abidjan) could most rationally be explained by the change in the listing procedures from 1987 onwards. As mentioned earlier, the listing process performed in 1984

randomly selected 64 households in each PSU to be listed. From among these 64, sixteen households were then selected to be interviewed. The listing procedure for CILSS 1987 and 1988 included all households in each PSU. The manner in which the earlier listing process was conducted gives room to the possibility that the more conspicuous and larger household buildings tended to be chosen rather than the less noticeable and smaller dwellings. Since larger dwellings in Africa tend to be associated with larger families, this listing process was likely to be biased toward selecting larger households. This was thought to be the most likely explanation for the drop in household size over the survey years. The Household Size weights described in this Chapter are designed to correct for biases resulting from changes in the listing operations. To see how application of the household size weights reduces the drop in household size over the years, refer to Demery and Grootaert (1992).

Methodology

Household size groups were computed for Block 1 and Block 2 data, by region. All households in a particular region, having a size of 1 were placed in Size Group 1, households with a size of two were placed in Size Group 2 and so on until Size Group 17. Size Group 18 contains all households within the region in question, with a household size of 18 or more. The frequency distribution of Size Group was obtained for Block 1 and Block 2 data. (Block 1 households in Abidjan were weighted with the Abidjan weights and Block 2 data was weighted with the Regional weights). In each region and for a particular Size Group, household size weights were then calculated by dividing the percentage of Block 2 households in that Size Group, by the percentage of Block 1 households in the same Size Group. The tables in the next few pages illustrate the process.

Since this particular selection error occurred in the choice of households within the clusters, the corrective weights are applied at the level of the household. The Household Size weights are applied to all Block 1 households - the weights differ, of course, by Region and by Size Group.

HOUSEHOLD SIZE WEIGHTS FOR ABIDJAN

	Percent of Block households (1)	Percent of Block 2 households (2)	Household Size Weight (2) / (1)
Household Size: 1	11.4159	11.1144	0.97359
Household Size: 2	7.8969	6.9079	0.87476
Household Size: 3	8.0458	8.5128	1.05804
Household Size: 4	6.5688	11.4035	1.73600
Household Size: 5	10.4073	13.1878	1.26717
Household Size: 6	11.9835	10.2472	0.85511
Household Size: 7	8.3376	10.3270	1.23860
Household Size: 8	6.7984	6.7185	0.98824
Household Size: 9	8.1710	5.6220	0.68805
Household Size: 10	4.8717	5.7815	1.18675
Household Size: 11	2.3523	4.3361	1.84335
Household Size: 12	2.9287	1.0367	0.35398
Household Size: 13	3.0888	1.2560	0.40662
Household Size: 14	2.1440	1.1962	0.55793
Household Size: 15	1.4052	0.4087	0.29085
Household Size: 16	0.3097	1.0965	3.54060
Household Size: 17	1.1361	0.2193	0.19303
Household Size: 18	2.1383	0.6280	0.29369

HOUSEHOLD SIZE WEIGHTS FOR OTHER CITIES

	Percent of Block households (1)	Percent of Block 2 households (2)	Household Size Weight (2) / (1)
Household Size: 1	5.4054	11.0291	2.04039
Household Size: 2	5.8559	8.3623	1.42802
Household Size: 3	7.6577	7.8752	1.02841
Household Size: 4	6.9820	10.8507	1.55410
Household Size: 5	6.6441	10.3877	1.56344
Household Size: 6	8.7838	11.9792	1.36378
Household Size: 7	9.5721	6.6069	0.69022
Household Size: 8	8.8964	8.3140	0.93454
Household Size: 9	7.5450	5.5604	0.73696
Household Size: 10	5.1802	3.7471	0.72335
Household Size: 11	4.0541	3.7712	0.93023
Household Size: 12	3.6036	2.3245	0.64504
Household Size: 13	3.6036	2.6331	0.73063
Household Size: 14	2.9279	2.0882	0.71319
Household Size: 15	3.1532	1.2683	0.40224
Household Size: 16	3.0405	0.7475	0.24584
Household Size: 17	0.7883	0.3086	0.39153
Household Size: 18	6.3063	2.1460	0.34030

HOUSEHOLD SIZE WEIGHTS FOR EAST FOREST

	Percent of Block households (1)	Percent of Block 2 households (2)	Household Size Weight (2) / (1)
Household Size: 1	3.8589	6.6106	1.71308
Household Size: 2	4.8512	7.8495	1.61806
Household Size: 3	7.3870	9.6616	1.30792
Household Size: 4	9.4818	10.0407	1.05894
Household Size: 5	8.4895	11.8251	1.39290
Household Size: 6	9.9228	11.9360	1.20289
Household Size: 7	8.4895	7.9512	0.93659
Household Size: 8	8.2690	8.8757	1.07337
Household Size: 9	7.0562	6.2777	0.88967
Household Size: 10	5.2922	4.7245	0.89273
Household Size: 11	4.7409	3.7629	0.79372
Household Size: 12	3.6384	2.2744	0.62512
Household Size: 13	4.4101	2.4871	0.56394
Household Size: 14	2.9768	0.6379	0.21430
Household Size: 15	1.7641	1.4238	0.80713
Household Size: 16	2.6461	1.1372	0.42977
Household Size: 17	2.0948	0.6379	0.30453
Household Size: 18	4.6307	1.8861	0.40731

HOUSEHOLD SIZE WEIGHTS FOR WEST FOREST

	Percent of Block households (1)	Percent of Block 2 households (2)	Household Size Weight (2) / (1)
Household Size: 1	2.5000	5.2995	2.11979
Household Size: 2	6.0714	9.5964	1.58058
Household Size: 3	6.0714	12.7474	2.09957
Household Size: 4	11.0714	14.2448	1.28663
Household Size: 5	9.1071	9.3490	1.02655
Household Size: 6	10.5357	16.0547	1.52383
Household Size: 7	8.3929	8.5026	1.01308
Household Size: 8	8.9286	5.0521	0.56583
Household Size: 9	10.0000	5.5469	0.55469
Household Size: 10	7.3214	3.8021	0.51931
Household Size: 11	5.7143	3.0469	0.53320
Household Size: 12	2.6786	1.4063	0.52500
Household Size: 13	2.3214	2.7995	1.20593
Household Size: 14	1.0714	1.0026	0.93576
Household Size: 15	1.7857	0.6510	0.36458
Household Size: 16	1.4286	0.2995	0.20964
Household Size: 17	1.0714	0.2995	0.27951
Household Size: 18	3.9286	0.2995	0.07623

HOUSEHOLD SIZE WEIGHTS FOR SAVANNAH

	Percent of Block households (1)	Percent of Block 2 households (2)	Household Size Weight (2) / (1)
Household Size: 1	5.4847	8.9728	1.63597
Household Size: 2	5.6122	10.9707	1.95477
Household Size: 3	9.1837	11.1386	1.21287
Household Size: 4	8.6735	11.6513	1.34333
Household Size: 5	11.4796	10.1574	0.88482
Household Size: 6	10.4592	8.0799	0.77252
Household Size: 7	7.3980	6.8600	0.92728
Household Size: 8	6.6327	6.3119	0.95164
Household Size: 9	7.1429	4.5173	0.63243
Household Size: 10	6.3776	4.8621	0.76238
Household Size: 11	4.4643	3.6687	0.82178
Household Size: 12	3.1888	2.0140	0.65980
Household Size: 13	3.4439	1.3260	0.38504
Household Size: 14	2.9337	1.9714	0.67198
Household Size: 15	2.1684	1.9360	0.89284
Household Size: 16	1.0204	1.0873	1.06559
Household Size: 17	1.1480	1.0873	0.94719
Household Size: 18	3.1888	3.2974	1.03406

ALLWAITN - The three weights combined

In order to facilitate the usage of the weights, the three weights are combined into one single normalized weight, called ALLWAITN, designed to be usable both for regional and national level analysis.

Normalization of weights is achieved by multiplying each weight by an appropriate normalization factor. The normalization factor is equal to N/(sum of weights); where N = the number of observations over which the normalization exercise is to take place. (If weights are to be normalized over a particular region, then N should equal the number of observations in that region whereas, if weights are to be normalized over the whole sample then N should equal the total number of households in the entire sample). The main outcome of this exercise is that it allows the degrees of freedom in the region/sample to remain the same as that found in the unweighted data. Normalization of weights ensures that the statistics measuring dispersion - variance and standard deviation, remain unbiased.

The manner in which the variable ALLWAITN was created, is as follows. First, the Abidjan weights and the Household size weights were multiplied together, and then normalized by region. (Where neither of these weights apply, as in the Block 2 data, the default value of the weight is 1). The resulting variable is called CSSIZWTN. The Regional weights being self-normalized by nature, we were now left with fully normalized weights for each year. Conceptually, no further manipulations would have been necessary at this stage. However, in order to create a single weight variable that could be used more convieniently at the analytic stage, we performed a final operation as follows. CSSIZWTN was multiplied into the Regional weights (again, the default value of the weight is equal to 1 if its application is not relevant to that Block of data) to yield the final variable ALLWAITN.

Contents of the CILSS Weights dataset

FILENAME : ALLWAITS FILETYPE: CI85, CILS2, CI87, CI88.

Variable Name	Description
CSWAITN	Normalized Abidjan Weights
PSUWAITN	Self-normalizing Regional Weights
SIZWAITN	Normalized Household Size Weights
CSSIZWTN	Normalized combination of the Abidjan weights with the Household Size weights
ALLWAITN	The three weights combined

V. REGIONAL AND TEMPORAL PRICE INDEX

Construction of Regional and Temporal Price Indices was motivated by the requirement that all monetized variables which were to be compared over the survey years, be expressed in constant prices and that regional price differences be taken into account. Earlier attempts at constructing regional price indices utilizing CILSS price data were beset by several limitations, the most troublesome of which was that CILSS price data is available for only a limited number of commodities. The limitations are particularly severe in the area of non-food commodities for which price information is available for only four items. Moreover, price data for 1987 and 1988 is only available for Abidjan and a few towns, not for rural areas.

Given all these considerations, it was clearly preferable to use the 1985 International Comparisons Project (ICP) price data for Cote d'Ivoire in which a wide variety of food and non-food items are represented. The ICP price data set contains 20902 prices for 971 products. Prices from Abidjan, Other Urban areas and the rural areas of East and West Forest and Savannah is available in the ICP data, thus enabling the required regional link with CILSS data. With the information obtained from the ICP data, reputed to be of higher quality than the CILSS price data, it was possible to obtain reliable estimates of regional price variation for at least one year and then to apply inflation rates to this set of price indices to obtain a price index over time and space. The resulting price index (presented in Table 5.3), obtained by multiplying the regional price index (presented in Table 5.1) with the temporal index, i.e. the CPI for 1985-1988 (presented in Table 5.2), is a single variable that can convieniently be used as a deflator in the later analytic stages. For more on this topic, see Grootaert and Kanbur (1992).

TABLE 5.1 : Regional Price Index, 1985-88

	1985	1986	1987	1988
Abidjan	100.0	100.00	100.00	100.00
Other Cities	92.84	93.62	91.49	92.57
East Forest	87.01	87.01	88.12	86.58
West Forest	78.25	74.66	75.64	72.42
Savannah	75.97	80.12	81.86	81.88

TABLE 5.2 : CPI, 1985-88

	1985	1986	1987	1988
CPI	100.00	107.30	107.75	115.31

Source: IMF, International Financial Statistics

Table 5.3: Price Index

	1985	1986	1987	1988
Abidjan	100.00	107.30	107.75	115.31
Other Cities	92.84	100.45	98.58	106.74
East Forest	87.01	93.36	94.95	99.84
West Forest	78.25	80.11	81.50	83.51
Savannah	75.97	85.97	88.20	94.42

Regional Price Index: Conceptual Background

The Regional Price Index was constructed as a Paasche Price Index.

Formula 1

$$C_r = \frac{\Sigma (p_{ri} * q_{ri})}{\Sigma (p_{ai} * q_{ri})}$$

where C_{ri} is the cost of living deflator for region r, p_{ri} is the price of item i in region r, and p_{ai} is the price of item i in the reference region, in this case Abidjan.

Formula 1 can be reduced to :

Formula 2

$$\frac{1}{\Sigma (S_{ri} * (p_{ai}/p_{ri}))}$$

where S_{ri} is the expenditure share of item i in the household's total budget.

This implies that all items to be included in the price index must have both expenditure shares and price relatives values available.

Methodology

The main steps involved in the calculation of the regional price index are as follows. First, expenditure items listed in the CILSS questionnaire were assembled into categories which could later be matched with the product codes in the ICP data. This categorization is presented in Appendix 5.1 (pp. 49-50) . In Appendix 5.2 (pp. 51-59), we present the outcome of the exercise which matched the CILSS expenditure categories with the ICP product codes.

The next step was to calculate expenditure shares for each of the selected categories listed in Appendix 5.1. This was done using the following Sub-Aggregate and Variable level expenditure files: BEC16, BEC17, BEC19, BEC20, IEC36. Expenditure shares for each region were computed by summing up weighted (using ALLWAITN) expenditures for each category (over all the households in the region) and dividing this sum by the sum of expenditures (weighted, using ALLWAITN) on all expenditure categories included in the price index (over all households in the region). If V_{ri} is the value of expenditure on item i in region r, then S_{ri}, expenditure share of item i in region r, can be formulated as:

$$S_{ri} = V_{ri} / \Sigma_i V_{ri}$$

Calculation of price relatives, using ICP data, followed next. First, the ICP data had to be 'cleaned' of observations with clearly erroneous data, inconsistent units of measurement and extreme values. Price relatives (i.e. Abidjan price divided by the price of the item in the region in question) were then calculated for each item in all categories. Next, the average of the price relatives over all items in each commodity group, was computed. Even if the price relative for a particular item within a particular category was missing, it was still possible to obtain the mean price relative for that region and category, as long as other items within the category had prices available. The problem arose only when prices for all items within a category were missing. In such rare cases, the missing price relative for that category, was replaced with the mean price relative of the same category in an analogous (usually, the corresponding urban) region. In the case of rural Savannah, a number of missing price values for non-food items were imputed with corresponding prices from urban Savannah on the grounds that rural Savannah inhabitants would, in any case, probably have to travel to the closest urban area in order to obtain such items. Table 5.1, below, documents the few cases which involved imputations requiring extra judgement.

TABLE 5.1

Product Category	Action
101	1. Other Urban Price Ratio = 1 2. Ratio in Rural East Forest = Ratio in Rural West Forest
106	1. Rural East Forest Ratios = Rural West Forest Ratios
132	Rural West Forest = Rural East Forest ratios
122,134	All Rural Price ratios = Other Urban price ratios
123,124,136,137	The Mean of prices from Other Urban and rural areas was used in computing this ratio

The next step in the process was to bring both the expenditure shares and price relatives together to calculate the price index as specified in Formula 2. Within each region, for each category, the price relative and expenditure share figures were multiplied together. This product was then summed over all categories used in the price index. The final outcome is the denominator of the price index formula specified in Formula 2.

APPENDIX 5.1

Categorization of Expenditure Items listed in Sections 11a and 11b of the CILSS Questionnaire

CILSS codes	Category description
101,331	FOOD & BEVERAGES CONSUMED AWAY FROM HOME
102	CIGARETTES, TOBACCO & COLA NUTS
103	COMMERCIAL OR HOME-MADE SOAP
104	OTHER PERSONAL CARE & HEALTH PRODUCTS
105	HOME MAINTENANCE PRODUCTS (BROOMS,TOILET PAPER ETC)
106	CHARCOAL
107	WOOD
108	OTHER FUEL FOR COOKING,LIGHTING
109,123	VEHICLE PARTS
115,116	SHOES
117/118/119	DOMESTIC/IMPORTED PAGNES,FABRIC FOR ADULT AND CHILDREN'S CLOTHING
120	ADULT CLOTHING
121	CHILDREN'S CLOTHING
122	PURCHASE OF CARS,BIKES & OTHER TRANSPORT
124	PUBLIC TRANSPORT, TAXIS ETC
125	HOME REPAIRS, PAINTING, INSURANCE ETC
128,129	MEDICAL SERVICES, DOCTOR, HEALER, OTHER MEDICAL EXPENSES
130	KITCHEN EQUIPMENT (CUPS,FORKS,PLATES,SAUCEPANS ETC)
131	FURNITURE (BEDS,TABLES,CUPBOARDS,CHAIRS,RUGS ETC)
132	LINEN (SHEETS, TOWELS, BLANKETS ETC)
133	ENVELOPES, WRITING PAPER, STAMPS
134	TELEPHONE, TELEGRAM, ETC
136	JEWELRY,WATCHES
137	ENTERTAINMENT (NOVELS,NEWSPAPERS,CINEMA,SPORTS,RECORDS,TAPES,TOYS)

Appendix 5.1 (Contd.)

CILSS codes	Category description
301	RICE
302	MAIZE (COB,GRAIN OR FLOUR)
303	MILLET, FONIO, SORGHUM, GRAIN OR FLOUR)
304	BREAD
305/306	CASSAVA AND ATTIEKE
307	MACARONI
308	COOKIES AND CAKE
309	YAM
310	PLANTAIN (RAW OR FLOUR)
311	TARO, SWEET POTATO, POTATO
312/313	OIL PALM NUTS AND PEANUTS (ROASTED, RAW OR BUTTER)
314	OTHER SEEDS (AVOCADO, COCONUT, NERE, OULEOULE,PUMPKIN) FRESH/DRIED
315	FISH AND SHELLFISH
316/317	GAME, CHICKEN, DUCK, PIGEON, TURKEY OR OTHER POULTRY
318	BEEF, MUTTON, GOAT, PORK, OTHER DOMESTICATED MEAT
319	CHICKEN EGGS
320/321	PALM OIL AND SHEA BUTTER AND REFINED OIL
322	BUTTER, MARGARINE
323	FRUIT (ORANGES,MANGOES, MANDARINS, PAPAYAS ETC)
324	SUGAR, CANDIES, HONEY, SUGARCANE
325	SALT
326	ALCOHOLIC BEVERAGES
327/328	NON-ALCOHOLIC BEVERAGES (TEA, COFFEE, SOFT DRINKS ETC), BOUILLION CUBES
329	TOMATO PASTE
330	LEAFY AND OTHER VEGETABLES
333	MILK PRODUCTS
334	OTHER FOODS

Note: EXPENDITURE ON EDUCATION (CILSS codes 126, 127) was obtained from CILSS Section 3.

Appendix 5.2

Matching ICP Product Categories with CILSS Expenditure Categories

CILSS CODE:101,331 Food and Beverages Consumed Away From Home

1830111B0 Meal local restaurant
1830111B1 Meal food kiosk/hawker
1830111C1 Ricard in a bar
1830111E0 Beer in a cafe
1830121A0 Breakfast in worker's cafeteria

CILSS CODE:102 Cigarettes, Tobacco & Cola Nuts

1130111A0 Cigarettes dark tobacco
1130111F0 Cigarettes light tobacco Marlboro
1130211E0 Kola nut

CILSS CODE:103 Commercial or Home-Made Soap

1450121A0 Household soap 72%
1450121B0 Household soap 72% 650 800 g
1450121C0 Household soap 60%

CILSS CODE:104 Other Personal Care & Health Products

1810111B0 Hairdresser w/o own establishment
1820111C0 Toothbrush
1820111D0 European comb
1820111O0 Razor blades

CILSS CODE:105 Home Maintenance Products (Brooms, Toilet Paper Etc.)

1450111A0 Toilet tissue 650 to 1000 sheets
1450121G1 Washing powder packet 100g
1450121H0 Bleach javel lacroix

CILSS CODE:106 Charcoal

1320511A0 Domestic charcoal

CILSS CODE:107 Wood

1320411A0 Firewood

CILSS CODE:108 Other Fuel For Cooking, Lighting

1320311A0 Paraffin

CILSS CODE:115,116

1220111I0 Men's boots
1220121D0 Ladies' plastic shoes

CILSS CODE:117 Domestic/Imported Pagnes, Fabric for Adult and Children's Clothing

1210111A0 Fabric wax print
1210111C0 Fabric wax block monochrome

CILSS CODE:120 Adult Clothing

1210211L0 Men's brief 100% cotton
1210211M0 Men's brief 2/3 polyamide 1/3 cotton
1210221F0 Women's brief 100% cotton
1210221I0 Brassiere
1210311B0 Men's handkerchief
1210311C0 Women's handkerchief

CILSS CODE:121 Children's Clothing

1210231A0 Boys' jeans
1210231E0 Boys' short
1210231H0 Boys' t shirt monochrome
1210231I0 Boys' t shirt printed
1210311A0 Zip fastener

CILSS CODE:122 Purchase of Cars, Bikes & Other Transport

1610111A0 Renault R4 L
1610111B0 Renault R5 GTL
1610111C0 Renault R9 GTL
1610111C1 Renault R9 GTC
1610111D0 Mazda 323
1610111E0 Peugeot 305 normal
1610111E1 Peugeot 305 GR
1610111F0 Peugeot 505 GL
1610111F1 Peugeot 505 GL diesel
1610111G0 Peugeot 504 normal
1610111H0 Toyota Corolla 1300 L
1610111I0 Toyota Corolla GL
1610111J0 Toyota Corolla GLS
1610111K0 Toyota Corolla 2000 GLE
1610111L0 Toyota Corolla 1800 XL
1610111M0 Nissan Sunny 1 3 std

1610111N0 Nissan Sunny 1 3 DX
1610111S0 Fiat Panda 45
1610111U0 Fiat Uno 55
1610211A0 Bicycle ladies's town
1610211B0 Bicycle men's town Raleigh
1610211C0 Bicycle men's town other makes
1610211D0 Motor cycle Peugeot 153 LSX
1610211D1 Motor cycle Peugeot 103 SPB
1610211E0 Motor cycle Motoconfort
1610211F0 Motor cycle Yamaha
1610211G0 Motor cycle Suzuki
1610211H0 Motor cycle Honda 185 S
1610211I0 Motor cycle Honda 125 S
1610211J0 Motor cycle Honda CG 125
1610211L0 Motor cycle Suzuki A 100
1610211M0 Motor cycle Suzuki TS 125
1610211N0 Motor cycle Yamaha YB 100

CILSS CODE:109,123 Vehicle Parts

1620111A0 Motor bicycle tyre Peugeot
1620111A1 Bicycle tyre size 26
1620111A2 Bicycle tube
1620111B0 Motor bicycle tyre Suzuki
1620111C0 Car tyre Michelin 165 SR 13
1620111C1 Car tyre 175 SR 14
1620111C2 Car tyre 175 SR 14 retread
1620111D0 Car tyre 165 SR 13
1620111D1 Car tyre 165 SR 13 retread
1620111D2 Car tyre 155 SR 14
1620111D3 Car tyre 155 SR 14 retread
1620111D4 Car tyre 155 SR 13
1620111D5 Car tyre 155 SR 13 retread
1620111D6 Motor cycle tyre 2.75
1620111D7 Motor cycle tyre 2.60
1620111D8 Motor cycle tyre 3.50
1620111D9 Motor cycle tyre 4.10
1620111E0 Battery locally manufactured 30ah
1620111F0 Battery locally manufactured 45ah
1620111G0 Imported battery 38ah
1620111G1 Imported battery 44a.m
1620111H0 Sparking plug Champion
1620111I0 Sparking plug Bosch
1620111J0 Brake cable
1620111K0 Tyre tube 175 SR 14
1620111K1 Tyre tube 165 SR 13
1620111K2 Tyre tube 155 SR 14
1620111K3 Tyre tube 155 sr 13
1620111L0 Distributor points r4 l
1620111L1 Distributor points r9 gtl
1620111L2 Distributor points mazda 323
1620211A0 Oil change and greasing official dealer
1620211B0 Oil change and greasing not official dea

1620211C0 Replacement water pump official dealer
1620211D0 Replacement water pump not official deal
1620211E0 Replacement shock absorbers official dea
1620211F0 Replacement shock absorbers not official
1620211G0 Engine tuning official dealer
1620211H0 Engine tuning not official dealer
1620211I0 Replacement brake linings/pads official
1620211J0 Replacement brake linings/pads not official

CILSS CODE:124 Public Transport, Taxis, Etc.

1630111A0 Taxi journey
1630111B0 Collective taxi journey
1630121A0 Bus journey minimum fare
1630121B0 Bus journey maximum fare
1630121C0 Bus fare student
1630121C1 Bus fare adult
1630131A0 Push cart charges
1630131B0 Informal sector journey 4 km
1630131C0 Informal sector journey 10km
1630211A0 Train journey 2nd class
1630211B0 Rail journey express
1630211B1 Train journey round trip
1630221A0 Bus journey
1630221A1 Journey by motorcoach
1630221B0 Journey by car or station wagon
1630221C0 Journey by mini bus

CILSS CODE:125 Home Repair

1310211B0 Portland cement
1310221A0 Painting of room w/ whitewash
1310221C0 Repair of a house roof
1410311B0 Replacement of top formica

CILSS CODE:127

1740111A0 School fees nursery
1740111B0 School fees Koran
1740111C0 School fees private vocational
1740111D0 School fees private primary
1740211A0 School fees private secondary
1740411D0 Exercise book
1740411E0 Pencil ordinary
1740411H0 Ruler

CILSS CODE:128,129

1510111C0 Aspirin upsa

1510121B0 Flavoquine 12 tablets
1510121C0 Nivaquine
1510121J0 T A O
1510211B0 Mercurochrome bottle 20 30 ml
1510211C0 Mercurochrome bottle 125 ml

CILSS CODE:130 Kitchen Equipment

1440111A0 Drinking glass
1440111B0 Tumblers
1440111E0 Soup plate
1440111G0 Saucepan
1440111M0 Knife
1440111O0 Soup spoon

CILSS CODE:131 Furniture (Beds, Tables, Cupboards, Chairs, Rugs, etc.)

1410111D0 Mattress 190x140 low density
1410111E0 Mattress without springs
1410111H0 Ordinary chair

CILSS CODE:132 Linen (Sheets, Towels, Blankets, etc.)

1420111A0 Bedsheet polyester
1420111B0 Bedsheet 100% cotton
1420111C0 Hand towel
1420111D0 Bath towel

CILSS CODE:133 Envelopes, Writing Paper, Stamps

1820311A0 Ordinary envelopes
1820311B0 Notepaper pad
1820311C0 Ballpoint pen

CILSS CODE:134 Telephone, Telegram, etc.

1640211A0 Telephone call public telephone box
1640211B0 Telephone call telephone subscriber
1640211C0 Monthly rental of one telephone
1640211D0 Telephone calls
1640221A0 Telegram

CILSS CODE:136 Jewelry, Watches

1820211A0 Gold ring 18 carat
1820211B0 Gold ring 14 carat
1820211C0 Men's wrist watch
1820211D0 Men's digital wrist watch
1820211E0 Watch maintenance

CILSS CODE:137 Entertainment (Novels, Newspapers, Cinema, Sports, Records, Tapers, Toys

1710311A0 Gramophone record
1710311A1 Gramophone record pop music
1710311B0 Unrecorded cassette 60 min
1710311B1 Unrecorded cassette 90 min
1710311C0 Unrecorded cassette other brands
1710311D0 Tennis balls
1710311E0 Building set ref n0 10
1710311E1 Building set ref n0 045
1710311F0 Bag of marbles
1710311G0 Children's ball
1710311H0 Film black and white
1710311I0 Film color 20 exposures
1710311I1 Film color 36 exposures
1710311J0 Slides
1710411A0 Services color tv technician
1720111A0 Cinema
1720111B0 Sports ground
1720211C0 Photographic developing and printing
1730111A0 Novel detective
1730111B0 Novel not detective
1730111C0 Newspaper
1730111D0 Magazine weekly
1730111E0 Magazine monthly

CILSS CODE:301 Rice

1B0 Long grained rice loose
1F0 100% broken rice

CILSS CODE:302 Maize (Cob, Grain, or Flour)

1110121E0 Yellow maize
1110121F0 White maize

CILSS CODE:303 Millet, Fonio, Sorghum, Grain or Flour)

1110121B0 Sorghum
1110121C0 Small millet souna variety
1110121D0 Small millet sagno variety

CILSS CODE:304 Bread

1110131A0 Fresh bread
1110131B0 Fresh bread baguette t

CILSS CODE:305,306 Cassava

1110821A0 Cassava fresh
1110821B0 Cassava dried

CILSS CODE:307 Macraroni

1110161A0 Spaghetti prepacked 500 g
1110161B0 Spaghetti prepacked 250 g
1110161C0 Egg noodles 500 g
1110161E0 Macaroni

CILSS CODE:308 Cookies and Cake

1110141H0 Biscuits sweet tea
1110141I0 Biscuits marie
1110141J0 Swiss roll

CILSS CODE:309 Yam

1110821D0 Yams fresh

CILSS CODE:310 Plantain (Raw or Flour)

1110611N0 Plantain
1110611O0 Plantain green

CILSS CODE:311 Taro, Sweet Potato, Potato

1110821E0 Taro (cocoyam)
1110821F0 Sweet potatoes fresh

CILSS CODE:312,313 Oil Palm Nuts and Peanuts (Roaster, Raw or Butter)

1110521M0 Peanut butter
1110631P0 Ground nuts dry roasted shelled

CILSS CODE:314 Other Seeds (Avocado, Coconut, Nere, Ouleoule, Pumpkin)

1110611G0 Avocado
1110611I1 Coconut without topshell

CILSS CODE:315 Fish and Shellfish

1110321B0 Crabs
1110331A0 Herring

CILSS CODE:316,317 Game, Chicken, Duck, Pigeon, Turkey or other Poultry

1110241A0 Live local chicken

CILSS CODE:318 Beef, Mutton, Goat, Pork, Other Domesticated Meat

1110211B0 Beef fresh w/ bone w/ offal
1110211C0 Beef fresh w/ bone no offal
1110251B0 Ox feet fresh
1110251D0 Beef tripes and offals

CILSS CODE:319 Chicken Eggs

1110451B0 Chicken eggs commercial fresh

CILSS CODE:320,321 Palm Oil and Shea Butter and Refined Oil

1110521H0 Palm oil unrefined
1110521H1 Palm oil refined loose
1110521L0 Shea nut butter

CILSS CODE:322 Butter, Margarine

2321110521K0 Margarine

CILSS CODE:323 Fruit (Oranges, Mangoes, Madarins, Papayas, etc.)

2331110611E0 Lemon
2341110611H0 Bananas
2351110611J0 Mango not grafted

CILSS CODE:324 Sugar, Candies, Honey, Sugarcane

1110911B0 Sugar cubes cardboard packet
1110911C0 Sugar cubes 10 lumps
1110911D0 Sugar refined white granulated cellophane
1111121C0 Acid fruit drops prepacked
1111121D0 Caramel sweets prepacked
1111121E0 Chewing gum Wrigleys
1111121F0 Chewing gum other makes

CILSS CODE:325 Salt

1111131A0 Salt coarse kitchen

CILSS CODE:326 Alcoholic Beverages

1120221D0 Wine red bottle 1 l
1120231C0 Beer Guinness
1120231G0 Beer local other

CILSS CODE:327,328 Non-alcoholic Beverages (Tea, Coffee, Soft Drinks, etc.), Bouillon

1110631Q0 Ginger Juice
1111011B0 Coffee beans roasted robusta
1111011F0 Coffee soluble instant
1111021C0 Tea black prepacked 20 bags
1111021D0 Tea black prepacked 100 bags
1120111A0 Natural mineral water
1120111B0 Soda water
1120121F0 Fanta orange bottle 33cl

CILSS CODE:329 Tomato Paste

1110721H0 Tomato puree 70 g

CILSS CODE:330 Leafy and Other Vegetables

2561110711M0 Okra fresh
2571110711Q1 Egg plant green fresh

CILSS CODE:333

1110411C0 Milk sterilized homogenized long life
1110421B0 Milk powdered skimmed 450 g
1110421C0 Milk sweetened condensed

CILSS CODE:334

The average of all ICP food items listed above

VI. PANEL DATA SETS

Panel data, i.e. data collected from the same households over a period of time, are particularly useful in monitoring changes that take place in the household economy. Comparison of time-series data from panel households can be made without regard to inter-household variation, thus enabling comparisons to be made more easily and definitively. The CILSS was designed to include the collection of panel data. Every year, one-half of the sample's households was replaced while the other half (the panel households) remained the same as that of the previous year's. This design yielded three sets of panel households (1985-86, 86-87, 87-88).

According to sample design, each panel data set was to contain 800 households. In practice, migration of households and individuals made it impossible for enumerators to access all the panel households and their members, as planned. In such cases, survey enumerators were asked to interview the new family now living in the same dwelling, while retaining the same household identification number (Cluster and NH) as the previous year's for that household. From this, it is immediatly obvious that simply matching household identification numbers between data for two consecutive years in an attempt to identify panel households, would result in the incorrect inclusion of non-panel households.

Section 17 of the CILSS questionnaire was especially built into the questionnaire in order to aid in the identification of panel households. Section 17 contains a variable called OPID - Old PID which provides the Personal Identification Number (PID) of each person who was also interviewed in the previous year. However, utilizing this section results in the identification of far less (701 households in the case of CILSS 1988) than the maximum possible size of the Panel (800 households). Therefore, other approaches were tried with the hope that they would result in the identification of larger numbers of panel households.

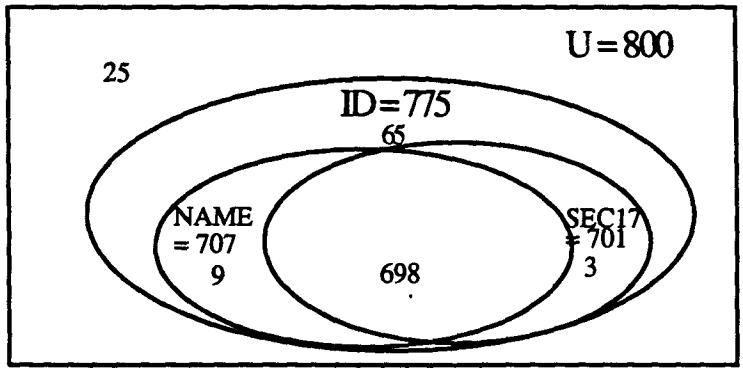
One method tried was to classify as panel households (from among those households

whose household identification numbers matched) only those households for which the household head's characteristics matched from year to consecutive year. The head's gender, age, place of birth, nationality, information on school attendance, highest level of education, highest diploma obtained and information on apprenticeship, were some of the key variables used. This approach proved inadequate, however, due to inconsistency of such data from year to year.

Other approaches, therefore, had to be considered. One possibility was to try and match the names of household members for all households whose household identification number matched (the necessary but not sufficient condition to qualify as a panel household). If the names of at least 50 percent of the household members could be matched from year to year, that household was considered part of the panel. Although the name-matching approach appeared generally satisfactory, the non-availability of names for the first two years of survey data (due to logistical reasons), meant that this approach could not be uniformly applied through out the four years. Thus, this approach also had to be ruled out.

Alternative approaches having been eschewed, we were brought back to the Section 17 approach. The Venn diagram below (using CILSS 1988 as an example) shows that the loss of observations resulting from the use of the Section 17 approach, as compared with the name-matching approach, is minor.

Venn diagram for three matchings



The Venn diagram shows that from a theoretical sample size of 800 panel households, 775 households were found to fulfill the primary criteria to qualify as a panel household, i.e. their household identification numbers matched from one year to the next. 707 households were identified through the Name-matching approach, while 701 households were matched through Section 17. The Venn diagram shows that 698 households were matched using either criteria. These 698 households may therefore be categorized as the most definitive set of panel households - their household IDs matched and names of at least 50 percent of household members could be matched and they qualified as panel households using Section 17. Given the authority of the 698 households as the most definitive, we can identify the degree of error made by using the Section 17 (we may be incorrectly accepting 3 households as panel when in fact they are non-panel) and with Name-matching approach (we may be incorrectly accepting 9 households as panel).

Methodology

Section 17 contains a variable called OPID, Old Person-ID, which can be used to identify the PID that was given to the same individual in the previous year's survey. If an individual appearing in Chapter 17 had not been interviewed in the previous year, the variable OPID would be missing in that case. Using OPID, we were able to identify all households for which at least one member was present in two consecutive years. These households were identified as Panel households.

The documentation on the next page describes the file structure and contents of the Panel data sets.

PANEL DATA SETS OF COTE D'IVOIRE CILSS
FROM 1985 TO 1988

Panel Sets (SAS file names: PANEL.SET)	
(SET01)	Non-Panel Households in 1985 - 795 households (hhs)
(SET02)	Unmatched households from 1st panel in 1985 - 79 hhs
(SET03)	1st panel households in 1985 - 714 hhs
(SET04)	1st panel households in 1986 - 714 hhs
(SET05)	Unmatched households from 1st panel in 1986 - 86 hhs
(SET06)	Unmatched households from 2nd panel in 1986 - 107 hhs
(SET07)	2nd panel households in 1986 - 693 hhs
(SET08)	2nd panel households in 1987 - 693 hhs
(SET09)	Unmatched households from 2nd panel in 1987 - 107 hhs
(SET10)	Unmatched households from 3rd panel in 1987 - 99 hhs
(SET11)	3rd panel households in 1987 - 701 hhs
(SET12)	3rd panel households in 1988 - 701 hhs
(SET13)	Unmatched households from 3rd panel in 1988 - 99 hhs
(SET14)	Non-panel households in 1988 - 800 hhs.

The panel data sets SET01 PANEL through SET14 PANEL are for the household level and the sets SET01IND PANEL through SET14IND PANEL are for the individual level as follows:

Panel Data Sets	Individuals	Panel Data Sets	Individuals
SET01IND	6,537	SET08IND	5,099
SET02IND	448	SET09IND	671
SET03IND	5,762	SET10IND	573
SET04IND	5,736	SET11IND	4,570
SET05IND	495	SET12IND	4,577
SET06IND	769	SET13IND	523
SET07IND	5,399	SET14IND	4,768

REFERENCES

Ainsworth, M. and J. Munoz (1986). "The Cote d'Ivoire Living Standards Survey: Design and Implementation," Living Standards Measurement Study Working Paper No. 26, World Bank, Washington D.C.

Daho, B. (1991). "Note Concernant l'Enquete Permanente Aupres Des Menages". Poverty and Social Policy Division, Africa Technical Department, World Bank, Washington, DC.

Demery, L. and C. Grootaert (1992). "Correcting for Sampling Errors in the Measurement of Welfare and Poverty: The Case of the Cote d'Ivoire Living Standard Survey." Poverty and Social Policy Division, The World Bank, Washington, D.C. Processed.

Glewwe, P. (1990). "Investigating the Determinants of Household Welfare in Cote d'Ivoire". Living Standards Measurement Study Working Paper No. 71, World Bank, Washington D.C.

Grootaert, C. (1983). "The Conceptual Basis of Measures of Household Welfare and their Implied Survey Requirements", Review of Income and Wealth, Series 29, pp.1-21.

Grootaert, C. (1986). "Measuring and Analyzing Levels of Living in Developing Countries: An Annotated Questionnaire." Living Standard Measurement Study Working Paper No. 24, The World Bank, Washington, D.C.

Grootaert, C. (1992). "The Evolution of Welfare and Poverty during Economic Recession and Structual Change-The Case of Cote d'Ivoire, 1985 - 1988." Poverty and Social Policy Division. The World Bank, Washington, D.C.

Grootaert, C. and R. Kanbur (1992). "A New Regional Price Index for Cote d'Ivoire Using Data from the International Comparisons Project." Poverty and Social Policy Division, The World Bank, Washington, D.C. Processed.

Johnson, M; A.D. McKay and J.I. Round (1990). "Income and Expenditure in a System of Household Accounts - Concepts and Estimation." Social Dimensions of Adjustment In Sub-Saharan Africa Working Paper No. 10, The World Bank, Washington, D.C.

Venkataraman, M. (1991). "Report on the Workshop on the Cote d'Ivoire Living Standards Survey Data. July 26 and 31, 1991. "Poverty and Social Policy Division, Africa Technical Department, The World Bank, Washington, D.C.

CORRECTIONS & REVISIONS TO OH & VENKATARAMAN
TABLE ON P. 32--ABIDJAN WEIGHTS

Note : Definition of housing types

Residentiel: villas, delux houses
Economique: modern type apartments
Evolutif: similar to traditional village compounds but set in city blocks
Spontane: Slums, squatter areas

Corrective Weights for Abidjan, 1985

Type of Housing	Cluster No.	Household Number	1985 Weights
Residentiel	7,8	All Households	0.329
Economique	1	31,61	
	6, 13-17	All Households	0.895
Spontane	12	All Households	4.182
Evolutif	1	All households except 31 and 61	0.899
	2-5, 9-11, 18-21	All Households	

Corrective Weights for Abidjan, 1986 and Panel 1986-87

Type of Housing	Cluster No.	Household Number	1986 and Panel 1986-87 weights
Residentiel	7,8	All Households	0.371
Economique	1	31	
	6, 13-17	All Households	0.840
Spontane	12	All Households	4.498
Evolutif	1	All households except 31	0.894
	2-5, 9-11, 18-21	All Households	