Appendix A

TECHNICAL NOTES

A1. Sampling and sampling errors

Sample design

The sample of households interviewed in the 2000/2001 HBS was selected in two stages. In the first stage, 1,161 small areas called Primary Sampling Units (PSUs) were selected throughout the country. In the second stage, 24 households were initially selected in each PSU.

The sampled households are located in the National Master Sample (NMS) of PSUs. The NMS is a generalised set of area units that can be used as PSUs for conducting various household surveys. It is a fixed sample of rural and urban clusters, which, among other things, make possible the performance of a continuous survey programme as well as ad hoc sample surveys. The NMS has four modules, A, A+B, A+B+C and A+B+C+D, which can provide urban and rural estimates at National, Zonal, Regional and District levels respectively[33] .

The HBS 2000/01 used Module A+B+C of the NMS comprising 621 urban EAs and 540 rural villages drawn from each of the 20 regions of Mainland Tanzania. In the second stage, 24 households were selected using systematic random sampling (SRS) from stratified lists of households compiled from each of the sampled PSUs. These lists were stratified into high, middle and low socio-economic groups based on socio-economic data collected during the listing exercise. The stratification and selection of households was conducted in the NBS head office and interviewers were supplied with a list of pre-selected households for interview.

Sample frame and sample selection

Rural frame. The initial rural NMS frame was based on the 1978 Population Census and later updated with information from the 1988 Population Census. At the beginning, a ward or a group of wards was used as a Primary Sampling Unit (PSU), but later a village was used instead. The rural frame of the NMS was divided into "normal," "large town surroundings" and "low density" strata. In total, 150 strata were created and 2 to 8 PSUs (villages) were selected from each stratum to come up with the sample of villages that can provide estimates for each region of Mainland Tanzania (Module A+B+C). These villages were selected using the probability proportional to size (PPS) selection procedure. The PSUs (villages) for Module A of the rural NMS are automatically included in the regional sample.

33. For further details see National Bureau of Statistics (1991): 'The Rural National Master Sample – Technical Report' and National Bureau of Statistics (1993) 'The Urban National Master Sample – Technical Report'.

Urban frame. The urban frame for the NMS was the sample used for the 1988 Population Census detailed questionnaire. For each district in a region, a list of the urban EAs was compiled and a specific number of EAs was selected from this frame using the systematic random sampling (SRS) procedure to produce the regional urban sample.

First stage sampling weights

Dar-es-Salaam. First stage sampling weights for Dar-es-Salaam are those used for Module A because the PSUs are the same. The EAs from the 1988 Census sample were stratified into proxy income levels and combined for all districts within the Dar-es-Salaam region. They were then selected independently within each level using the SRS procedure. Details on how these weights were calculated are found in "*The National Master Sample* (NMS) – *Technical Report*" (cited above). The formula for calculating the weights is:

$$W_{hk} = \frac{\sum_h \frac{V_h}{A_{hk}} \text{x } N_h}{\sum_h \frac{V_h}{A_{hk}} \text{x } n_h} \text{x } \frac{\sum_h A_{hk}}{a_k}$$

where:

W_{hk}= First stage weight for an EA in stratum *k* of cluster *h*
V_h/A_{hk} = the proportion of the sample that falls into district *h* to the selection interval
N_h = number of EAs in district *h*
n_h = number of sampled EAs in district *h*
A_{hk} = number of EAs in district h and NMS stratum *k*
a_k = number of sampled EAs in NMS stratum *k*

When the multiple of the selection interval is completely within stratum *k* of district *h*, the proportion V_h/A_{hk} becomes 1.

Other Urban. For other urban areas, a sample of about 30 EAs was targeted for each region. Each district within the region contributed a certain proportion of the 30 EAs. The EAs were then selected independently from each district in the region using the SRS procedure. EAs representing municipalities and other urban areas in Module A of the NMS were automatically included in the regional sample. The formula for calculating the weights for an EA in district j of region i is given by:

$$W_{ij} = \frac{T_{ij}}{C_{ij}} \text{x } \frac{C_{ij}}{S_{ij}} = \frac{T_{ij}}{S_{ij}}$$

where:

W_{ij} = First stage weight for a selected EA in district *j* of region *i*
T_{ij} = total number of urban EAs in district *j* of region *i*
C_{ij} = number of selected urban EAs for the census sample in district *j* of region *i*
S_{ij} = number of selected urban EAs for the NMS (Module A+B+C) in district *j* of region *i*

Rural. The rural NMS (Module A+B+C) has been used by a number of previous agricultural surveys. The first stage selection of PSUs was done using the PPS sampling procedure. The formula for the first stage weights is as follows:

$$W_{ij} = \frac{P_i}{P_{ij} \text{ x } n_i}$$

where:

W_{ij} = First stage weight for a selected village *j* from stratum *i*
n_i = number of villages selected from stratum *i*
P_i = 1998 population of stratum *i*
P_{ij} = 1998 population of village *j* from stratum *i*

Second stage sampling weights

The basic second stage weights are given by:

$$W_{khl} = \frac{M_{kh}}{m_{kh}}$$

where:

W_{khl} = Second stage weight for a selected household *l* in socio-economic group *h* (High, Middle or Low) of PSU *k*
M_{kh} = total number of households in socio-economic group *h* (High, Middle or Low) of PSU *k*
M_{kh} = number of households interviewed in socio-economic group *h* (High, Middle or Low) of PSU *k*

Adjustments

These weights were adjusted for the reduction in the sample during the second half of the survey. This adjustment inflated the weight of rural households in the second half of the survey so that estimates of variables with seasonal patterns are unbiased.

In addition, it was found that the sum of the weights (multiplied by household size) did not equal the projected population of Tanzania. A constant adjustment factor was incorporated to correct this sum. In practice this will have no effect on the estimates, since totals are never estimated directly from the data. It is not clear what the cause of the problem was. If it were due to under-listing of households, there would be concern if households of different types were under-represented to different degrees. A similar problem was experienced in the 1991/92 HBS.

An adjustment for household non-response is included in the specification of the second stage weights given above.

1991/92 Household Budget Survey

The 1991/92 HBS used Module A of the NMS, which comprised 122 urban EAs, drawn independently from the City of Dar-es-Salaam (52 EAs,) Municipalities (40 EAs,) and Other Urban Centres (30 EAs;) and 100 villages for the rural sample. First and second stage weights were calculated as specified above for the 2000/01 HBS, although the adjustment for non-response was incorporated at the PSU level rather than by socio-economic group within a PSU. During the re-analysis of the 1991/92 data, the effect of adjusting for non-response within each group was examined and found to be negligible.

However, the original weights were found to under-represent the rural population in the final sample analysed; this was large enough to have some effect on estimates. The weights were adjusted to increase the weight given to rural households and bring their share of total population into line with the 1988 Census. The share of the population constituted by each area in the final weighted sample of each survey is given in Table A1.1.

Both surveys used the urban-rural classification of PSUs given in the NMS. As a result, the only way in which the urban share of the weighted sample can increase is through differential growth of the PSUs, since no PSUs have been reclassified. This means that there may be a small number of peri-urban PSUs that are classed as rural in this analysis but would be considered urban if classified now. The effect on the estimates is likely to be small. The 2000/01 Dar es Salaam estimates also include a small number of rural households, whereas the 1991/92 sample included only urban households; the effect of this difference is very small.

TABLE A1.1 POPULATION SHARE OF WEIGHTED SAMPLE (%)

Area	1991/92	2000/01
Dar es Salaam	5.3	5.8
Other urban areas	12.6	13.8
Rural areas	82.1	80.4
Total	100.0	100.0

Sample size, losses and replacement

The final sample analysed for the 2000/01 HBS consisted of 22,178 households, a large sample for any household budget survey. Three PSUs were lost entirely from the sample. Households were included in the analysis if they had at least one record in both the roster and the monthly diary. The weights were calculated for this group of households.

Field supervisors were supplied with a list of twelve 'replacement' households, drawn as a separate sample at the same time as the main household sample, to be used if a sampled household could not be interviewed for the duration of the survey. The 2000/01 HBS sample had a high level of replacement of households that were not interviewed – around 12 per cent.

A total of 4,823 households were analysed for the 1991/92 sample. Losses were higher; levels of replacement were lower (Table A1.2). In both surveys, households that were part of the initial selection constitute around 85 per cent of the sample analysed.

TABLE A1.2 HOUSEHOLDS SAMPLED, LOST AND REPLACED

	2000/01	1991/92
No of PSUs included in final sample analysed	1,158	222
No of households selected in final sample	22,584	5,328
Total number analysed	22,178	4,823
No of first selections interviewed	19,500	4,466
No of replacements interviewed	2,678	357
Total number analysed as a percentage of initial sample	98.2	90.5
No of first selections interviewed as a per cent of initial sample	86.3	83.8
Replacements as a percentage of initial sample	11.9	6.7

The distribution by region of the 2000/01 HBS sample is given in Table A1.3.

TABLE A1.3 NUMBER OF HOUSEHOLDS INTERVIEWED BY REGION (HBS 2000/01)

Region	Urban	Rural	Total
Dodoma	710	502	1,212
Arusha	693	432	1,125
Kilimanjaro	642	435	1,077
Tanga	701	416	1,117
Morogoro	716	408	1,124
Pwani	700	287	987
Dar Es Salaam	1,167	58	1,225
Lindi	718	379	1,097
Mtwara	695	396	1,091
Ruvuma	718	384	1,102
Iringa	730	395	1,125
Mbeya	720	412	1,132
Singida	714	358	1,072
Tabora	715	381	1,096
Rukwa	682	324	1,006
Kigoma	744	406	1,150
Shinyanga	668	406	1,074
Kagera	700	449	1,149
Mwanza	712	404	1,116
Mara	706	395	1,101
Total	14,551	7,627	22,178

Sampling errors

Table A1.4 shows standard errors and confidence intervals around a number of estimates, calculated in STATA. It also presents the results of statistical tests for a significant difference between the 2000/01 and 1991/92 estimates, for the total population and each of the three areas. While STATA allows the specification of sample design in the calculation of sampling errors, identifying the strata and PSUs used, it is not possible to specify fully the complexity of the design of the HBS 2000/01. The standard errors, confidence intervals and tests are therefore approximate.

TABLE A1.4 STANDARD ERRORS AND CONFIDENCE INTERVALS AROUND SELECTED ESTIMATES

Estimate		Estimate	SE	95% Confidence Intervals		Significance of diff. 91/92 - 00/01 (p)
				Lower	Upper	
Percentage of female-headed households						
2000/01	Total	0.229	0.009	0.212	0.246	0.000
	Dar es Salaam	0.209	0.019	0.172	0.246	0.009
	Other urban	0.279	0.012	0.255	0.302	0.140
	Rural	0.221	0.011	0.200	0.242	0.001
1991/92	Total	0.176	0.011	0.154	0.197	--
	Dar es Salaam	0.141	0.017	0.107	0.175	--
	Other urban	0.239	0.024	0.192	0.286	--
	Rural	0.167	0.013	0.142	0.192	--

Percentage of households owning a radio						
2000/01	Total	0.518	0.013	0.493	0.543	0.000
	Dar es Salaam	0.794	0.019	0.757	0.831	0.976
	Other urban	0.713	0.014	0.685	0.741	0.002
	Rural	0.457	0.015	0.428	0.486	0.000
1991/92	Total	0.374	0.020	0.335	0.412	--
	Dar es Salaam	0.795	0.027	0.741	0.849	--
	Other urban	0.557	0.048	0.462	0.652	--
	Rural	0.306	0.023	0.261	0.352	--
Percentage of adults with no education						
2000/01	Total	0.252	0.010	0.232	0.271	0.869
	Dar es Salaam	0.076	0.011	0.053	0.098	0.400
	Other urban	0.131	0.007	0.116	0.147	0.992
	Rural	0.290	0.012	0.266	0.314	0.618
1991/92	Total	0.249	0.122	0.225	0.273	--
	Dar es Salaam	0.090	0.012	0.066	0.113	--
	Other urban	0.130	0.020	0.091	0.169	--
	Rural	0.280	0.015	0.252	0.309	--
Percentage of children aged 7-13 years reported as studying						
2000/01	Total	0.614	0.015	0.585	0.642	0.086
	Dar es Salaam	0.760	0.024	0.713	0.807	0.010
	Other urban	0.765	0.019	0.727	0.803	0.012
	Rural	0.581	0.017	0.549	0.614	0.395
1991/92	Total	0.574	0.018	0.538	0.609	--
	Dar es Salaam	0.657	0.031	0.596	0.718	--
	Other urban	0.636	0.047	0.542	0.729	--
	Rural	0.559	0.021	0.518	0.599	--
Percentage of households with piped or protected water sources						
2000/01	Total	0.555	0.020	0.514	0.595	0.016
	Dar es Salaam	0.936	0.023	0.891	0.981	0.212
	Other urban	0.880	0.016	0.849	0.910	0.340
	Rural	0.459	0.025	0.411	0.508	0.020
1991/92	Total	0.459	0.034	0.394	0.525	--
	Dar es Salaam	0.968	0.011	0.946	0.990	--
	Other urban	0.837	0.042	0.753	0.920	--
	Rural	0.349	0.040	0.270	0.428	--
Percentage of adults in agriculture (main economic activity)						
2000/01	Total	0.633	0.013	0.607	0.658	0.000
	Dar es Salaam	0.030	0.007	0.016	0.045	0.441
	Other urban	0.269	0.017	0.236	0.303	0.006
	Rural	0.758	0.011	0.737	0.780	0.000
1991/92	Total	0.728	0.012	0.705	0.752	--
	Dar es Salaam	0.023	0.006	0.010	0.035	--
	Other urban	0.430	0.056	0.321	0.540	--
	Rural	0.834	0.011	0.812	0.857	--
Percentage of households within 2km of a primary school						
2000/01	Total	0.635	0.017	0.600	0.669	0.393
	Dar es Salaam	0.813	0.039	0.736	0.890	0.364
	Other urban	0.823	0.022	0.779	0.866	0.243
	Rural	0.583	0.022	0.541	0.625	0.277
1991/92	Total	0.663	0.029	0.607	0.719	--
	Dar	0.866	0.042	0.782	0.949	--
	Other urban	0.765	0.044	0.677	0.852	--
	Rural	0.628	0.035	0.558	0.697	--

Percentage of households within 6km of dispensary / health centre						
2000/01	Total	0.755	0.020	0.716	0.794	0.972
	Dar es Salaam	0.981	0.009	0.963	0.998	0.097
	Other urban	0.978	0.005	0.968	0.987	0.551
	Rural	0.693	0.025	0.645	0.741	0.842
1991/92	Total	0.754	0.033	0.690	0.818	--
	Dar es Salaam	0.950	0.016	0.918	0.982	--
	Other urban	0.968	0.015	0.937	0.998	--
	Rural	0.702	0.040	0.624	0.781	--
Percentage of households within 1km of drinking water						
2000/01	Total	0.549	0.018	0.514	0.584	0.174
	Dar es Salaam	0.840	0.046	0.749	0.930	0.393
	Other urban	0.732	0.022	0.688	0.776	0.343
	Rural	0.490	0.021	0.448	0.531	0.237
1991/92	Total	0.499	0.032	0.436	0.562	--
	Dar es Salaam	0.885	0.026	0.833	0.937	--
	Other urban	0.668	0.064	0.543	0.794	--
	Rural	0.438	0.038	0.362	0.513	--
Mean expenditure per adult equivalent (in 2000/01 prices)						
2000/01	Total	10,884	250	10,393	11,374	0.003
	Dar es Salaam	15,944	779	14,415	17,472	0.000
	Other urban	13,533	496	12,560	14,506	0.078
	Rural	10,064	273	9,528	10,601	0.099
1991/92	Total	9,746	285	9,186	10,306	--
	Dar es Salaam	10,640	387	9,874	11,407	--
	Other urban	11,865	806	10,283	13,447	--
	Rural	9,362	326	8,722	10,002	--
Percentage of individuals below the food poverty line						
2000/01	Total	0.187	0.014	0.159	0.215	0.215
	Dar es Salaam	0.075	0.017	0.042	0.108	0.020
	Other urban	0.132	0.018	0.097	0.168	0.697
	Rural	0.204	0.017	0.171	0.238	0.341
1991/92	Total	0.216	0.019	0.178	0.254	--
	Dar es Salaam	0.136	0.020	0.098	0.174	--
	Other urban	0.150	0.041	0.068	0.231	--
	Rural	0.231	0.023	0.186	0.276	--
Percentage of individuals below the basic needs poverty line						
2000/01	Total	0.357	0.016	0.325	0.389	0.293
	Dar es Salaam	0.176	0.027	0.124	0.229	0.008
	Other urban	0.258	0.022	0.215	0.301	0.604
	Rural	0.387	0.020	0.349	0.425	0.520
1991/92	Total	0.386	0.023	0.342	0.430	--
	Dar es Salaam	0.281	0.028	0.226	0.336	--
	Other urban	0.287	0.050	0.187	0.386	--
	Rural	0.408	0.026	0.357	0.460	--

A2. Calculating the consumption aggregate and defining the poverty lines

This appendix outlines the cleaning of the consumption data, the calculation of the consumption aggregate and the setting of the poverty lines.

Cleaning the consumption data

The data on consumption were the most problematic because such a large volume was collected – there were 5.6 million records in the files containing the data from the monthly diaries. Despite the consistency checks built into the data entry and checking programmes, a substantial number of households had problems in the consumption/expenditure component of the final data set, which had to be resolved at the beginning of the analysis. The same cleaning procedures were carried out on the 1991/92 data.

Food consumption data

Most errors were identified in the food consumption data. These were largely due to a number of identifiable, simple errors, most of which could be corrected. An important quality indicator used was daily calorie consumption per person. This was calculated for all households and out-of-range households were investigated. About 1,000 households had a consumption of over 8,000 calories per person per day and a smaller number had very low calorie consumption.

The main causes of high calories were clear: interviewers miscoding as consumption bulk items that had been bought for re-sale or storage; miscoding as consumption harvested food that was stored or sold; and miscoding the units of quantity (grams as kilograms etc).

Cleaning took place in a number of stages:

1. A programme written in Visual Basic identified extreme outlying unit prices and automatically corrected the units of quantity. These were identified when the *z*-score of the unit price was brought within range by transformation by a factor of 103 or 10^{-3} or powers of the same. This identified gram/kilogram and similar errors very effectively.
2. Less extreme outliers where then identified by examining consumption per person for each commodity. Outliers with consumption over 20 times the median were flagged and the quantity was replaced with the median level of consumption for the size of the household. For some commodities, a larger multiple than 20 was used as a cut-off, to be sure that plausible values were not being replaced. This dealt with much of the mis-coding of items bought for trade or harvested and stored/sold.

3. Households with low calorie consumption were also investigated. In a number of cases, unit miscoding of staple carbohydrates could be identified as the cause, through the presence of outlying unit prices for these items. The correct quantities were imputed using the reported expenditure on the item consumed.

After cleaning, calorie consumption was re-calculated. Far fewer households were found to have outlying values.

Non-food items

Some problems were also identified in the non-food items. It is not possible to calculate a unit value for these items. Instead, outliers were identified using two criteria: that the expenditure per capita on that item/category was high (more than 10 times the mean value), and that the budget share of the item was also high in comparison with the average share. The latter criterion helps ensure that wealthy households with genuinely high expenditure on a range of items are not mistakenly identified as outliers. For example, a household would clearly not spend 90 per cent of its budget for laundry soap: the mistake could be a miscoding of items bought for trade or some erroneous extra digit. Outlying expenditure values for a given item were replaced with the mean expenditure calculated across all households that had consumed that item.

Reporting of household size and the number of transactions

It was found that average household size dropped significantly during the fieldwork of the 2000/01 survey (Figure A2.1). This is not a true reflection of actual household size and is too large to be a consequence of changes in the sample enumerated over the period. It is almost certainly an outcome of enumerator fatigue. It could potentially have a very damaging effect on consumption and poverty estimates. However, it was found that the average number of transactions also declined over the period (Figure A2.2). As a result, the average number of transactions per person does not show a downward trend over the period, suggesting that the biases have to some degree cancelled out. Furthermore, the 1991/92 data show the same trends (Figures A2.3 and A2.4). Note that the 1991/92 data files contained the transaction data already summed to item level, so Figure A2.4 is not directly comparable to Figure A2.2. Figure A2.5 shows more comparable data for 2000/01.

FIGURE A2.1 MEAN NUMBER OF HOUSEHOLD MEMBERS BY MONTH OF SURVEY (HBS 2000/01)

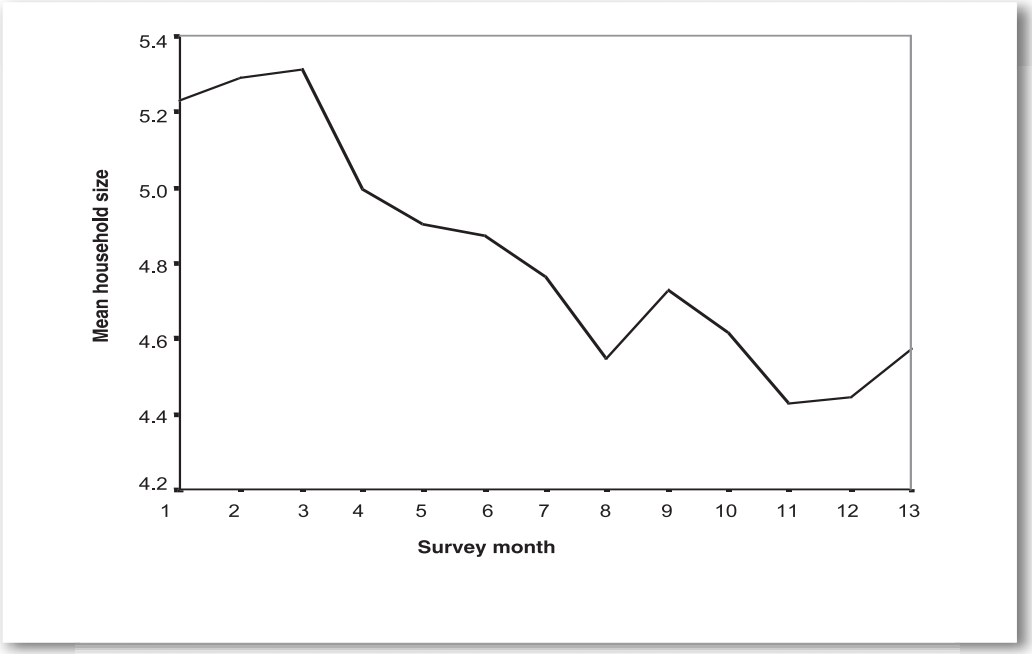


FIGURE A2.2 MEAN NUMBER OF TRANSACTIONS RECORDED IN THE DIARY BY MONTH OF SURVEY (HBS 2000/01)

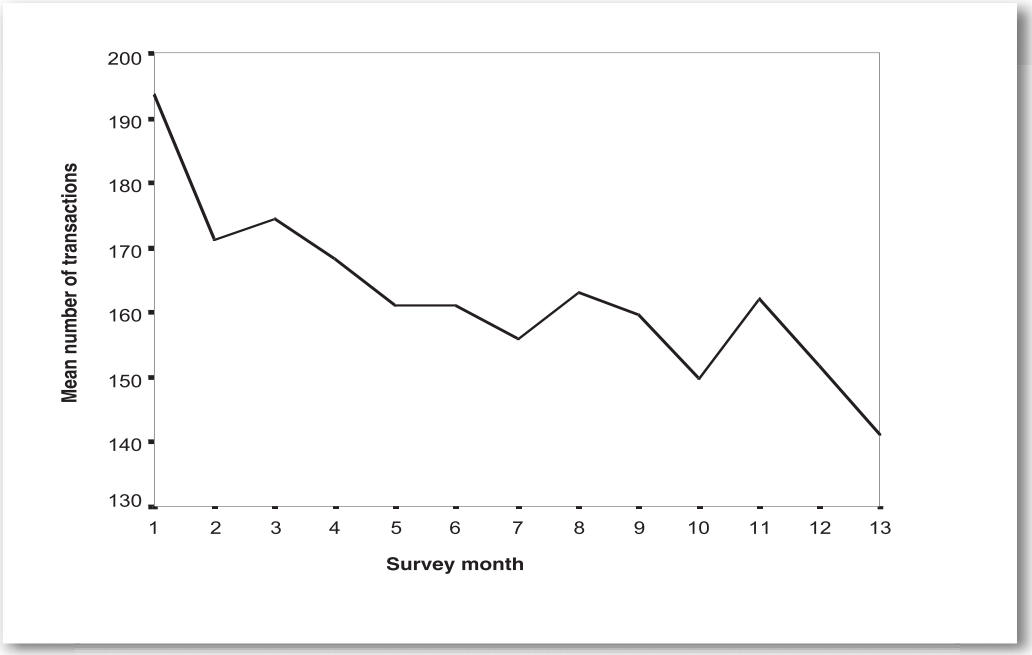


FIGURE A2.3 MEAN NUMBER OF HOUSEHOLD MEMBERS BY MONTH OF SURVEY (HBS 1991/92)

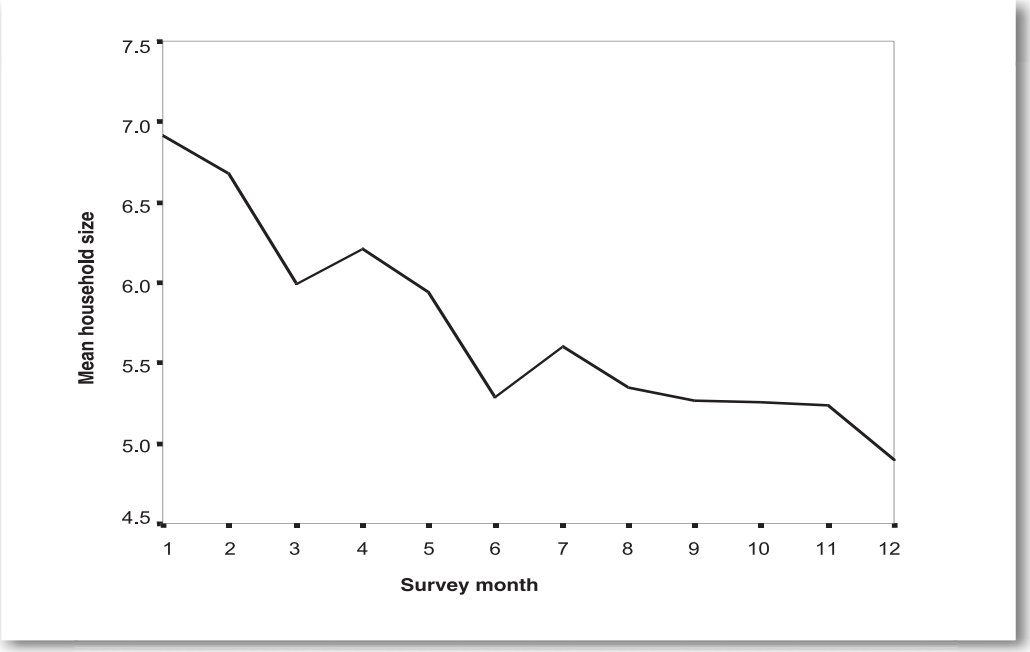


FIGURE A2.4 MEAN NUMBER OF TRANSACTIONS (AT COMMODITY LEVEL) RECORDED IN THE DIARY BY MONTH OF SURVEY (HBS 1991/92)

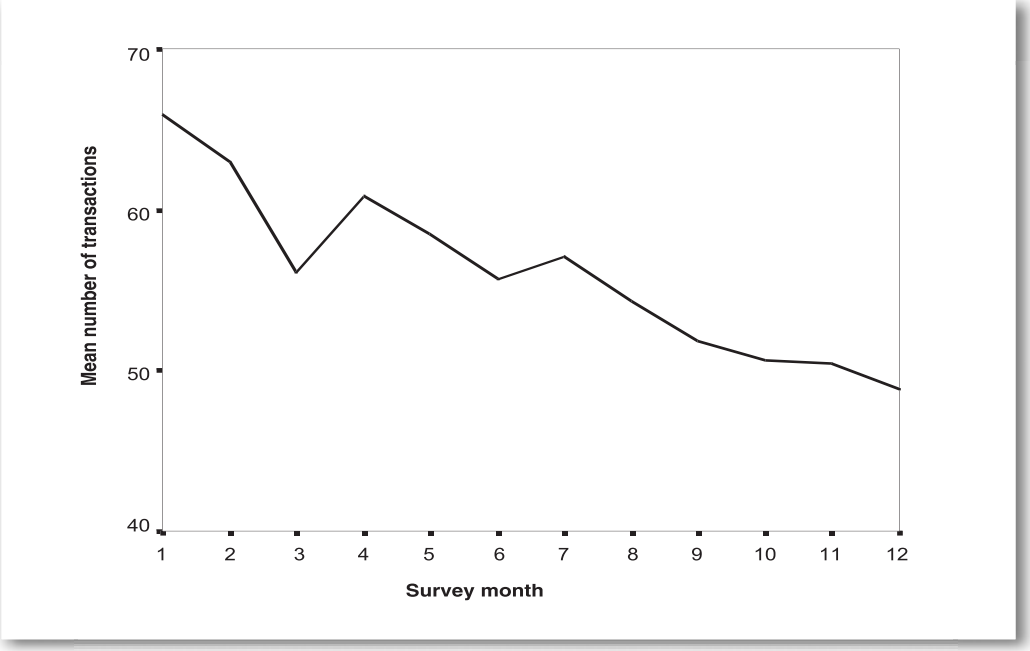
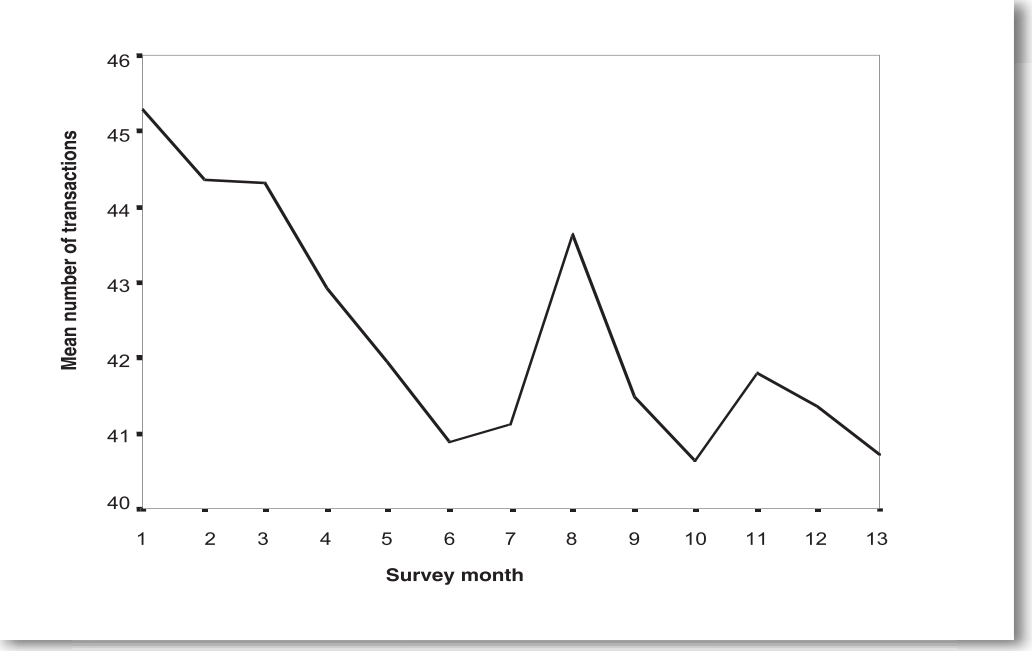


FIGURE A2.5 MEAN NUMBER OF TRANSACTIONS (AT COMMODITY LEVEL) RECORDED IN THE DIARY BY MONTH OF SURVEY (HBS 2000/01)



Calculation of the consumption aggregate

The consumption aggregate sums the value of all items consumed. This includes purchased and home produced items, as well as items received as payment in kind or as gifts. It includes food and drinks, clothing and personal effects, most household effects and consumables (but excluding durables), recreation and transport. Non-consumption expenditure – such as production costs – is excluded.

For a number of items, information on consumption was collected in both the monthly diary and in the annual recall section. The question arises as to which is the better source of information for inclusion in the consumption aggregate. This was decided by examining three measures. These were: the proportion of households reporting consumption of that item in the annual recall; the ratio of the number of households reporting expenses on an item in the annual reports to the number reporting expenses in the diary; and the ratio of the reported amount spent in the annual recall over the amount reported in the diary (the latter multiplied by 12). They can be used to decide which appears to be the more reliable source for each item[34].

34. See Blaizeau, D, 'Household Expenditure in the seven UEMOA countries,' (mimeo).

This analysis was carried out for both years. A number of items were identified as having been paid for much more often in 2000/01 than in 1991/92. This included health, education, water, postage and telephone charges. These are all services that have seen increased cost recovery over the 1990s. On the assumption that these changes largely reflect an increase in payment for similar services (rather than a large increase in the quantity or quality of services supplied), it was decided to exclude them from the consumption aggregate used in the poverty analysis.

Since the diaries were completed over one calendar month, the consumption measure is standardised to 28 days. It is also standardised for the demographic composition of the household, adjusted for the consumption needs of different individuals using the adult equivalence scale shown in Table A2.1.

TABLE A2.1 ADULT EQUIVALENCE SCALE

Age groups	Sex			
	Male		**Female**	
0 – 2	X1	0.40	X2	0.40
3 – 4	X3	0.40	X4	0.48
5 – 6	X5	0.56	X6	0.56
7 – 8	X7	0.64	X8	0.64
9 – 10	X9	0.76	X10	0.76
11 – 12	X11	0.80	X12	0.88
13 – 14	X13	1.00	X14	1.00
15 – 18	X15	1.20	X16	1.00
19 – 59	X17	1.00	X18	0.88
60 +	X19	0.80	X20	0.72

Information is collected in the roster on the number of days that members were present in the household during the month that the diary is completed. It would be possible to adjust the consumption aggregate for the number of days that household members were present during the month of survey, either using this information or looking at the pattern of entries into the diary itself. It was thought that this could have explained some of the low, out-of-range calorie consumption values. In fact there was little relationship between the information on days present, the number of days in the month with transaction data and households with low calorie intake. Given how unreliable this information was, it was decided not to adjust the consumption aggregate. As a sensitivity check, households with a low number of days present per person were removed from the analysis – one per cent of the sample. As it had no appreciable impact on the poverty estimates, these households were left in the analysis.

Adjusting for prices: the Fisher Index

The consumption aggregate provides a measure of how much was spent by a household per adult equivalent. However, the goods and services that can be purchased with that expenditure depend on the prices faced by that household. The poverty line (or equivalently, the consumption aggregate) must be adjusted to reflect this. This is done using the Fisher Ideal Index[35].

The value of any price index will depend on the goods included in it. Consumption patterns vary between different areas, and over time, so a particular consumption basket cannot represent average consumption patterns everywhere. Different consumption patterns will tend to reflect differences in prices, as households substitute a more expensive good with a cheaper one. For example, if the relative price of one staple carbohydrate increases over time, households may shift to another one. A price index that failed to reflect this would overestimate the prices faced by households at the later time, because it would fail to reflect the change in consumption patterns. A Laspeyres index has this disadvantage. A similar argument applies to differences in consumption patterns in different geographical areas.

For this reason, the consumer price index (CPI) was not used to adjust for prices between the two surveys. In addition to the theoretical disadvantage of being a Laspeyres index, the CPI has a number of practical limitations. The consumption basket used is based on the 1991/92 HBS and so will tend to be out-of-date. It is explicitly an urban index. Dar es Salaam weights very heavily in the index, at about 40 per cent, while it is only around 6 per cent of the national population. This means it not representative of prices faced in rural areas.

Instead a Fisher Index is used to adjust for price variation both over time and across different geographical areas. It can be thought of as representing a sort of 'average' consumption pattern between the two populations being compared.

The index is calculated using the price and quantity information from the surveys themselves. Respondents were asked to provide information on how much they spent on each item and on the quantity consumed. The ratio of expenditure to volume provides a measure of price, or more precisely, a measure of unit value. There are half a million unit price observations in the 2000/01 data. In 1991/92 there is information on 272,178 transactions. It is therefore possible to construct a price index both between the surveys and between geographical areas within each year. For 2000/01 it is possible to construct price indexes across regions, although the sample size in 1991/92 does not allow this level of disaggregation. In the 2000/01 data, a separate Fisher Index is calculated for the urban and rural

35. For a detailed description of the Fisher Index and its benefits see Deaton, A. and Tarozzi, A. (1999) 'Prices and Poverty in India', Princeton University (mimeo).

populations of each region, with the national population as the reference population. For the 1991/92 data, indexes are calculated for Dar es Salaam, other urban areas and rural areas, relative to the national population. A separate Fisher index is calculated to measure price differences between the 1991/92 and 2000/01 surveys.

In each case, the largest possible basket of goods is compared between the two populations, subject to there being at least five observations for that item in each population. Items with fewer than this number are excluded from the index. Likewise, the index that is calculated to compare prices between 1991/92 and 2000/01 excludes items that are absent from one or other data set.

The HBS contains information on approximately 128 food items. Items that are not measured in standard units (grams, kilograms, litres and millilitres) are excluded, with the exception of eggs that are measured in 'number' and are included. This removes some 6 per cent of transactions. For a few commodities it is effectively impossible to measure quantities and so these are dropped from the calculation, for example, '10403 – syrups, jams, marmalades, jellies, chocolates, sweets'. Some non-food items were also included in the Fisher index. They were: fuel and lighting; cigarettes and tobacco; laundry soap; and petrol. These are included in the index, although there is little difference in the index if they are excluded.

Records that are missing information on quantity are excluded and the quantity of each item is then standardised to a common unit (grams to kilograms, etc). A unit price is calculated for each transaction record, dividing amount spent (or its equivalent) by the quantity. A small number of outlying unit prices are removed and median unit prices are then calculated for each item. The Fisher Index is then a weighted average of these unit prices, with the median quantity of each item consumed constituting the weights.

Between 1991/92 and 2000/01 the index is 2.49. By contrast, the CPI shows prices rising by a factor of about 4.4 over a similar period. Fisher Indexes for each region in the 2000/01 data are shown in Table C27.

In the analysis, the Fisher Index is used to adjust the consumption aggregate for price variation between the different geographical areas. It is also used to assess whether there has been growth in household consumption in real terms over the 1990s. However, the comparison of poverty levels over time does not use the Fisher Index. During stakeholder consultation, it was decided that separate poverty lines should be set, one for each year. This is outlined below.

Setting the poverty lines

The poverty lines define the minimum expenditure necessary to meet basic human needs. The food poverty line represents the expenditure necessary to eat sufficient calories. The basic needs poverty line includes the cost of other essential items of expenditure.

Food poverty line

Following Ravallion and Bidani (1994)[36], Ravallion (1998)[37] and others, the food poverty line was based on the food basket consumed by the poorest 50 per cent of Tanzanians. An infinite variety of food baskets, differing in price, are consistent with attaining a given level of calories. It would be unreasonable to set a poverty line based on the cheapest possible basket - a diet purely of the staple that provides the cheapest calories – since other nutrients are also necessary for survival. However, choosing which items to include in the basket can be very arbitrary. It was therefore decided to include all items consumed by the poor in the food basket, thus avoiding this problem. Only alcoholic drinks and items that cannot be assigned a calorific value are excluded.

Median quantities consumed per adult equivalent were estimated for every food item. Median unit prices were also calculated. The approximate calorific values of these foods were calculated. The food basket gives the share of consumption accounted for by each item. The level is set so that the sum of calories is 2,200 per day, the minimum necessary for survival. The food basket defined by these two parameters is then priced to give the food poverty line. The calculation of the consumption baskets for each year is shown in Tables A2.2 and A2.3. Note that the quantities shown are the reported quantities consumed, not the food basket quantities, which would need to be obtained by scaling down by the same factor as the total cost (ie around 0.4).

Basic needs poverty line

Adjusting the food poverty line to allow for non-food consumption gives the basic needs poverty line. This is done by calculating the share of expenditure that goes on food in the poorest 25 per cent of households. Multiplying the food poverty line by the inverse of this share inflates it to allow for non-food consumption. The food share was 75 per cent in 1991/92, and 73 per cent in 2000/01.

36. Ravallion, M (1998) 'Poverty Lines in Theory and Practice', Living Standards Measurement Study, Working Paper No. 133, Washington DC.
37. Ravallion, M and Benu Bidani (1994) 'How robust is a poverty line?' World Bank Economic Review 8(1):75-102.

Poverty lines are therefore set separately for the two surveys, although price adjustments within a survey year are made using Fisher Indexes. A poverty line set using the Fisher Index calculated to compare prices between the surveys would imply a slightly lower poverty line in 2000/01 or equivalently a slightly higher line in 1991/92, although the difference would be small - the food poverty line would be 2,126 TShs in 1991/92, for example.

TABLE A2.2 REPORTED CONSUMPTION BY ITEM AND THE CALCULATION OF THE FOOD POVERTY LINE FOR 2000/01

Item code	Item	Median quantity in g/ person /month	Median price per kg	Calories per 100g	Calories per day	Cost per day
10101	paddy	1250	155	361	161	193
10102	rice, husked	1149	400	364	149	459
10103	green maize cob	1136	139	165	67	158
10104	maize, grain	3333	100	368	438	333
10105	maize, flour	7183	165	368	944	1185
10106	millet, grain	439	200	350	55	88
10107	millet, flour	424	200	350	53	85
10108	sorghum, grain	1596	100	341	194	160
10109	sorghum, flour	7355	104	341	896	767
10110	wheat, grain	495	300	323	57	149
10111	wheat, flour	347	350	341	42	122
10201	bread	157	562	261	15	88
10202	baby food excl. milk	85	400	380	12	34
10203	biscuits	15	2000	450	2	30
10205	macaroni, spaghetti	144	577	342	18	83
10301	cassava fresh	1637	86	149	87	140
10302	cassava dry	1587	100	344	195	159
10303	cassava flour	4035	110	344	496	444
10304	sweet potatoes	1695	96	105	64	163
10305	yam. cocoyam	1067	100	118	45	107
10306	potatoes	753	143	79	21	108
10307	cooking bananas,	2564	91	135	124	234
10308	other starches	439	167	79	12	73
10401	sugar	589	598	400	84	352
10402	honey	120	585	400	17	70
10501	peas, dry	362	260	343	44	94
10502	beans, dry	962	294	333	114	283
10503	lentils & other	439	240	338	53	105
10504	pulse product	168	200	127	8	34
10601	groudnuts in shell	206	278	567	42	57
10602	groundnuts, shell	187	400	567	38	75
10603	coconuts, mature	828	168	376	111	139
10604	coconuts, immature	318	105	376	43	33
10605	cashewnuts	120	400	567	24	48
10606	almond & other nuts	231	200	567	47	46
10701	sesame seeds	130	400	605	28	52
10702	sunflower seeds	358	172	605	77	62
10703	products from nuts	51	500	567	10	25
10801	carrots	81	333	43	1	27
10802	radishes, beets,	272	200	43	4	54
10803	garlic	41	500	34	1	21
10804	onion	143	375	34	2	54

10805	leeks	123	250	34	1	31
10806	spinach	294	202	22	2	60
10807	lettuce	189	200	13	1	38
10808	cabbage	355	133	16	2	47
10809	other leafy vegetables	403	221	16	2	89
10810	tomatoes	417	239	19	3	100
10811	bitter tomatoes	118	200	13	1	24
10812	ladies finger	114	333	16	1	38
10813	cauliflower	142	263	13	1	37
10814	cucumber, pumpkin	974	100	26	9	97
10815	brinjals, eggplants	106	222	26	1	23
10816	green peas	287	242	36	4	70
10817	green beans	805	200	36	10	161
10818	fresh green pepper	23	417	25	0	10
10819	cultivated	373	222	13	2	83
10820	other wild vegetables	293	200	13	1	59
10821	dried vegetables	295	248	13	1	73
10822	canned vegetable	58	513	13	0	30
10901	sweet bananas	221	173	92	7	38
10902	orange, tangerine	187	150	47	3	28
10903	grapefruits, lemon	57	200	29	1	11
10904	mangoes, avocado	237	143	65	6	34
10905	pawpaw	439	91	39	6	40
10906	pineapples	446	114	49	8	51
10907	melons	998	62	32	11	61
10909	jack fruit	364	100	49	6	36
10910	apples, pears	149	160	49	3	24
10911	other cultivated	162	200	29	2	32
10912	other wild fruit	95	200	29	1	19
10913	dried fruits	31	250	238	3	8
10914	canned fruits	13	3000	238	1	39
11001	goat, sheep	347	700	122	15	243
11002	cattle meat	526	722	115	22	380
11003	pork	279	733	114	11	205
11004	other domestic animals	184	700	115	8	129
11005	wild animal	362	500	115	15	181
11006	offal	154	650	123	7	100
11007	dried, salted meat	110	1111	115	5	122
11008	canned meat	142	500	225	11	71
11009	other meat products	238	500	225	19	119
11010	chicken & other	338	925	139	17	313
11011	wild birds & insects	198	600	139	10	119
11012	eggs	1	50000	158	0	49
11201	fresh fish	692	400	82	20	277
11202	shell fish	143	1000	89	5	143
11203	fresh dried fish	183	667	225	15	122
11204	dried or salted	387	657	225	31	254
11205	canned fish/shellfish	72	1000	238	6	72
11301	fresh milk	741	200	61	16	148
11302	cream	118	714	355	15	84
11303	cheese	126	1125	355	16	142
11304	yoghurt	704	200	61	15	141
11305	canned milk	291	1700	134	14	494
11306	milk powder	17	2143	362	2	36
11401	cottonseed oil	90	1000	884	28	90
11402	groundnuts oils	63	1000	884	20	63
11403	sesame/sunflower	159	1000	884	50	159
11404	coconut cooking	27	1000	884	9	27

11405	other cooking oil	179	958	884	57	172
11406	butter, ghee	67	1050	717	17	70
11407	margarines cooking fat	85	1000	719	22	85
11408	other oil & fat	63	1000	717	16	63
11501	red pepper/black	32	625	25	0	20
11502	curry powder	11	1000	25	0	11
11503	other spices	19	857	25	0	16
11504	salt	329	200	25	3	66
Total					5,512	13,266
Ratio of calories needed per day (2,200) to reported consumed					0.40	
Cost of food basket adjusted by this ratio: food poverty line						5,295

Note: Full descriptions of each item code are given in the HBS 2000/01 interviewers' manual.

TABLE A2.3 CALCULATION OF THE FOOD POVERTY LINE FOR 1991/92

Item code	Item	Median quantity in g/ person /month	Median price per kg	Calories per 100g	Calories per day	Cost per day
10101	paddy	1217	67	361	157	81
10102	rice, husked	1369	146	364	178	199
10103	green maize cob	878	54	165	52	47
10104	maize, grain	5123	47	368	673	242
10105	maize, flour	5282	75	368	694	396
10106	millet, grain	565	80	350	71	45
10107	millet, flour	763	59	350	95	45
10108	sorghum, grain	1776	56	341	216	99
10109	sorghum, flour	4795	68	341	584	326
10110	wheat, grain	352	120	323	41	42
10111	wheat, flour	344	200	341	42	69
10201	bread	187	251	261	17	47
10202	baby food excl.	134	88	380	18	12
10203	biscuits	31	300	450	5	9
10205	macaroni, spaghetti	176	400	342	22	70
10301	cassava fresh	1451	39	149	77	56
10302	cassava dry	1386	40	344	170	55
10303	cassava flour	3623	53	344	445	193
10304	sweet potatoes	1747	37	105	66	64
10305	yam. cocoyam	1036	43	118	44	44
10306	potatoes	862	50	79	24	43
10307	cooking bananas,	3521	30	135	170	106
10308	other starches	1036	50	79	29	52
10401	sugar	517	241	400	74	124
10402	honey	182	200	400	26	36
10501	peas, dry	487	90	343	60	44
10502	beans, dry	1055	100	333	125	106
10503	lentils & other	545	74	338	66	40
10504	pulse product	167	80	127	8	13
10601	groundnuts in shell	287	100	567	58	29
10602	groundnuts, shell	214	150	567	43	32
10603	coconuts, mature	500	94	376	67	47
10604	coconuts, immature	267	60	376	36	16
10605	cashewnuts	174	161	567	35	28
10606	almond & other nuts	89	140	567	18	12

10701	sesame seeds	136	100	605	29	14
10702	sunflower seeds	247	117	605	53	29
10703	products from nuts	52	167	567	11	9
10801	carrots	53	133	43	1	7
10802	radishes, beets,	181	100	43	3	18
10803	garlic	66	86	34	1	6
10804	onion	152	111	34	2	17
10805	leeks	48	140	34	1	7
10806	spinach	225	100	22	2	23
10807	lettuce	242	119	13	1	29
10808	cabbage	357	52	16	2	19
10809	other leafy vegetables	355	96	16	2	34
10810	tomatoes	309	100	19	2	31
10811	bitter tomatoes	113	83	13	1	9
10812	ladies finger	96	130	16	1	13
10813	cauliflower	190	100	13	1	19
10814	cucumber, pumpkin	1116	35	26	10	39
10815	brinjals, eggplants	143	83	26	1	12
10816	green peas	489	80	36	6	39
10817	green beans	503	75	36	6	38
10818	fresh green peppers	22	300	25	0	7
10819	cultivated	277	94	13	1	26
10820	other wild vegetables	296	100	13	1	30
10821	dried vegetables	188	133	13	1	25
10822	canned vegetable	15	615	13	0	10
10901	sweet bananas	305	50	92	10	15
10902	orange, tangerine	250	50	47	4	13
10903	grapefruits, lemon	84	67	29	1	6
10904	mangoes, avocado	252	55	65	6	14
10905	pawpaw	375	25	39	5	9
10906	pineapples	342	50	49	6	17
10907	melons	648	11	32	7	7
10909	jack fruit	172	64	49	3	11
10910	apples, pears	110	93	49	2	10
10911	other cultivated	136	80	29	1	11
10912	other wild fruit	395	33	29	4	13
10913	dried fruits	20	500	238	2	10
11001	goat, sheep	285	200	122	12	57
11002	cattle meat	644	209	115	26	134
11003	pork	510	240	114	21	122
11004	other domestic animals	223	300	115	9	67
11005	wild animal	298	150	115	12	45
11006	offal	250	200	123	11	50
11007	dried, salted meat	250	100	115	10	25
11008	canned meat	196	300	225	16	59
11009	other meat products	321	140	225	26	45
11010	chicken & other	415	250	139	21	104
11011	wild birds & ins	230	125	139	11	29
11012	eggs	1	19474	158	0	23
11201	fresh fish	573	165	82	17	94
11202	shell fish	188	209	89	6	39
11203	fresh dried fish	192	211	225	15	40
11204	dried or salted	310	250	225	25	77
11205	canned fish/shellfish	25	80	238	2	2
11301	fresh milk	768	63	61	17	48
11302	cream	283	230	355	36	65
11303	cheese	1168	20	355	148	23
11304	yoghurt	679	70	61	15	48
11305	canned milk	155	50	134	7	8

11306	milk powder	201	120	362	26	24
11401	cottonseed oil	100	383	884	32	38
11402	groundnuts oils	114	384	884	36	44
11403	sesame/sunflower	87	417	884	27	36
11404	coconut cooking	43	320	884	13	14
11405	other cooking oil	152	444	884	48	67
11406	butter, ghee	39	800	717	10	31
11407	margarines cooking fat	135	474	719	35	64
11408	other oil & fat |	103	333	717	26	34
11501	red pepper/black	56	150	25	0	8
11502	curry powder	16	500	25	0	8
11503	other spices	42	286	25	0	12
11504	salt	323	92	25	3	30
Total					5,418	5,130
Ratio of calories needed per day (2,200) to reported consumed					0.41	
Cost of food basket adjusted by this ratio: food poverty line						2,083

Note: Full descriptions of each item code are given in the HBS 1991/92 interviewers' manual.