ESTIMATING THE CONTRIBUTION OF INFORMAL SECTOR ACTIVITY TO THE GROSS DOMESTIC PRODUCT OF GHANA

H. Coulombe, A.D. McKay and J.I. Round

No.449

WARWICK ECONOMIC RESEARCH PAPERS

THE UNIVERSITY OF WARWICK

DEPARTMENT OF ECONOMICS

ESTIMATING THE CONTRIBUTION OF INFORMAL SECTOR ACTIVITY TO THE GROSS DOMESTIC PRODUCT OF GHANA

H. Coulombe*, A.D. McKay** and J.I. Round*

No.449

February 1996

H. Coulombe* & J. Round*
Department of Economics
University of Warwick
Coventry CV4 7AL
England

**A.D. McKay
Department of Economics
University of Nottingham
University Park
Nottingham NG7 2RD
England

ESTIMATING THE CONTRIBUTION OF INFORMAL SECTOR ACTIVITY TO THE GROSS DOMESTIC PRODUCT OF GHANA

H Coulombe, A D McKay and J I Round[1]

1. Introduction

The conceptual basis underlying the identification of household economic activity, and of informal sector activity in particular, suggests that estimates of value added ought to be derived at an economy-wide level for different categories of activity. The new 1993 SNA makes a distinction between household enterprises owned and operated by own-account workers, that is, *family enterprises*; and household enterprises of employers, which we refer to as *micro-enterprises*. But there is also a sectoral dimension, starting with the distinction between *farm* and *non-farm enterprises*, although a much finer disaggregation of activities is desirable for national accounts purposes.

This paper sets out a general methodology for deriving estimates of the gross domestic product of household economic activity in Ghana together with some estimates derived from the Ghana Living Standards Surveys (GLSS 1 and GLSS 2). It follows directly from two earlier papers, the first of which (McKay and Round, 1994) established the conceptual basis for the identification of household economic activity (that is, production activity) in Ghana, while the second (Coulombe, McKay and Round, 1994b) set out some sample estimates on the basis of GLSS 1 and GLSS 2 survey results. However, the methodology also relies on earlier work in which sample estimates of household incomes and expenditures were derived from the GLSS data (Coulombe, McKay and Round, 1993) and some identified differences in the sample results for GLSS 1 and GLSS 2 were analysed (Coulombe, McKay and Round, 1994a).

The methodology begins from the point where GLSS sample results of household economic activity have been obtained (Coulombe, McKay and Round, 1994b). There are then two stages in the methodology which follow. The observed shortfall in the

---

[1] This is a revised version of a paper presented in a workshop on 'Measuring Informal Sector Activity in Ghana' held in Accra, January 1995, and based on a project financed by the Overseas Development Administration. It is to be published by the Ghana Statistical Service as Chapter 4 in a monograph with the same title. All results are provisional and should not be quoted without proper attribution.

estimates of total household income relative to total household expenditure suggests that at least some components of income may be underestimated and/or some components of expenditure may be overestimated. However, there are good reasons for believing that the vast majority of the shortfall is a consequence of an underestimation of income (Coulombe, McKay and Round, 1994a). Thus at the first stage a set of adjustment factors are derived so as to compensate for possible underrecording of income in the sample estimates. At the second stage the (adjusted) sample estimates then have to be grossed up to represent the population as a whole. For this we rely on reciprocal sampling fractions to represent the grossing up factor.

The outline of the paper is as follows. After briefly reviewing, in section 2, the aggregate income and expenditure components that have been identified in earlier work, section 3 focuses on an analysis of the identified shortfall of household income relative to household expenditure. This is important, not only as an interesting independent statistical analysis, but it also to form an integral part of the adjustment methodology. Section 4 describes a range of proposed methodologies for deriving adjustment factors to account for the underrecording of income. Sets of these adjustment factors are then presented and discussed in section 5. Then, in section 6, the adjusted GLSS sample estimates are multiplied by grossing up factors so as to provide some preliminary estimates of all-Ghana informal sector GDP. Finally, section 7 concludes by reviewing the strengths and weaknesses of the methodology that has been used and offers some reflections on the overall results achieved.

2. Household economic activity and the GLSS

Income and expenditure components

It will be useful to consider the measurement of household production activity within the context of a set of household income and expenditure accounts. In general households derive their incomes from a variety of sources and make a range of expenditures and other outlays from this income, leaving savings as the component which balances household incomes and outlays. We shall leave to one side all considerations about what constitutes income and expenditure, the distinction between current and capital items, and therefore what should or should not be included in the accounts. It can be noted that in our earlier work (Coulombe, McKay and Round, 1993) the following aggregates (or components) of household income and expenditure were considered sufficient to provide an adequate framework for the estimation of various income and expenditure totals. As already noted, household savings is the balancing item for the household current accounts.

Income

1. Employment (employee compensation)
2. Household agriculture
3. Non-farm self-employment
4. Rent
5. Remittances received
6. Other income

Expenditure

7. Food (actual)
8. Housing (actual and imputed)
9. Other expenditure (actual)
10. Food (imputed)
11. Other expenditure (imputed)
12. Remittances paid out

These components are broadly in line with those referred to in United Nations (1991, pp 13-18), and with the current accounts of the household sector specified in the 1993 SNA, although obviously some regrouping of the underlying subaggregates is necessary for complete accordance. They have hitherto proved to be suitable for a range of descriptive analyses and are useful as a framework for this study too.

As already indicated, households engage in informal sector activity in a variety of ways. However, the majority of the income from household economic activity classified as informal sector activity shows up in two income categories in particular: 'household agriculture' and 'non-farm self-employment'. Obviously some income will also be received by household members who are engaged as employees in micro-enterprises owned and operated by other households. However, as already explained (McKay and Round, 1994), in order to derive economy-wide estimates of household production activity the relevant information base should be confined to informal activities included in component 2 (agricultural activity) and component 3 (non-farm activity), with additional allowance for the employee compensation *paid out* by micro enterprises, usually to members of other households. In other words the inclusion of component 1 (employee compensation) *received* by households from (informal sector) micro enterprises could lead to some double counting of income and product under the sampling scheme. Examining this more closely, it can be seen that two possibilities may arise. On the one hand employee compensation received by one household in the sample may have been included in the enterprise income generated by some other household in the sample. On the other hand, even if the enterprise paying the employee compensation is excluded from the sample, the sample design ought to have accommodated enterprise activity representative of that type elsewhere in the sample. However, in spite of our need to focus on just two of the household income components, it will be shown that the remaining aggregates are also required to implement the general methodology developed in subsequent sections.

Alternative estimates of the income components

Our earlier discussion of the GLSS sample results (Coulombe, McKay and Round, 1994b) outlined some of the problems encountered in deriving the estimates of the components. In particular we noted that three alternative estimates of non-farm enterprise income and two estimates of household agricultural income could be obtained from different responses and different sections of the survey. It will be

useful to review briefly what are these alternative estimates in order to assist in subsequent discussion.

1. *Non-farm enterprises*

Net revenue (NFSEY1) defined as total revenue minus total current costs. Total revenues consist of payments in cash or kind as well as the value of any output consumed domestically. The costs comprise total current input expenditures, excluding at this stage any allowances for depreciation.

Profits (NFSEY2) defined as revenue used for household purposes plus retained 'profits'.

Earnings (NFSEY3) defined as the total self-employment income derived from the activities module (section 5) but limited to informal sector activities.

2. *Farm enterprises*

Net revenue (HHAGINC1) defined as the difference between total revenues and total costs can be estimated separately for crops, food products from homegrown crops, livestock, and animal products.

Earnings (HHAGINC2) defined as self-employment income derived from the activities module but limited to informal sector activities.

Comparison between estimates

As already indicated there is no sound conceptual basis for choosing between the alternative estimates. Clearly the alternatives do not necessarily measure precisely the same quantity; thus, enterprise profit, self employment income and net revenue are each different concepts and are very unlikely to record identical estimates.

Tables 2.1, 2.2, and 2.3 record some summary statistics relating to the alternative estimates. Table 2.1, for instance, shows the mean incomes calculated across those households recording a receipt of the appropriate enterprise income. It is immediately noticeable from Table 2.1 that, for all households recording non-farm enterprise

Table 2.1: Conditional means

	GLSS 1	GLSS 2
Non-farm enterprise		
NFSEY1	-208430	-161509
NFSEY2	118288	123464
NFSEY3	122529	151507
Agriculture		
HHAGINC1	161385	124082
HHAGINC2	140245	124445

Table 2.2: Standard deviation

	GLSS 1	GLSS 2
Non-farm enterprise		
NFSEY1	998205	1017531
NFSEY2	154524	177735
NFSEY3	217280	268273
Agriculture		
HHAGINC1	314705	154578
HHAGINC2	167734	143038

Table 2.3: Percentage of households recording positive income

	GLSS 1	GLSS 2
Non-farm enterprise		
NFSEY1	39.1	42.4
NFSEY2	100.0	100.0
NFSEY3	100.0	100.0
Agriculture		
HHAGINC1	94.9	94.1
HHAGINC2	100.0	100.0

activity, the mean value of estimated net revenue (conditional mean NFSEY1) across those households is negative. Indeed, Table 2.3 shows that around 60 per cent of households show negative net revenues (60.9 per cent in GLSS 1 and 57.6 per cent in GLSS 2). By contrast, for household agricultural enterprise income, the conditional mean HHAGINC1 is positive, and only about 5 per cent of households show negative net revenue (Table 2.3). It should be noted that in all cases the standard deviations are large (Table 2.2). While there is undoubtedly a large amount of variation of enterprise income across households, standard deviations (and means) are bound to be affected by outliers. However, in deriving the aggregates, some outliers have already been removed and re-estimated and this will already be reflected in these summary statistics. Obviously, for both agricultural and non-farm enterprise activity, the 'profits' and 'earnings' estimation methods necessarily record positive incomes.

The earlier summary statistics can be augmented with Pearson correlation coefficients calculated for pairs of alternative estimates. The correlation coefficients between HHAGINC1 and HHAGINC2 are reported in Table 2.4.

Table 2.4: Correlation coefficients between household agricultural income estimates (HHAGINC1 and HHAGINC2)

	GLSS1	GLSS2
Full sample	0.528	0.825
Outliers removed	0.678	0.825

In spite of the fact that over 5 per cent of households have reported negative HHAGINC1 estimates the correlation between estimates for the full sample is quite high, and is even higher in the case of GLSS 2 than GLSS 1. However a plot of the correlations for both non-farm and household agricultural incomes identified some household observations as outliers, and these may have distorted the correlation coefficient values. Table 2.4 shows that after removing two outliers from the GLSS 1

sample and three from GLSS 2 the correlation coefficient for GLSS 1 rises to 0.678 although the coefficient for GLSS 2 remains unchanged.

A similar set of correlation coefficients between the estimates of non-farm income yielded the results shown in Table 2.5.

Table 2.5: Correlation coefficients between non-farm income estimates

	GLSS1	GLSS2
Full sample		
NFSEY1 and NFSEY2	-0.305	-0.154
NFSEY1 and NFSEY3	-0.141	-0.101
NFSEY2 and NFSEY3	0.558	0.523
Outliers removed		
NFSEY1 and NFSEY2	-0.367	-0.181
NFSEY1 and NFSEY3	-0.128	-0.118
NFSEY2 and NFSEY3	0.557	0.522

In this case the correlations are generally much weaker and highlight a particular problem with the NFSEY1 estimates. In each sample not only do more than half the households record negative incomes but Table 2.5 also indicates that this is unlikely to be accounted for simply by a scale factor, that is, an underrecording of non-farm income in all households. This is because the correlation coefficient between NFSEY1 and each of the NFSEY2 and NFSEY3 aggegates is negative in both GLSS 1 and GLSS 2 - whether these aggregates are measured with or without outliers. On the other hand the correlation coefficient between NFSEY2 and NFSEY3 is positive and is relatively high.

The choice of which pair of estimates to use is not an easy one to make. In earlier uses of the survey results different combinations have been selected in different

circumstances. For example, for the derivation of the income and expenditure aggregates (Coulombe, McKay and Round, 1993) the choice was HHAGINC1 and NFSEY3, while for the descriptive analysis (Coulombe, McKay and Round, 1994b) the choice was HHAGINC1 and NFSEY2. Clearly there would also be a good case for using HHAGINC2 and NFSEY3 as these are consistently defined and yield non-negative estimates throughout the sample. However, the case against using HHAGINC2 and NFSEY3 is that they are individual-based rather than enterprise-based estimates, and may not therefore be appropriate for all purposes. Furthermore, difficulties can be expected to arise in translating individual responses on self employment income into accurate and clearly defined components of household income. In summary, therefore, it is difficult to make an optimum choice: no one pair of estimates appear to have a clear superiority to any other pair. For present purposes it was decided to use HHAGINC2 and NFSEY2, although in some of the ensuing discussion some comparisons alternative estimates will help to ascertain the robustness of results obtained.

3. Income-expenditure discrepancies in the GLSS samples.

Overall discrepancies between incomes and expenditures

In many household surveys there is a tendency for incomes to be under-recorded relative to expenditures. In common with other Living Standards Surveys, earlier estimates derived from the GLSS surveys suggested that a similar phenomenon exists with these surveys too. Tables 3.1, 3.2, 3.3 and 3.4 summarise some results on the basis of the GLSS. In all four tables the farm income component is based on the earnings estimate (HHAGINC2). Table 3.1 shows the mean household total income and total expenditure for GLSS 1 under two alternative combinations of the estimates, as well as the shortfall of income relative to expenditure, which is referred to throughout as the 'discrepancy'. The first panel of the table shows the results based on the non-farm income component estimated from profits (NFSEY2), while in the second panel this component is estimated from earnings (NFSEY3). Table 3.2 shows the corresponding results for GLSS 2. Tables 3.3 and Table 3.4 are similar to Table 3.1 except that those components which are common to both incomes and expenditures (imputed items of production for own consumption, including rents on owner-occupied dwellings) have been removed before the mean incomes, expenditures and discrepancies are calculated.

The term 'discrepancy' needs to be heavily qualified for the ensuing analysis. There is no allowance made for household savings beyond that which arises when estimated household income exceeds expenditure. The information on savings available from the GLSS was considered too meagre and unreliable to be of use. The positive discrepancies, which suggest that, on average, incomes are *less* than expenditures, and which occur in all panels of both tables, are obviously too large and too universal to be attributable to dissaving alone. So an underrecording of income in at least some of the components is the only possible acceptable explanation, as it is highly unlikely that such a widespread overrecording of expenditure could have occurred. Clearly, the underrecording of income could materially affect our use of these estimates in measuring household production activity, and informal sector activity in particular. Therefore it is important to try to attempt to ascertain which income components are affected, and to what extent, and hence to try to generate some correction factors to apply to the sample results.

Table 3.1: Income-Expenditure Discrepancies (including imputations): GLSS 1

(i) NFSEY2

	Accra	Other Urban	Rural Coastal	Rural Forest	Rural Savannah	Ghana
Income (NFSEY2)	211244	219081	191882	197716	273104	217977
Expenditure	452480	335906	289106	267467	299172	314534
Discrepancy	241236	116825	97224	69751	26068	96557
Discrepancy (%)	53.3	34.8	33.6	26.1	8.7	30.7
Saving < 0 (%)	90.7	80.4	81.6	78.5	69.4	79.1

(ii) NFSEY3

	Accra	Other Urban	Rural Coastal	Rural Forest	Rural Savannah	Ghana
Income (NFSEY3)	187299	213263	172123	195298	269386	209513
Expenditure	452480	335906	289106	267467	299172	314534
Discrepancy	265180	122643	116983	72169	29785	105021
Discrepancy (%)	58.6	36.5	40.5	27.0	10.0	33.4
Saving < 0 (%)	90.7	82.7	87.3	80.9	70.4	81.5

Table 3.2: Income-Expenditure Discrepancies (including imputations): GLSS 2

(i) NFSEY2

	Accra	Other Urban	Rural Coastal	Rural Forest	Rural Savannah	Ghana
Income (NFSEY2)	294608	251894	206410	208091	253523	236378
Expenditure	502448	386684	332046	295700	308156	349285
Discrepancy	207840	134789	125636	87609	54632	112907
Discrepancy (%)	41.4	34.9	37.8	29.6	17.7	32.3
Saving < 0 (%)	85.9	81.4	84.4	82.3	75.0	81.5

(i) NFSEY3

	Accra	Other Urban	Rural Coastal	Rural Forest	Rural Savannah	Ghana
Income (NFSEY3)	301822	230668	231199	202671	254697	234186
Expenditure	502448	386684	332046	295700	308156	349285
Discrepancy	200626	156015	100847	93029	53458	115099
Discrepancy (%)	39.9	40.3	30.4	31.5	17.3	33.0
Saving < 0 (%)	86.2	85.2	82.5	83.6	77.8	83.1

Table 3.3 Income-Expenditure Discrepancies (excluding imputations): GLSS 1

(i) NFSEY2

	Accra	Other Urban	Rural Coastal	Rural Forest	Rural Savannah	Ghana
Income (NFSEY2)	195720	169063	117585	95845	102958	130661
Expenditure	436955	285888	214809	165596	129026	227218
Discrepancy	241236	116825	97224	69751	26068	96557
Discrepancy (%)	55.2	40.9	45.3	42.1	20.2	42.5
Saving < 0 (%)	90.7	80.4	81.6	78.5	69.4	79.1

(i) NFSEY3

	Accra	Other Urban	Rural Coastal	Rural Forest	Rural Savannah	Ghana
Income (NFSEY3	171775	163244	97826	93426	99241	122196
Expenditure	436955	285888	214809	165596	129026	227218
Discrepancy	265180	122643	116983	72169	29785	105021
Discrepancy (%)	60.7	42.9	54.5	43.6	23.1	46.2
Saving < 0 (%)	90.7	82.7	87.3	80.9	70.4	81.5

Table 3.4 Income-Expenditure Discrepancies (excluding imputations): GLSS 2

(i) NFSEY2

	Accra	Other Urban	Rural Coastal	Rural Forest	Rural Savannah	Ghana
Income (NFSEY2)	274369	210898	135094	124593	103762	160887
Expenditure	482209	345687	260730	212203	158394	273794
Discrepancy	207840	134789	125636	87609	54632	112907
Discrepancy (%)	43.1	39.0	48,2	41.3	34.5	41.2
Saving < 0 (%)	85.9	81.4	84.4	82.3	75.0	81.5

(i) NFSEY3

	Accra	Other Urban	Rural Coastal	Rural Forest	Rural Savannah	Ghana
Income (NFSEY3)	281583	189672	159883	119174	104936	158695
Expenditure	482209	345687	260730	212203	158394	273794
Discrepancy	200626	156015	100847	93029	53458	115099
Discrepancy (%)	41.6	45.1	38.7	43.8	33.8	42.0
Saving < 0 (%)	86.2	85.2	82.5	83.6	77.8	83.1

The size and pattern of the discrepancies shown in Tables 3.1 to 3.4 are clearly significant. For Ghana as a whole, and measuring discrepancy as a percentage of total expenditure, it can be seen that overall discrepancies are about 30 to 33 per cent in both years regardless of which estimation method is used for the non-farm income component. Also, the magnitude and pattern of income - expenditure discrepancies changes little if the NFSEY2 estimate is used instead of NFSEY3. However a disaggregation by locality reveals much more variation in discrepancies, there being even more regional variation on the basis of GLSS 1 than on GLSS 2. In both surveys the discrepancies appear to be significantly higher in households from urban areas than from rural areas. Clearly this regional variation will be due to intrinsic regional features and characteristics. Some of this will be associated with the regional mix of income components in the sense that the individual components might be under-recorded to different degrees thereby affecting the overall discrepancy in the localities to different extents.

Analysis of discrepancies

To investigate the pattern of discrepancies further, the discrepancies at the household level have been analysed by conducting a series of regressions in which a range of explanatory variables were selected, including both locality and SEG dummies and income-related variables. For this and all our subsequent analyses it proved useful to confine our income variables to just four categories of income as follows:

(i) employee compensation,
(ii) household agricultural income,
(iii) non-farm business income, and
(iv) all other income.

The last category is simply an aggregation of the rent, remittance and other income categories from the income aggregates set out earlier. It will therefore constitute an especially heterogeneous category of income sources. However, one advantage of this classification is that it provides a reasonable balance in terms of the proportions of income derived from the four sources across the household sector as a whole.

The first aim of this analysis is to try to ascertain which variables, related to household characteristics, best explain the variation in the size of the discrepancies. Throughout the analysis discrepancies are measured as 'expenditure minus income' and are

expressed as a percentage of expenditure. As noted earlier the preferred choice of estimates of the household agricultural and non-farm income components in all cases was HHAGINC2 and NFSEY2, primarily because these ensure that income estimates for both of the components are non-negative, and hence is non-negative for household income as a whole.

For each of GLSS 1 and GLSS 2 two separate data sets have been defined. Incomes and expenditures have been measured inclusive and exclusive of those imputed items common to each side of the household accounts. Clearly, although the treatment of imputations will not affect the size of the discrepancy in absolute terms it will affect our measure in percentage terms at the household level. To pursue this analysis further a series of regressions were carried out in which household discrepancies were separately regressed on dummy variables representing locality, socioeconomic group (SEG) and time of interview (in terms of quarter, or 'season'). Socioeconomic groups are defined according to which main source of income among the four listed above constitute the main source for each household, regardless of how large is the proportion of total income it represents. The seasonal dummy was introduced to try to establish whether discrepancies exhibited a significant seasonal pattern.

Regressions containing only dummy variables are equivalent to carrying out a one-way analysis of variance, and as each of the regressions contain three dummy variables, representing a four-way classification of the sample, a direct comparison can be made between the explained sum of squares in each case. Table 3.5 shows that the between group variation is

Table 3.5: R^2 values for alternative household groupings

	GLSS1		GLSS2	
Imputations	included	excluded	included	excluded
Locality	0.026	0.019	0.010	0.009
Socioeconomic group	0.033	0.018	0.017	0.017
Season	0.007	0.008	0.003	0.006

low relative to the total variation but is nevertheless statistically significant in all cases. Therefore it can be concluded that the locality and SEG groupings explain relatively more of the total variation than do the seasonal groupings. Table 3.6 shows comparable results to those in Tables 3.1 to 3.4 in respect of a disaggregation of the sample by socioeconomic group instead of by locality. It suggests a substantially lower discrepancy in those households primarily dependent on non-farm self-employment income relative to other groups, and that this is largely independent of whether imputations are included or excluded from the accounts.

Further analyses of the factors influencing the discrepancies were carried out although the detailed results are not reported here. For example, the inclusion of household total income significantly increased the explained variation (the adjusted R^2 increases from around 0.02 to 0.22 in the case of the regression for GLSS 2 with SEG dummies, inclusive of imputations) and there is also a suggestion that the degree of household dependence on a mixture of income sources rather than a sole income source may also be a relevant factor in explaining income - expenditure discrepancies.

If it is indeed the case that income is being under-reported on a wide scale then, in the light of the above results, it is possible to set out a number of conjectures, although it is more difficult to translate these into testable hypotheses. One possibility, for example, is that households which claim to be primarily dependent on wage (that is, employment) income and are classified as such in Table 3.6, may actually be underreporting, say, enterprise income. If this had been captured the this would not only reduce (or even eliminate) the discrepancy of certain wage-earning households but it may have shifted them into another SEG category altogether.

One conclusion that can be drawn from the above analysis is that income source may be an important factor in 'explaining' the size and pattern of income-expenditure discrepancies across households. It helps us to ascertain whether any particular income source is more strongly associated with the observed discrepancies than any other and hence whether this income is more likely to be under-recorded. The significance of the SEG variable in the analysis of variance does suggest that the household's main income source matters and we shall exploit this result in seeking to estimate adjustment factors so as to correct for the underestimation of household income in the sample.

Table 3.6: Income-Expenditure Discrepancies, by socioeconomic group

(i) GLSS 1 (including imputations)

	Wage	Agric	Non-farm	Other	Total
Income (NFSEY2)	182344	215883	269877	150816	217977
Expenditure	343380	291068	352003	284811	314534
Discrepancy	161036	75185	82126	133995	96557
Discrepancy (%)	46.9	25.8	23.3	47.0	30.7
Saving < 0 (%)	84.0	81.1	68.8	86.2	79.1

(ii) GLSS 2 (including imputations)

	Wage	Agric	Non-farm	Other	Total
Income (NFSEY2)	227946	215647	299987	173669	236378
Expenditure	394197	312468	400700	285372	349285
Discrepancy	166251	96822	100713	111703	112907
Discrepancy (%)	42.2	31.0	25.1	39.1	32.3
Saving < 0 (%)	86,4	84.4	72.8	81.8	81.5

(iii) GLSS 1 (excluding imputations)

	Wage	Agric	Non-farm	Other	Total
Income (NFSEY2)	163498	78421	218163	133983	130661
Expenditure	324534	153606	300289	267978	227218
Discrepancy	161036	75185	82126	133995	96557
Discrepancy (%)	49.6	48.9	27.3	50.0	42.5
Saving < 0 (%)	84.0	81.1	68.8	86.2	79.1

(iv) GLSS 2 (excluding imputations)

	Wage	Agric	Non-farm	Other	Total
Income (NFSEY2)	202755	93084	248795	151173	160887
Expenditure	369006	189906	349508	262876	273794
Discrepancy	166251	96822	100713	111703	112907
Discrepancy (%)	45.1	51.0	28.8	42.5	41.2
Saving < 0 (%)	86.4	84.4	72.8	81.8	81.5

4. A methodology for deriving adjustment factors

Our interest here is to seek a way of adjusting the sample estimates of household income and its components so as to obtain a more credible set of estimates in the specific sense that they are more consistent with the level of expenditure at the household level. The foregoing analysis suggests that the adjustment factors should vary by component. However the only information that is available for assessing the degree of underrecording of income is to observe the income-expenditure discrepancies for individual households.

The information available from the Ghana household surveys relate almost entirely to expenditures, incomes and household production. Many, if not most, households are likely to save some of their income. However there is no reliable (or even usable) information on the level of household savings in the GLSS1 or GLSS2 surveys[2]. Therefore, the best that can be achieved at the micro level is to determine those scale factors which would (on average) at least equalise household incomes with our estimates of household expenditures, and hence to disregard household savings, at least for the GLSS1 and GLSS2 data sets on which we rely.

Formally, the problem can be stated as one where the income for household i (Y_i) which is initially

$$Y_i = \Sigma X_{ik} \qquad (1)$$

and where, for most households reported income is less than reported expenditure, i.e.

$$Y_i < E_i \qquad (2)$$

The aim is to estimate a set of scale factors β_{ik} in order to generate revised household income estimates Y_i^* such that

$$Y_i^* = \Sigma \beta_{ik} X_{ik} \qquad (3)$$

and which satisfy the desirable condition that

$$Y_i^* \geq E_i \qquad (4)$$

---

[2] There is more information on household savings in GLSS3.

Clearly, it is infeasible to generate scale factors that are both household- and component-specific, but it does serve to establish the general problem and helps to set out a way to proceed.

Method 1

The first method is simply to accept both the level and pattern of household incomes derived from the sample without any further adjustment. This obviously implies that

$$\beta_{ik} = 1 \qquad \text{all i and k} \qquad (5)$$

which violates the condition that incomes should, on average, at least equal expenditures ((4) above), and, in consequence, it must raise questions about the integrity of the expenditure estimates. Formally, therefore, method 1 simply establishes a 'no adjustment' benchmark.

Method 2

The next most obvious method is to scale up all income components by the same amount in those households where total income falls short of total expenditure. At the individual household level this means defining β_i where

$$\begin{aligned} \beta_i &= E_i/Y_i \qquad \text{if } Y_i < E_i \\ &= 1 \qquad \text{if } Y_i > E_i \end{aligned} \qquad (6)$$

This means that sample values of all income components for each household i would be scaled by the same factor. Thus the pattern - but not the level - of incomes at the individual household level is maintained. An alternative would be to aggregate across certain household groups, say urban and rural households, localities, or SEGs, and to determine *average* scale factors applicable to all income components across all households within those groups. There is no particular advantage in grouping households, however, as the method can be applied equally well at the individual household level.

The main disadvantage of this method is that no account can be taken of the evidence described in the previous section which suggests that income components may be

under-recorded to different extents. Hence, it is preferable to seek estimates of component-specific rather than household-specific scale factors as determined by this method.

Method 3

An alternative method is to estimate β_k, on the basis of the condition that, on average, household incomes are scaled to equal (or exceed) household expenditures. Formally

$$E_I = \Sigma \beta_k X_{Ik} \qquad (7)$$

where I is a group of households in the sample so that, for example, E_I is either the total expenditure of the group or, equivalently, it might be defined as the mean total expenditure of that group. Clearly if 'I' were to be defined to be all the households in the sample then there is no unique solution to the problem of estimating β_k. For instance if there are four income groups (k = 1,...,4) then the problem reduces to the solution of one equation in four unknowns. However this does suggest a viable way to proceed.

Suppose the sample of households is subdivided into four mutually exclusive groups (I = 1,...,4). If E' = [E_1, E_2, E_3, E_4] is a vector of total expenditures of each household group, and X is a matrix of 'group by component' incomes where, for example, X_{Ik} is the income of component k received by household group I, then

$$E = X \beta \qquad (8)$$

where β' = [β_1, β_2, β_3, β_4] is the set of unknown scale factors necessary to equate group incomes to match group expenditures.

Providing household groups are chosen so that between-group income patterns differ then X is not only a square matrix but is non-singular so that β may be derived (uniquely) as follows

$$\beta = X^{-1} E \qquad (9)$$

Other than ensuring that X is non-singular the choice of the four income groups is open. However, it is interesting to conjecture whether or not the values of the

adjustment factors, β, are sensitive to the choice of household groups. In line with the analysis of discrepancies discussed in the previous section one way of grouping households is according to socioeconomic group. This appears to capture the most between-group variation in household discrepancy among the factors considered.

Method 3 can be applied separately to each of the GLSS 1 and GLSS 2 samples, or to a combined sample. Alternatively it could be applied to subsets of the sample so as to derive, say, locality-specific, or season-specific sample adjustment factors if there is good evidence for believing that these might differ substantially from those at an all-Ghana level.

Method 4

In assessing the relative size of household income and expenditure (equations (1) to (3)) it will be recalled from section 2 that certain items will be common to both the income and expenditure sides of the accounts at the household level. These items are those that are produced and consumed within the household and are therefore simultaneously included as an income and expenditure. It will include rent on dwellings of owner-occupiers, but it will also include imputations of items produced and consumed within the household, as part of agricultural or non-farm enterprise activity.

If these items are removed then, in absolute terms, the income-expenditure discrepancies will be unaffected, although discrepancies expressed as a percentage of (revised) expenditures will increase. Also, it is reasonable to expect that any sample adjustment factors should be derived from (and applied to) only those parts of the income components which are not imputed, because any discrepancy must be the result of a shortfall between 'actual' income and 'actual' expenditure. Method 4 is therefore a modified version of method 3 in this regard.

Formally, a revised set of sample adjustment factors can be derived by excluding those imputed components that are common to both income and expenditure at the household level. Likewise, the resulting factors should only be applied to the 'actual' incomes recorded at the household level. Imputed items of income are multiplied by adjustment factors equal to one. Clearly, as these adjustments are applied to incomes at an individual household level the effective difference between methods 3 and 4 cannot be ascertained in advance of their application to a particular data set.

Method 5

Finally, sample adjustment factors can also be estimated in a variety of ways using regression methods. The general procedure is to regress total expenditure against the four income components as independent variables, excluding any intercept term, and treating households as independent sample observations. The slope coefficients can then be directly interpreted as sample adjustment factors. That is, they are the scale factors that would need to be applied to each income component in order to provide, in a least squares sense, the best overall fit of total income to total expenditure in the sample. As in the case of methods 3 and 4 above, the scale factors are derived to equate income and expenditure 'on average' so there is no guarantee that income will match expenditure at the individual household level.

In principle the regression method can be applied in several different ways. First, it can be applied to income and expenditure components inclusive of both actual and imputed items (c.f. method 3); second it can be applied to income and expenditure components exclusive of those (imputed) items common to both sides of the accounts for each household (c.f. method 4). Thirdly, it is possible to split the sample (by region or by SEG) and hence to derive separate estimates of the scale factors for each group so defined.

As is the case in methods 3 and 4 above, there is no guarantee that the estimated scale factors will be greater than one. In principle therefore a further alternative is to carry out least squares in which all b_k are constrained to be at least one. But this may be unnecessary if the unconstrained least squares procedure generates estimates that are in the admissible range. In any case, it could reasonably be argued that such a modification is over complex under the circumstances given that there is no clear methodological preference for it.

5. Estimates of the adjustment factors

Adjustment factors, which are applied to the sample estimates prior to any further analysis or grossing up of these estimates to the population (all Ghana) level, have been estimated according to each of the methods described in the previous section. Obviously *method 1*, which is to leave the sample estimates unchanged, requires no further explanation. But the other methods which each yield a whole range of estimates of adjustment factors, and their application can potentially make substantial differences to the population-wide estimates that ensue.

The adjustment factors derived by *method 2* are household-specific and therefore are not reported here, although some population estimates of the income components after applying method 2 adjustment factors to the sample estimates are reported in the next section.

Sample adjustment factors have been derived according to *methods 3 and 4* and are based on a number of variations of the basic methods. As already indicated, there is no compelling reason for choosing one household grouping rather than another. In principle, the method can formally apply to any aggregation of households providing they are classified into four groups. However, a classification based on socioeconomic groups (SEG) defined according to main income source does have a special attraction in view of the fact that these groups are so closely aligned to the income components to which the adjustment factors apply, as well as offering the greatest explanatory power in the analysis of discrepancies across households. The classifications, and the nature of the solutions one might expect to equation (9) which yield the adjustment factors, can be explored further.

There is a strong preference for the adjustment factors β_{ik} that eventually emerge from the methodology should be positive. They are, after all, to be applied as scale factors to income variables so negative values would be meaningless. However if we consider the simple mathematics of equation (9) which is the basis of methods 3 and 4 then, obviously, even with $X > 0$ and $E > 0$ there is no general guarantee that $\beta > 0$. Recall that X is defined as a matrix of income payments by component to household groups. Therefore, in the case where the groups are defined according to main source of income (SEG) the matrix X has a dominant diagonal, and this is sufficient to ensure that $\beta > 0$ in the application of methods 3 and 4. Similarly, for all practical purposes

and based on the correlation structure between income components and expenditures, method 5 also ensures $\beta > 0$.

There is a separate issue as to whether or not one would expect $\beta \geq 1$. At first sight, as total income is generally less than total expenditure across households then the adjustment factors for each income component might be expected to be greater than one. But it might be appropriate to scale one or more components downwards if there is reason to believe they have been overestimated in the sample results. Let us briefly consider some *a priori* possibilities in the light of the GLSS sample design and our previous analysis of the sample results (Coulombe, McKay and Round, 1994a).

There are two main grounds for expecting employee compensation to have been underestimated in the sample responses. First, the in the GLSS 1 and GLSS 2 surveys households were asked for details about their primary and secondary jobs only. If household members have had more jobs income received would not have been reported. This deficiency does not apply to GLSS3. Secondly, there is some difficulty, arising from the LSMS questionnaire design, in the treatment of responses to questions about jobs undertaken in the past seven days and occupations in the past twelve months. The line taken in deriving the income aggregates on which the current estimates are based is that the 'seven day'-based and 'twelve month'-based estimates should be considered as alternatives (Coulombe, McKay and Round, 1993). Differences between the reported primary and secondary occupations in the two cases will affect some households, as it will depend crucially on when the interview took place. Therefore a bias will be introduced if the estimates are combined, and in any case total employee compensation may be under-recorded whether 'seven day', 'twelve month' or combined estimates are used. In this case the 'twelve month' estimates were used so as to be entirely consistent with related analyses and other uses of the GLSS results.

The component 'other income' is clearly subject to substantial underrecording, and hence underestimation, as it is a catch-all for incomes not listed or included in specific questions elsewhere. The two remaining components 'household agricultural income' and 'non-farm enterprise income' are our main concern for the purpose of estimating informal sector activity. For these components our *a priori* expectations are more uncertain. It might be expected that these incomes are universally under-recorded, either because of evasion, missing items in the responses from which the household production accounts are assembled, or again because the survey limited the number of business activities that could be reported upon. However, a previous analysis of the

sample results from GLSS 1 and GLSS 2 (Coulombe, McKay and Round, 1994a) noted a significantly larger decrease in both the share and absolute levels of reported household agricultural income than would have been expected over a twelve month period. This could have been due to differences in the classification of activity (or incorrectly classified activity) but this reason is hard to accept in view of the clear distinction drawn between agricultural and non-farm activities in different sections of the questionnaire. It should also be noted in this context that there is some evidence, notably from the GLSS3 results, that the contribution of agricultural income to total income is more accurately represented by GLSS 2 than by GLSS 1, which has led us to prefer to base our estimates on GLSS 2 for 1988/89.

Methods 3, 4 and 5 for deriving sample adjustment factors described in the previous section have been applied to different sample sets and some results are shown in Table 5.1. The columns of the table refer to the different sample sets and it can be seen that two broad sets of results are reported: those for GLSS 1 and those for GLSS 2. The first column in each case (that is, columns 1 and 5 in the table) show the results for method 3. This means that the adjustments have been derived for all four income components where all incomes and expenditures for each household are measured inclusive of both actual and imputed items. The adjustment factors are all greater than one, with the highest factor being associated with the 'other income' component in GLSS 1 and 'employee compensation' in GLSS 2. The estimated adjustments for both of these components in both surveys are high, although the adjustments for the two components of immediate concern here are relatively modest. It is particularly striking to note the similarity between these two columns (columns 1 and 4) in terms of their orders of magnitude and their relative patterns.

Columns 2 and 6 in Table 5.1 show what adjustment factors result if *method 4* were applied, in the specific case of (at least) equating incomes and expenditures and excluding imputed items from the calculation. The argument here is that as the imputed items are added to each side of the income-expenditure equation the adjustment should not apply to them. Any underrecording of these items will not affect the size of the discrepancies in absolute terms. Again the comparison between GLSS 1 and GLSS 2 suggests that the estimates are close. In this case, although the adjustment factors for 'employee compensation' and 'other income' are similar to the previous estimates the adjustment factor for 'household agriculture' rises dramatically.

Columns 3 and 7 of Table 5.1 show the values of adjustment factors that arise if only the panel elements of GLSS 1 and GLSS 2 are used. As these represent essentially

Table 5.1: Sample adjustment factors estimated on the basis of SEG household groups (Methods 3, 4 and 5)

	GLSS 1				GLSS 2			
	Method 3	Method 4	Method 4 (panel)	Method 5	Method 3	Method 4	Method 4 (panel)	Method 5
Employee compensation	2.030	2.053	1.998	1.562	1.834	1.870	1.891	1.558
Household agriculture	1.292	2.161	2.306	1.053	1.432	2.383	2.161	1.105
Non-farm business income	1.190	1.246	1.239	0.953	1.238	1.287	1.208	1.059
Other income	2.071	2.128	2.338	1.697	1.702	1.750	1.971	1.334

common samples of households they might be expected to indicate some intrinsic features of the differences between GLSS 1 and GLSS 2. The adjustment factors for 'employee compensation' and 'non-farm enterprise income' are very close, although the corresponding factors for 'household agriculture' and especially 'other income' show some decline over time. Hence the results do suggest that there may be some unexplained changes in the level and patterns of either incomes or expenditures, or both, between GLSS 1 and GLSS 2, as revealed by the differences in the adjustment factors for all components. Some of this might be due to the probable relative overestimation of household agricultural income in GLSS 1, and is our prime reason for concentrating on the GLSS 2 results in generating GDP estimates in the next section.

Methods 3 and 4 were also applied to groupings of households by locality and by season, in order to check on the sign, size and robustness of the adjustment factors. The results showed a considerable variation in the values of the adjustment factors, many of which were implausible, and are not reported here. Therefore the grouping by SEG not only has intrinsic appeal on *a priori* grounds but also generates consistent and reasonably plausible results.

Finally, columns 4 and 8 show the results obtained by applying method 5, the basic regression method, to the GLSS 1 and GLSS 2 data sets. The most striking observation is that these estimates are considerably lower than those obtained under methods 3 and 4. As method 5 is applied to the full set of items of income and expenditure (i.e. inclusive of imputed items) the results are most directly comparable with those for method 3. The relative magnitude of the scale factors for both methods is similar in both data sets. But the absolute size of the estimates is significantly lower in method 5 than in method 3. Indeed, the scale factor for 'non-farm business income' turns out to be less than one, and is therefore below the realistic lower bound. Further work is required in order to apply other variants of method 5 and, in particular, to income and expenditure data exclusive of imputed items. Present evidence does suggest, however, that the regression methods will consistently generate lower scale factors than simultaneous equation methods (methods 3 and 4) and it may well be that regression methods also have more desirable properties. All of this remains to be investigated in future work.

6. Estimates of household economic activity in Ghana

Two stage methodology

In principle, universe (or population) estimates of household economic activity can be derived simply by grossing up sets of sample estimates according to some appropriately chosen grossing-factor, such as the reciprocal sampling fraction. Indeed, this is the basis of the general methodology which is applied and discussed in this section. However, the previous discussion has shown that the sample estimates ought first to be adjusted in various ways so as to account for some possible underrecording of income on a component by component basis. Therefore an application of the general methodology potentially can generate a variety of estimates depending upon the choice of (i) which particular combinations of sample estimates of the income components are selected (that is, the choice of either NFSEY1, NFSEY2 or NFSEY3, combined with either HHAGINC1 or HHAGINC2); and (ii) which sample adjustment methodology is selected (that is, methods 1, 2, 3 , 4 or 5). As already suggested in section 2, we have a marginal preference for HHAGINC2 (household agricultural income) and NFSEY2 (household non-farm business income) mainly on the grounds that both of these sets of estimates are non-negative throughout the sample, and because NFSEY2 also has the advantage that it retains some enterprise-based features. Therefore the results that follow are based exclusively on these sets of estimates.

It should be recalled that our earlier analysis of the estimates of household incomes and expenditures revealed a marked change in the patterns of income (across income components) between GLSS 1 and GLSS 2. A comparative analysis of these results, in conjunction with comparable (preliminary) results for GLSS 3, suggested that GLSS 2 might be more reliable, both in terms of the levels and the patterns of incomes. Therefore, the results reported in this section will be confined to GLSS 2 (1988/89) and to the estimates for HHAGINC2 and NFSEY2 although sets of results could also be generated for GLSS 1 (1987/88) and based on alternative sample estimates.

In spite of limiting the number and range of alternative estimates that, in principle, could be derived, our application of the two-stage general methodology still gives rise to five alternative estimates according to which of the five sample adjustment methods is used. It will be recalled that Table 5.1 reports several alternative estimates of

sample adjustment factors relating to methods 3, 4 and 5. Two of these adjustment factors are relevant to and are utilised in the current application. These are the factors for 'household agriculture' and 'household non-farm enterprise activity' and estimates are confined to GLSS 2, and to methods 3 , 4 and 5. There has been no attempt to extend the regression methods further at this stage. It should also be recalled that method 2 adjusts the sample estimates at the individual household level while method 1 introduces no adjustment to the sample estimates whatsoever.

The final stage is a fairly straightforward grossing-up procedure. The grossing-up factor is determined on the assumption that the GLSS is a self-weighted sample and is representative of the population as a whole. The grossing-up factor is therefore calculated as the ratio of the estimated 1988 population (13.8 million) and the number of individuals included in GLSS 2 (1988/89) which is 15,369. This fraction is therefore 897.91.

Estimates at an aggregate level

Table 6.1 reports some aggregate results of the application of our methodology based on methods 1 to 5. The 'baseline' estimate of the value added of household (informal sector) non-farm activities at an all-Ghana level, in the case where the grossing-up factor is applied directly and where there is no adjustment of the sample estimates (method 1), is 218216 million cedis. The corresponding 'baseline' estimate of household agricultural activity is 339404 million cedis. Neither of these estimates makes any allowance for possible underrecording of income, and as they are the direct consequence of grossing-up the (unadjusted) sample estimates they are referred to here as 'baseline' estimates.

Table 6.1 also shows some disaggregation of these components. For non-farm enterprise activity the chosen disaggregation is between family enterprises (own-account workers) and microenterprises (employers). These categories are further disaggregated between fixed and itinerant family enterprises and between different sizes of microenterprises. Fixed-location family enterprises account for almost half of the baseline household non-farm enterprise income, while 69.6 per cent accrues from family enterprises as a whole. For household agricultural activity the disaggregation is according to categories of crop production, processing, animals (livestock), and a non-attributable category. The dominant contributor is crop production, with

Table 6.1: Estimate of (all Ghana) value added in household non-farm enterprises and household agricultural enterprises: alternative methodologies (in 1988/89, millions cedis)

(i) Non-farm enterprise income

	Method 1	Method 2	Method 3	Method 4	Method 5
Family-fixed	107525	162486	133116	134601	113869
Family-itinerant	44284	65669	54824	55891	46897
Micro<5	51654	75888	63947	65534	54702
Micro 5-9	11079	24059	13716	14084	11733
Micro>=10	3673	4948	4548	4593	3890
Total	218216	333050	270151	274704	231091

(ii) Agricultural income

	Method 1	Method 2	Method 3	Method 4	Method 5
Crops	29634	492184	424924	419839	333049
Processing	22668	41920	32460	40239	25442
Livestock	16435	27245	23536	25267	18447
Non-attributable	3567	7411	5108	8500	4004
Total	339404	568759	486027	493845	380942

processing and livestock each contributing less than 10 per cent of the baseline value added for this sector.

Before embarking on a more detailed examination of these and alternative estimates, it might be useful to compare the baseline estimates of household (informal) production activity with the latest available estimate of Ghana GDP, which is for 1988. GDP (purchasers value) for 1988 has been estimated to be 1,057,868 million cedis.[1] The earlier review of present National Accounts methodology regarding informal economic activity (Powell, Debra, Amable and Tonhie, 1995) suggested that coverage and measurement of such activity varies considerably from sector to sector. The most optimistic view which can be drawn from this is that informal sector is already 'covered' in the trade, construction, road transport, mining and manufacturing sectors although the 'quality' of the estimates may be poor. It can also be deduced that informal activity in the service sector may not even be adequately covered. On the other hand, a more pessimistic view can be posited that the current coverage of non-farm activity may, at best, include only micro-enterprises (small and medium scale enterprises) and therefore may exclude much of household 'own-account' enterprise activity. However, as regards agriculture, the current estimation procedures are quite different and it is likely that most activity is already covered although there may be a significant downward bias (an undermeasurement) in the estimates achieved.

Our baseline estimate of non-farm activity amounts to 20.6 per cent of GDP in 1988, while the equivalent percentage for agriculture is 32.1 per cent. Economy-wide value added estimates based on alternative estimation methods, that is, by applying sample adjustment factors derived by methods 2 to 5, are also summarised in Table 6.1. Method 2, in which incomes are scaled to at least match total expenditures at the individual household level, gives rise to the largest increase compared with the baseline estimate (method 1). In this case the estimate of household non-farm enterprise income would amount to 31.4 per cent of GDP, while the equivalent figure for agricultural income would be 53.8 per cent. Methods 3, 4and 5 yield estimates which lie between the percentages derived under methods 1 and 2 with method 5 giving the smallest adjusted increase over the baseline estimate.

---

1 Ghana Statistical Service (1989), Table 76.

Table 6.2: Household non-farm enterprise output by industrial sector
(in 1988/89, millions cedis)

(i) Method 1

	Agro	Mining/ Quarrying	Manufac- turing	Construc tion	Trade	Transport	Services	Total
Family-fixed	1665	1700	27930	354	68568	610	6699	107525
Family-itinerant	2665	769	4707	1670	27991	2974	3508	44284
Micro <5	1863	49	13497	2888	19091	9926	4340	51654
Micro 5-9	361	0	3902	209	3642	351	2625	11079
Micro >9	0	0	347	819	970	511	1026	3673
Total	6554	2518	50383	5941	120260	14372	18188	218216

(ii) Method 2

	Agro	Mining/ Quarrying	Manufac- turing	Construc tion	Trade	Transport	Services	Total
Family-fixed	2213	2848	42428	534	104278	751	9434	162486
Family-itinerant	2921	1926	7925	2565	41306	3958	5067	65669
Micro <5	2169	136	19788	4016	27573	16374	5832	75888
Micro 5-9	361	0	11474	504	7313	508	3899	24059
Micro >9	0	0	473	1491	970	701	1314	4948
Total	7664	4909	82088	9110	181440	22292	25547	333050

(iii) Method 3

	Agro	Mining/ Quarrying	Manufac- turing	Construc tion	Trade	Transport	Services	Total
Family-fixed	2061	2104	34578	439	84887	755	8293	133116
Family-itinerant	3299	952	5827	2068	34653	3682	4343	54824
Micro <5	2306	61	16710	3575	23634	12288	5373	63947
Micro 5-9	447	0	4831	259	4508	435	3237	13716
Micro >9	0	0	429	1014	1201	633	1271	4548
Total	8113	3117	62375	7354	148882	17792	22517	270151

(iv) Method 4

	Agro	Mining/ Quarrying	Manufac- turing	Construc tion	Trade	Transport	Services	Total
Family-fixed	1920	2186	35179	288	85651	753	8589	134566
Family-itinerant	3161	989	5811	2136	35454	3810	4510	55874
Micro <5	2309	63	17110	3714	24114	12769	5452	65532
Micro 5-9	437	0	4991	269	4610	452	3325	14085
Micro >9	0	0	446	1054	1114	658	1321	4593
Total	7828	3238	63538	7462	150943	18441	23196	274647

(v) Method 5

	Agro	Mining/ Quarrying	Manufac- turing	Construc tion	Trade	Transport	Services	Total
Family-fixed	1763	1800	29578	407	72614	646	7094	113869
Family-itinerant	2822	814	4984	1769	29642	3149	3715	46897
Micro <5	1973	52	14293	3148	20217	10512	4596	54702
Micro 5-9	382	0	4132	221	3857	372	2769	11733
Micro >9	0	0	367	867	1027	541	1987	3890
Total	6941	2667	53356	6292	127355	15220	19261	231091

Estimates at a disaggregated level

(i) Household non-farm enterprise income

Table 6.2 reports some more detailed estimates of household non-farm enterprise income with respect to the GDP contributions by sector in 1988. This table is directly comparable with the sample results shown in Coulombe, McKay and Round (1994b), Table 4.7. As noted in our discussion of the sample results, the sample size will permit only a modest disaggregation which is limited here to just seven sectors. It was also indicated that the sectoral classification of activity is unreliable. For example, some trade activities could have been allocated elsewhere, so too much credence should not be placed on the detailed results. Nevertheless, they suggest that, under each of the estimation methods, more than half of the informal sector output in this category stems from trade sector activity, which is an entirely credible result. Manufacturing, services and transport, together generate the majority of the remaining contribution to the output of non-farm enterprise activity.

(ii) Household agricultural income.

Table 6.3 shows more detailed estimates of the output of household agricultural activity. The table shows the all-Ghana estimates that are comparable with the sample estimates in Coulombe, McKay and Round (1994b), Table 5.4. The outputs attributable to different kinds of crops are shown at a very detailed level, although the sample size is not really large enough to sustain such a high degree of disaggregation, so the results must be treated with particular caution. However, it is worth noting that the proportionate contributions of individual crops to total crop output varies considerably in percentage terms according to the estimation method.

The second panel of Table 6.3 shows some detailed estimates for agricultural processing. Again, the estimates should be treated with some caution because of the relatively small sample size. There is a marked difference here between methods 3 and 4, which is obviously the result of wide variations between imputed and non-imputed items at the household level. No further disaggregation can be provided for the 'livestock' and 'non-attributable' categories.

Table 6.3: Household agricultural output (in 1988/89, million cedis)

	Method 1	Method 2	Method 3	Method 4	Method 5
Cocoa	33841	65084	48460	74219	37394
Coffee	44	170	63	104	49
Coconut palm	901	1604	1291	1600	996
Oil Palm	6337	9933	9075	10270	7002
Wood	131	896	187	299	144
Cola nut	1144	1921	1638	2630	1264
Kenef	37	90	53	35	41
Cotton	960	1131	1375	2223	1061
Peanut	6637	10082	9505	10395	7334
Tobacco	1128	1733	1615	2381	1246
Pineapple	1014	1980	1452	1515	1120
Sugarcane	838	1627	1201	1904	926
Plantains	24086	40962	34491	32808	26615
Bananas	2668	4944	3820	4187	2948
Oranges	2558	4687	3663	3880	2827
Other fruit	3390	5617	4854	3880	3746
Cassava	44737	79522	64063	56763	49434
Yam	24475	34745	35048	23395	27045
Cocoyam	16829	28262	24100	18473	18596
Potato	738	1246	1057	1004	815
Maize	47976	77296	68701	64807	53013
Rice	11205	20109	16045	20587	12382
Sorghum	28461	37050	40756	30354	31449
Tomato	8581	13794	12288	15169	9482
Okro	4308	7164	6169	5533	4760
Garden egg	2585	5069	3702	3360	2856
Beans	6398	9948	9161	8714	7070
Pepper	10092	17568	14452	14166	11152
Leafy vegetable	3870	6858	5542	4204	4276
Other vegeatble	663	990	950	861	733
Other crops	269	417	385	514	297
Total	296902	492499	425164	420237	328077

Export crops	53013	96251	75915	107577	58579
Food crops	243889	396248	349249	312660	269497
Total	296902	492499	425164	420237	328077

(ii) Processing

	Method 1	Method 2	Method 3	Method 4	Method 5
Garri	3381	6866	4841	6968	3736
Peanut	3722	6586	5330	6123	4112
Pito	1597	1964	2288	3093	1765
Maize flour	738	1234	1056	1758	815
Cassava flour	2585	6024	3701	6159	2856
Fufu	280	1385	400	666	309
Banku	654	862	937	1559	723
Achekie	5457	9440	7814	5457	6030
Kenkey	2434	4347	3486	4540	2690
Other	1820	3211	2607	3917	2011
Total	22668	41920	32460	40239	25048

Table 6.4: Actual and imputed income (in 1988/89, millions cedis)

(i) Imputed non-farm income

	Method 1	Method 2	Method 3	Method 4	Method 5
Family-fixed	13306	20678	16473	13306	14091
Family-itinerant	3911	5729	4842	3911	4142
Micro-enterprise	4374	6172	5416	4374	4633
Total	21592	32578	26731	21592	18866

(ii) Actual non-farm income

	Method 1	Method 2	Method 3	Method 4	Method 5
Family-fixed	94219	141808	116643	121259	99778
Family-itinerant	40373	59941	49982	51960	42755
Micro-enterprise	62032	98723	76795	79835	65691
Total	196624	300472	243420	253054	208224

(iii) Agricultural income

	Method 1	Method 2	Method 3	Method 4	Method 5
Imputed	227733	369573	326114	227733	251645
Actual	111671	199186	159913	266112	118260
Total	339404	568759	486027	493845	369905

(iii) Actual and imputed items

The estimation procedure permits some disaggregations of estimated value added according as to whether the output is 'actual' or has been imputed. Table 6.4 summarises the results at an aggregate level; more detailed results are available although again their reliability would be more doubtful. Considering the baseline estimates as the benchmark for comparison it can be seen that imputed income represents 9.9 per cent of total (informal) non-farm enterprise income, and 67.1 per cent of total agricultural income. So imputations account for only a small part of the likely increase in estimated GDP, as most of the agricultural income (imputed and actual) ought, in principle, to be included in GDP already. Departures from these baseline percentages for alternative estimation methods are quite dramatic, because of the way in which imputed items are treated in deriving sample adjustment factors (especially method 4).

Household formal and informal sectors compared.

All of the estimates presented and discussed so far relate to household *informal* sector activity. Although the majority of household sector production activity is defined as being informal, some activity in the non-farm enterprise sector has been classified as formal and has therefore been excluded from the sample before deriving the above estimates. However all household agricultural income is considered to have been derived from 'informal' activity according to our working definition. It is therefore interesting to see the extent to which households engage in formal sector production activity, as reported in the GLSS surveys. Table 6.5 summarises the estimates of informal and formal sector output for non-farm enterprises, classified by each of the enterprise categories, and derived by each of the four estimation methods. Considering the baseline estimates it can be seen that formal sector accounts for only 5.4 per cent of the total estimated income generated from household production activity. This percentage is higher for method 2 (which is 7.9 per cent) but for other methods it is otherwise similar to the percentage for the baseline estimate. This accords with prior expectations that household formal sector activity was relatively small in Ghana at this time although it is not negligible.

In summary, it would be desirable to express a view on which might be the preferred method of the four methods proposed above. If it is accepted that there is strong evidence that income has been under-recorded, then method 1 (baseline estimates) is

Table 6.5: Household formal and informal sector output (non-farm enterprise only), (in 1988/89, millions cedis)

(i) Informal sector

	Method 1	Method 2	Method 3	Method 4	Method 5
Family-fixed	107525	162486	133116	134601	113869
Family-itinerant	44284	65669	54824	55891	46897
Micro<5	51654	75888	63947	65534	54702
Micro 5-9	11079	24059	13716	14084	11733
Micro>=10	3673	4948	4548	4593	3890
Total	218216	333050	270151	274704	231091

(ii) Formal sector

	Method 1	Method 2	Method 3	Method 4	Method 5
Family-fixed	3770	6566	4668	4839	3992
Family-itinerant	1278	2316	1582	1639	1353
Micro<5	2484	4509	3075	3163	2609
Micro 5-9	3168	7978	3923	3997	3355
Micro>=10	1934	7302	2394	2487	2048
Total	12635	28671	15642	16125	13380

likely to be a lower bound estimate of informal sector activity. Method 2 is a general method of scaling at the household level and therefore scales all components equally. Method 4 is more sensitive than method 3 with regard to imputed items and would seem to have the most desirable features overall.

7. Conclusions

The aim of this paper has been to present estimates of the contribution of the informal sector (and of household production activity in general) to the gross domestic product in Ghana. This has been a complex exercise for several reasons, but two reasons predominate.

First, it has been recognised in earlier estimation and analytical exercises that a multiplicity of alternative estimates of key aggregates may be derived from the GLSS results. No one set of estimates predominate or can be considered to be the most credible in all circumstances or for all purposes. Secondly, our analysis of the individual household accounts strongly suggests some underrecording of incomes or overreporting of expenditures (or both), because of the high preponderance of implied negative savings. These results do not seem credible and it strongly suggests that incomes are under-recorded for one reason or another.

In the light of these two factors, a two-stage methodology has been developed which attempts to provide an adjustment to the sample estimates at the first stage prior to grossing-up the sample estimates at the second stage. There appears to be very little alternative other than to apply the grossing-up factor derived from the sample and universe population estimates. However, even if one accepts the likely underrecording of income, the methods proposed take no additional account of household savings, and therefore use the expenditure estimates as the lower bound of household incomes for most households in the sample.

The estimates of informal sector income and output in Ghana that emerge suggest it to be a sizeable percentage (at least 20 per cent in the case of non-farm enterprises) of the current published estimates of GDP. The more important question as to whether this production activity has already been accounted for in the existing estimates is a quite separate issue. The issue has been addressed in an earlier paper although not yet definitively resolved. Indeed there are a whole range of methodological issues that require further consideration before these estimates can be fully utilised in national accounts practice.

References

Coulombe, H., A. McKay and J.I. Round (1993) *The Estimation of Components of Household Incomes and Expenditures from the first two rounds of the Ghana Living Standards Surveys, 1987/88 and 1988/89*, Ghana Statistical Service, Accra.

Coulombe, H., A. McKay and J I Round (1994a) 'An investigation of some identified differences between rounds 1 and 2 of the Ghana Living Standards Surveys', DIERC University of Warwick, for the Ghana Statistical Service, February 1994.

Coulombe, H., A. McKay and J I Round (1994b) 'A descriptive analysis of the informal sector in Ghana based on the first two rounds of the GLSS surveys', DIERC University of Warwick, for the Ghana Statistical Service, November 1994. (Chapter 3 in this volume).

Ghana Statistical Service (1989) *Quarterly Digest of Statistics*, Vol VII No 1, Republic of Ghana, Accra.

Johnson, M., A. McKay and J.I. Round (1990) *Income and Expenditure in a System of Household Accounts: Concepts and Estimation*, SDA Working Paper No. 10, The World Bank, Washington D C.

McKay, A. and J.I. Round (1994), 'A Conceptual Background to Measuring Informal Sector Activity in Ghana', DIERC University of Warwick, for the Ghana Statistical Service, October 1994. (Chapter 1 in this volume).

Powell, M., P Debra, D. Amable and R Tonhie (1995) 'The Informal Sector in the Ghana National Accounts'. Chapter 2 in this volume.

United Nations (1991) *Surveys of Household Economic Activities A Technical Study*, National Household Survey Capability Programme, Department of Technical Cooperation for Development and the Statistical Office, New York.

Vijverberg W P M (1991) 'Measuring Income from Family Enterprises with Household Surveys' *Living Standards Measurement Study, Working Paper No 84*, The World Bank, Washington D C.