

MEASURING INEQUALITIES: DO THE SURVEYS GIVE THE REAL PICTURE? STUDY OF TWO SURVEYS IN COTE D'IVOIRE AND MADAGASCAR¹.

Charlotte Guénard
Paris 1 University - IEDES,
DIAL – IRD Research Unit
guenard@univ-paris1.fr

Sandrine Mesplé-Somps
DIAL - IRD Research Unit
mesple@dial.prd.fr

May 2005

RÉSUMÉ

Outre les choix méthodologiques effectués en amont du calcul des agrégats de consommation et de revenu et le non apurement des fichiers de base, plusieurs sources de biais conditionnent la mesure des niveaux de vie et leur distribution. Dans la présente étude appliquée à deux enquêtes malgache et ivoirienne, ces biais ont été identifiés en effectuant plusieurs scénarios de calculs des agrégats de niveau de vie, en analysant la cohérence interne des enquêtes et en confrontant les données d'enquêtes avec d'autres sources de données (comptes nationaux et balances des paiements). Les questions de méthode n'ont finalement pas de grandes incidences, à l'exception de la prise en compte ou non des prix régionaux. Les biais de sous-déclaration, bien qu'importants, ne sont pas aisés à redresser, notamment par la méthode d'imputation multiple. Cependant, on montre que les niveaux de revenu moyens dans les deux enquêtes étudiées semblent sous-évalués de 15 à 50 %. Les différentes corrections amènent les niveaux d'inégalité des deux pays vers des niveaux similaires à ceux des pays les plus inégalitaires tels que le Brésil.

ABSTRACT

Measurements of standards of living and its distribution are affected by methodological choices made before the consumption and income aggregates are calculated and by failure to correct the primary databases, but several sources of bias can also have an impact. This study, based on surveys held in Madagascar and Côte d'Ivoire, aimed to detect these biases by applying several scenarios for calculating living standards aggregates, by analysing the internal coherency of the surveys and by confronting the survey data with other sources of data, namely the National Accounts and the Balance of Payments. Methodology was found to have little impact, except for the question of whether or not regional prices were taken into account. Although there was significant bias due to under-declaration, this was not easy to correct, notably with the multiple imputation method. However, the results show that average income levels appear to be underestimated by 15 to 50% in the two surveys in question. The different corrections bring inequality levels in both countries nearer to levels in the most inegalitarian countries such as Brazil.

¹

We would like to thank Denis Cogneau and Jean-Pierre Cling for their support and helpful comments and all those who took part in the internal Dial seminar. We are entirely responsible for any errors that may remain.

1. Introduction

In recent years, there has been a large increase in primary data from household surveys. Now in the public domain, the data is used as a basis for the statistics on inequality rates compiled by international databases² and has sparked off much research on trends in world poverty and inequalities. There is a striking lack of consensus in the vast amount of literature on the subject. Apart from ideological disagreements, most of the divergence comes from the different methods used to measure living standards and from the choices of statistical sources, reviving an old debate between national accountants and household survey statisticians. This methodological controversy encouraged the researchers to re-examine the aims of the surveys carried out in developing countries in the past 20 years and the diagnoses drawn from them³.

The discussions are mainly about trends in average living standards between countries, but with little mention of the difficulties relating to measuring inequalities within countries, although this is a vital issue. As it has been demonstrated that high inequality levels reduce the growth-elasticity of poverty (Bourguignon, 2002, Cling *et al.*, 2004), a re-evaluation of inequality levels can call into question the expected effects of growth on poverty reduction. It is also essential to examine the scale of inequalities and their origins before introducing redistributive measures.

These surveys contain a large number of biases, which is far from being specific to developing countries, or to Africa in particular. There are numerous sources of potential biases: (i) collection methodology, data entry errors, and choice of how the well-being aggregates are calculated; (ii) incorrect sample design, selective observations (“non compliance”); (iii) missing values (item non-response) or underestimates of certain items in the questionnaire.

Through a scrupulous analysis of two surveys, the EPM93 survey in Madagascar and the ENV98 survey in Côte d’Ivoire, and in the spirit of Pyatt (2003), we decided to review the different biases that are likely to have an impact on measurements of average living standards and inequalities. In the second section, we assess the impact of the methodological choices underlying standard of living measurements. Although methodological and conceptual choices do have repercussions, their impact on living standards distributions is not as high as that of the data entry errors detected. Once the latter have been corrected and the method for calculating living standards aggregates has been chosen, we consider the veracity of the survey results, examining the structures of consumption and income and carrying out a decomposition of inequalities by source of income. Although it is difficult to make a diagnosis, several elements point to the fact that some of the results on living standards and inequality indicators may come from measurement errors.

Although it is now widely accepted that it is difficult to reconcile data from surveys and data from National Accounts and that both sources of information are marred by errors, we nonetheless attempt to compare them, in the third section, in a view to gaining a better assessment of the quality of household survey data.

We then try to detect and correct biases arising from sample design (section 4) and under-declarations or missing values (section 5). Sample design based on housing automatically eliminates homeless people, who are amongst the most deprived groups, meaning that the poorest people tend to be under-represented. In addition, in most household surveys, certain households selected at the sampling stage do not actually take part in the survey⁴. High-income households are likely not to take part, either because their time has a high opportunity cost or in order to protect their private lives. Enumerators are therefore obliged to replace certain wealthy households with households that are more conciliating, but which may also have more modest living standards. We detect these sample design biases by comparing certain elements of the survey with data from censuses (type of housing, nationality of

² For example Deininger and Squire (1996), WIDER (2000). For a critique of these databases, limited to OECD countries, see Atkinson and Brandolini, 2001.

³ Bhalla, 2002; Chen and Ravallion, 2004; Deaton, 1997, 2001; 2004; Ravallion, 2000, 2001.

⁴ These cases of non-response (unit “non-response”) can represent up to 30% of the initial sample in Anglo-Saxon and American surveys (see studies quoted by Mistiaen and Ravallion, 2003).

residents, etc.). The usual correction, which consists in re-stratifying the survey *a posteriori*, is not only questionable, but the results it gives in Madagascar and Côte d'Ivoire show that the real biases do not come from the sample design but from underestimates of certain types of income, which we try to rectify, or from refusals to answer certain items on the questionnaire.

Whereas cases of non-response are easy to identify, this is not always true of under-declarations. An attempt is made to assess the scale of such errors by examining the internal coherency of the surveys using an analysis of residual savings rates. It is not easy to rectify all these errors. We discuss several methods and try to detect whether or not there is a selection bias in households which refused to answer questions on their income or which were identified as having under-declared their income. Various corrections are then made on the basis of these results. In conclusion, we propose a summary of the results.

2. Impact of the quality of surveys and of methodological choices on the assessment of inequalities.

The surveys analysed in this study are the Permanent Household Survey carried out in Madagascar in 1993 (EPM93) and the Living Standards Survey carried out in Côte d'Ivoire in 1998 (ENV98).⁵ Our examination of the data aims to evaluate the impact of data availability constraints and methodological choices on the assessment of living standards.

2.1. Overall quality of the surveys

Living standards distribution can be sensitive at the extremes, which are sometimes the result of insufficient data correction. For example, before corrections were made to the data for Madagascar, the income of the wealthiest household was 15 times higher than the income generated by its one professional activity, which was an informal enterprise. The amount declared for production drawn by the household was found to be a yearly amount instead of the daily amount required in the questionnaire. Another similar example concerned a primary school teacher who declared her annual income instead of her monthly income, meaning that her household was rated among the ten richest households.

In both the surveys studied here, errors seem to be far more pronounced for incomes than for consumption (cf. Table 1). If the income data is not corrected and households with nil income due to non-responses are maintained in the data base⁶, the average income in the Ivorian and Madagascan studies increases by 11% and 78% respectively and the Gini index by 5 and 19 points!!

These errors are usually identified by examining the extremes of the distributions only. They are then corrected by the so-called "Winsorisation" method, which consists in attributing average spending and/or income levels to households with levels that are considered too extreme (i.e. levels superior or inferior to the average, plus or minus five standard deviations). In our view, this method is not very convincing as it tends to reduce the level of inequalities in a totally artificial manner. In the case of the survey in Côte d'Ivoire, this method reduces the average consumption level by 2% and the Gini index by one point (cf. Table 2, line 6).

⁵ See Appendix A for a brief presentation of the surveys analysed.

⁶ The data contains 33 households for which each of the items that make up total income was nil in Côte d'Ivoire, and only 6 households of this sort in Madagascar.

Table 1: Impact of data capture errors on consumption and income aggregates.

	Without correction	With correction
Côte d'Ivoire		
Average per capita consumption	349 [337.6 – 360.5]	349 [337.5 – 360.5]
Gini Index	43.6 [42.4 – 45.6]	43.6 [42.1 – 45.2]
Average per capita income	422.5 [373 – 471.7]	380.7 [359 – 402.2]
Gini Index	57.0 [52.6 – 61.5]	52.2 [50.3 – 55.6]
Madagascar		
Average per capita consumption	301.3 [287.6 – 315]	296.6 [284.9 – 308.4]
Gini Index	46.4 [44.1 – 48.4]	45.6 [44.1 – 47.1]
Average per capita income	640.7 [86.3 – 1 195.5]	358.5 [347 – 370]
Gini Index	69.1 [42.6 – 83.5]	40.9 [39.8 – 42.3] (b)

In thousands of F CFA or millions of current FMG

Comparison made on the basis of definitions for consumption aggregate: n° 10, Table 2 for Côte d'Ivoire; n° 6, Table 3 for Madagascar; and income aggregate n° 1, Table 15 for Côte d'Ivoire.

Confidence intervals in brackets, calculated by bootstrap.

Sources: EPM93 Madagascar, ENV98 Côte d'Ivoire, our own calculations.

2.2. Measuring living standards and inequalities: methodological choices and potential bias

Contrary to National Accounts, there is no international protocol for household surveys defining the methods for data collection and the calculation of living standards aggregates⁷. But methodological choices can have a real impact. For example, calculating an imputed rent for Madagascan households living in housing which they own increases the level of per capita consumption by 8% and reduces the Gini index by over 6 points, as poor people often own their housing. The methods of calculation must be stated explicitly, otherwise it is difficult to know whether differences in living standards observed, for example between two countries, are the result of real gaps or of different methodologies.

Apart from these issues of definition,⁸ other problems are raised. For instance, we noted that too small a number of product categories or of income sources and too long reference periods tend to lead to an underestimation of spending and income due to lapses of memory. The results of work done by Visaria (2000) on India, quoted by Deaton (2001), are striking in this respect: traditionally, the period of reference for all expenditure is a month for Indian surveys; by reducing this period to 7 days for expenditures on food, as is most often the case in other countries, poverty rates fall from 43% to 24% in rural areas and from 33% to 20% in urban areas, reducing the numbers of poor by 175 million! The way expenditure is annualised can also have an impact on the calculation of annual expenditure levels. A study of Chinese data shows that the annualisation of monthly declarations of expenditure by multiplying by twelve increases the poverty rate by 16 percentage points and the Gini index by 13 points compared with the levels calculated with statements of expenditure for the twelve months (Gibson, Huang and Rozelle, 2003).

In this respect, the Madagascan questionnaire gives far more detailed information on current consumption than the Ivorian questionnaire. For example, the Madagascan nomenclature includes 69 food items compared with 37 for Côte d'Ivoire. In addition, there are specific modules for annual expenditure on education and on consumer durables, as well as self-consumption, expressed as a quantity of goods consumed for which the household assesses the value. The survey also helps

⁷ We should point out that comparing consumption levels in the two countries is biased by the simple fact that there are great differences in the precision of the questionnaires

⁸ See Appendix A for the definition of the consumption aggregate used in this study.

measure own account production and consumption of non-food goods by individual enterprises. The respondent household chooses, for each product, the reference period for which it wishes to declare its purchases and own consumption, i.e. day, week, month or year, and is then asked how often these purchases and own consumption are made during the year.

In Côte d'Ivoire, the households declare their spending in the previous seven days and the previous month and give the number of months during the year that the product is consumed. The National Institute of Statistics (INS) calculates an average of the weekly spending (translated into monthly spending) and the declared monthly spending, and multiplies it by the declared number of months of consumption. Not taking into account the frequency with which the products are consumed during the year increases average levels of consumption (cf. Table 2 and Table 3). As this survey was carried out at a single point in time between mid-August and mid-December 1998, this method of annualisation was intended to solve the problem of the seasonal variations in expenditures. However, there is a risk of underestimating expenditure for households which declare that they have not consumed the product in the previous seven days or the previous month, but do consume it during the year. The problem of the seasonal variations in spending is not therefore completely resolved.

In addition, there can be a declaration bias regarding the number of months of consumption of each product depending on whether the household is rich or poor, or even depending on the subjective perception of its living standards. Jones and Ye (1997) show that there is a positive correlation between the number of months of consumption declared and the monthly level of expenditure. If this is the case, if the data does not take into account how frequently the products are consumed but simply multiplies the monthly figures by 12, this could give a false picture of the distribution of standards of living, in particular by giving more weight to the poorest peoples' spending than it has in reality. We checked this bias relating to the time unit of the declaration in the Madagascar survey.⁹ On average, the shorter the period chosen to declare the expenditure, the higher the declared levels. The most striking example is the consumption of rice (which represents around 18% of current spending in Madagascar): the average levels of rice consumption declared for the year are 50,000 and 35,000 FMG lower than the monthly and daily amounts declared in urban areas, *ceteris paribus*. As this gap is halved in rural areas, we suspect that the choice of the period of declaration depends on the respondents' level of wealth: hence, the impact of memory is related to an income effect.

Seasonal variations in prices and quantities of products consumed are difficult to control and can lead to biases in comparisons of time-space living standards. With respect to the seasonal nature of consumption, Jones and Ye (1997) identified the phenomenon in the case of Côte d'Ivoire. Farming households producing cash crops (coffee, cocoa and cotton) surveyed between December and March, i.e. after harvests, have significantly higher spending than the others. Among households producing staple foods and with high levels of self-consumption, spending is higher in April and May and lower from December to March. Whereas seasonality of expenditure appears to be more or less the same for coffee and cocoa producers whether their per capita spending exceeds or is lower than 100,000 F CFA, poor cotton producers are far more sensitive to seasons than rich ones. Correcting declarations for this seasonality bias actually has relatively little impact on the distribution of living standards (cf. Table 2). For Madagascar, it appears that whether or not households are surveyed during a harvest period (December to May) has no impact on declarations, except for households in the first half of the living standards distribution or those that live in the regions of Fianarantsoa and Toamasina.¹⁰

⁹ Per capita consumption levels for certain products were regressed on the variable indicating the choice of frequency of declarations, whilst controlling the seasonality of expenditures with the variable of the survey period.

¹⁰ The per capita expenditure logarithm was regressed on the indicator for the period of the household survey, regional indicators, type of crop and level of self-consumption, checking certain features of the household such as size and level of education.

Table 2: Impact of methodological choices on consumption aggregate, Côte d'Ivoire

	Average per capita (current F CFA)	Poverty rate *	Gini Index	Inter-regional inequality
1	315 [304.3 – 325.7]	17.0	45.4 [44.2 – 47.4]	22.1
2	317 [306.5 – 327.3]	16.3	45.1 [43.5 – 46.6]	21.9
3	311 [300.8 – 321.5]	17.3	45.2 [44.0 – 46.6]	21.6
4	318.8 [306 – 331.5]	17.8	46.0 [44.2 – 48.0]	22.6
5	327 736 [314.8 – 340.7]	16.8	46.1 [44.4 – 48.0]	22.8
6	308.3 [298.4 – 318.2]	17.1	44.3 [42.9 – 45.8]	21.9
7	311.2 [300.8 – 32.5]	17.4	45.0 [43.8 – 46.9]	21.7
8	301.6 [291 – 312]	21.3	46.8 [45.7 – 48.3]	22.9
9	297 [287 – 307.3]	21.7	46.4 [44.9 – 47.8]	23.0
10	349 [337.5 – 360.5]	10.1	43.6 [42.1 – 45.2]	17.3
11	356 [344.4 – 367.4]	9.4	43.4 [42.0 – 44.7]	17.1
12	319	16.2	45.2	nd

1. Current expenditure, incl. imputed rents, transfers and consumer durables; Average of weekly and monthly declarations * number of months of consumption declared for each product (average of 3 & 4)
2. Definition n°1, but monthly declarations * 12
3. Definition n°1 but monthly declarations * number of months of consumption declared for each product
4. Definition n°1 but weekly declaration (*) * number of months of consumption declared for each product
5. Definition n°1 but weekly declaration (*) * 12
6. Definition n°1 with correction for extremes by “Winsorization” method
7. Definition n°1 without consumer durables
8. Definition n°1 without imputed rents
9. Definition n°1 without imputed rents or consumer durables
10. Definition n°1 with regional price deflators. Definition chosen subsequently.
11. Definition n°1 with regional price deflators and correction for seasonality of declarations
12. Sources: Average consumption and poverty rates: <http://www.worldbank.org/research/povmonitor/>
Gini Indexes: *World Development Indicators*, World Bank, 2004.

* poverty line: 110.7 thousand F CFA equivalent to 1 US\$ PPP 85 per day.

Confidence interval between brackets, calculated by bootstrap; in thousands of F CFA or millions of current FMG.

Sources: ENV98 Côte d'Ivoire, our own calculations

A final point to be discussed is price differentials. First, infra-annual inflation can have a non-negligible impact on the level of inequalities. The EPM93 survey in Madagascar took place over a 10-month period in which inflation amounted to roughly 40%, which could have an impact on the calculations for average aggregates and on estimates for inequalities. Unfortunately, as intra-annual price indexes are not available, we have been unable to impute part of the inequalities to relative price differentials over time. In addition, price differentials within countries are often badly known and can have a significant impact on calculations of levels of poverty and inequalities and, of course, on the way they are evolving (Appleton, 2003).

In the end, out of all the possible methodological choices, given the structure of the surveys and the data available and apart from the issue of defining aggregates, the question of whether or not the regional price differentials¹¹ are taken into account has the greatest impact on the living standards calculations (Table 2, line 10 for Côte d'Ivoire: Table 3, line 6 for Madagascar). If these relative price differentials are taken into account, the average consumption level increases significantly, by more than 10% in both cases, and poverty rates are reduced by 7 percentage points. The effects on the global inequality levels amount to 2 points in the case of Côte d'Ivoire and to 0.8 in the case of Madagascar. However, these differences are not significant, although the fact that the relative price differentials are,

¹¹ In the case of Côte d'Ivoire, they are far removed from reality, given that the latest regional prices dates from 1985. On the contrary, in the case of Madagascar, the deflators used are precisely those calculated by the Madagascar Institute of Statistics when it processed the national survey (EPM) for 1993.

or are not, taken into account may explain the difference in the inequality indicators found with our own calculations and those available in the World Development Indicators.

Overall, fewer methodological choices have to be made for income than for consumption. Apart from the question of whether or not differences in relative prices are taken into account, which is posed for consumption and income alike, choices for calculating the income aggregate were above all concerned with the way of assessing farming income in the Ivorian case.¹²

Whatever the case may be, except for the question of regional prices, these methodological choices do not have a great deal of impact on the form of living standards distributions. Instead, it is data entry errors and their corrections that have a significant impact on the assessment of living standards and inequalities. Nonetheless, it is quite legitimate to question whether the structure of consumption and income and the levels of inequalities produced by these surveys are realistic.

Table 3 : Impact of methodological choices on consumption aggregate, Madagascar.

	Average per capita (current FMG)	Poverty rate *	Gini Index	Inter-regional inequality
1	264.7 [253.9 – 275.6]	59.5	46.8 [44.8 – 48.8]	14.6
2	242 913 [233.6 – 252]	62.1	53.1 [51.5 – 54.8]	14.5
3	257 661 [248 - 267]	59.8	43.7 [42.4 – 44.8]	13.9
4	235.8 [227.41 – 244]	62.8	50.4 [49.3 – 51.9]	13.0
5	465.6 [401 – 530]	26.9	46.7 [43.8 – 50.7]	13.4
6	296.6 [284.9 – 308.4]	52.8	45.6 [44.1 – 47.1]	11.5
7	237	46.4	47.45	nd

1. Current expenditure, incl. imputed rents, transfers and consumer durables;
2. Definition n°1 without imputed rents
3. Definition n°1 without consumer durables
4. Definition n°1 without imputed rents or consumer durables
5. Definition n°1 without taking into account frequency of purchase during the year
6. Definition n°1 with regional price deflators. Definition chosen subsequently.
7. Sources: Average consumption and poverty rates: <http://www.worldbank.org/research/povmonitor/>
Gini Indexes: *World Development Indicators*, World Bank, 2004.

* poverty line: 203,241 FMG equivalent to 1\$ PPP 85.

Confidence interval between brackets, calculated by bootstrap; in thousands of F CFA or millions of current FMG.

Sources: EPM93 Madagascar, our own calculations.

2.3. Is the initial picture of inequalities and the structure of living standards a realistic one?

2.3.1. Structure of consumption and income

The structure of current consumption does appear to be typical of developing countries (cf. Table 16, Appendix B): the share of food products in total expenditure exceeds 50% on average and, as can be expected, the share of self-consumption falls with the level of total expenditure. For both countries, there is a negative relation between the food budget share and the standard of living, thus confirming Engel's law.¹³ However, the share of food products exceeds 50% even for the higher deciles and it is difficult to know whether the fact that self-consumption accounts for a higher share of food consumption in Madagascar than in Côte d'Ivoire (40.4% of food, i.e. 24.5% of total expenditure) stems from a methodological difference or from a real difference between the two countries.

As for the structure of incomes, it is even more difficult to assess the veracity of the survey results without referring to other sources of information such as the National Accounts (cf. *infra*). Whatever

¹² See Appendix A for the definition of the income aggregate used in this study and a summary of the impact of these different choices.

¹³ Housing is the second item of expenditure (10% of the total in Madagascar and 14% in Côte d'Ivoire). Rent accounts for 50% of housing expenditure in Madagascar and 62% in Côte d'Ivoire; this share falls more strongly with the standard of living in Madagascar than in Côte d'Ivoire. The budget share of expenditure on consumer durables, domestic care, health and leisure logically increases with wealth, once again in a more pronounced way in Madagascar.

the case may be, the surveys gave the following results. In both countries, the share of wages increases with living standards (cf. Table 17 in Appendix B): by 8 percentage points between the first and last quartile in Madagascar and by 12 points in Côte d'Ivoire. Whereas in Madagascar, the share of farm revenues in non salary income remains stable and represents over 50% of total income whatever the level of wealth, in Côte d'Ivoire this share is halved between the first and last quartile, falling from 41% to 19%. In the latter country, the income of the first three deciles is dominated by farming activities (40% of the total), whereas this income only represents a third of the total for the middle of the distribution and less than a quarter for the last three deciles. However, an examination of the inequalities suggests that part of the results may possibly be due to measurement errors.

2.3.2. Level of inequality and decomposition of inequalities by source of income

The greatest income disparities seem to be found in Côte d'Ivoire (cf. Table 4, and Table 17, Appendix B): the Gini index is higher (52.2 compared with 40.9 in Madagascar) and the ratio of last decile over first decile (d10/d1) is 35 compared with 10 for Madagascar. On the other hand, in terms of consumption, there is greater dispersion in Madagascar than in Côte d'Ivoire. In Madagascar, the fact that inequalities are greater in terms of consumption rather than income is surprising and contrary to results usually found in other countries. We can also note the low level of the highest incomes. In Côte d'Ivoire, the maximum incomes correspond to the average formal wage of a non African, whereas in Madagascar they are at the level of the average annual salary of a Madagascan senior executive working in a formal industrial company. These two observations seem to indicate that the highest incomes are under-estimated, particularly when it comes to non-salary incomes.

Table 4: Summary of differences in living standards and distribution extremes.

	Côte d'Ivoire (1998)		Madagascar (1993)	
	On expenditure	On income	On expenditure	On income
Current € per capita per month				
Minimum	3	0	1	- 27
C1 (1 st centile)	7	2	1	- 0
P10 (upper bound, 1 st decile)	12	8	2	3
P90 (lower bound, 10 th decile)	103	93	20	25
C100 (last centile)	713	823	182	129
Maximum	2 249	3 697	1 504	429
Average per capita	44	48	13	16
In \$ PPP 1998				
Minimum	7	1	1	-32
C1 (1 st centile)	16	6	1	0
P10 (upper bound, 1 st decile)	29	19	3	4
P90 (lower bound, 10 th decile)	240	218	23	30
C100 (last centile)	1 669	1 926	218	154
Maximum	5 262	8 651	1 802	514
Average per capita	104	113	16	19
GDP per capita per month (\$ PPA 98)	123		64	
Coefficient de Gini (%)	43.6	52.2	47.5	40.9
Confidence interval	[42.7 – 45.2]	[50.20 – 54.52]	[45.9 – 50.0]	[39.8 – 42.3]
P90/P10	8	12	8	8

Not corrected; Confidence intervals between brackets, calculated by bootstrap.

Sources: EPM93 Madagascar, ENV98 Côte d'Ivoire, our own calculations.

Although they present similarities, Madagascar and Côte d'Ivoire are different in terms of structures of income and sources of inequalities (Table 5). These differences can be explained by structural differences between the two countries but also by different survey methods and by measurement errors. In order to determine the reasons for these differences, we carried out a decomposition of inequalities by source of income.

The most commonly used method for decomposing inequalities by source of income is the so-called “natural” method of Gini index decomposition, which adds three elements for each income component: its share in total income, its Gini index and a correlation coefficient that captures the relationship between each of the income components and its distribution¹⁴. Salaries represent 22% of total income in Madagascar and 28% in Côte d'Ivoire, whereas income from non-salary activities represents 66% and 60% respectively. However, in Madagascar, it is farming revenues which are most important (88% of income from work activity compared with only 42% in Côte d'Ivoire). Public and private salaries make relatively small contributions to global inequalities in the two countries, due to their small share in total income, and despite representing relatively high levels of inequality. In Côte d'Ivoire, inequalities in farming revenues are quite high (Gini index of 72 compared with 54 for Madagascar), but farming revenues are only responsible for 20% of total inequalities whereas their contribution exceeds 50% in Madagascar. On the contrary, income from non-farming activities makes a very low contribution to total inequalities in Madagascar (8%) whereas it contributes around 43% in Côte d'Ivoire. As this difference seems very large, a comparison must be made with other data sources in order to hone the diagnosis of the quality of the surveys.

Table 5: Decomposition of inequalities by source of income.

Sources of per capita income	Gini Coefficient	Share of total income	Relative contribution	Absolute contribution
Côte d'Ivoire				
Private salary	88.2	17.7	18.4	9.6
Public salary	95.3	7.6	8.7	4.6
Farming activities	71.9	27.5	19.8	10.3
Non-farming activities	80.9	35.2	43.8	22.9
Assets	74.4	5.8	3.8	2
Pension, insurance	98.1	1.5	1.8	0.9
Public transfers	94.5	0.7	0.6	0.3
Private transfers	92.4	2.8	1.78	0.9
Other income	97.8	1.1	1.2	0.6
Total	52.2	100	100	52.2
Madagascar				
Private salary	83.2	15.4	15.6	6.4
Public salary	95.5	6.8	10.6	4.3
Farming activities	54.6	58.5	52.9	21.7
Non-farming activities	90.3	7.5	8.2	3.4
Assets	57.6	5	4.2	1.7
Pension, insurance	99.2	0.08	0.08	0.03
Public transfers	98.5	1.04	1.3	0.5
Private transfers	22	1.7	1.6	0.7
Other income	93.	3.9	5.3	2.2
Total	40.9	100	100	40.9

Sources: EPM93 Madagascar, ENV98 Côte d'Ivoire, our own calculations, “natural” decomposition using DADS software.

3. Coherency with National Accounts?

Given the differences of methods and cover,¹⁵ there is clearly no real reason why the two sources of information, i.e. survey data and National Accounts aggregates, should arrive at a similar assessment of household consumption and/or income levels and in fact it is hardly surprising that this is not the case. Following work by Ravallion (2001), Deaton (2004) shows that in 277 surveys carried out

¹⁴ See, for example, Lerman and Lerman (1989) for a presentation of this method and its application to American data.

¹⁵ For a discussion on the conceptual differences in the aggregates, see Guénard and Mesplé-Somps (2004).

throughout the world, the per capita consumption found in the surveys is underestimated compared with the National Accounts, the ratio between the two data sources averaging 86% (with a standard deviation of 31%). This average ratio amounts to 78% (standard deviation 10%) for OECD countries, although they are renowned for having better statistical resources than other countries.

It is worrying that there should be such a wide gap between the two sources and also that this gap is widening all the time, whether it be in rich countries such as the United States and Great Britain, or in developing countries. In a sample of non-OECD countries from 1990 to 2000, the growth rate of consumption found in the surveys was, on average, half of that found in the National Accounts (Deaton 2004). This confirms the diagnosis that household surveys have difficulty in capturing the top end of income distributions. This is particularly true in the case of developing countries in phases of high economic growth, such as India, where the emergence of new wealthy social classes is completely overlooked by the surveys (Banerjee and Piketty, 2003).

For the two countries studied here, the two sources of information show incoherencies, mainly in the case of Madagascar, which suggest that income (but also consumption) has been underestimated in the household survey, but also problems in the National Accounts. In the Ivorian case, despite the small amount of information that can be pieced together¹⁶, the two data sources are relatively coherent (Table 18, Appendix C). It was found that the surveys underestimated consumption by only 8%¹⁷ and income by 16.8%¹⁸. These results are in line with averages for African countries, where gaps between the two data sources are relatively small (around 15% according to Deaton, 2004).

Data from the Madagascan survey greatly underestimates household consumption compared with national aggregates (cf. Table 18, Appendix C).¹⁹ In the end, it was only possible to match 53.5% of the National Accounts data with the survey data, with large disparities from one budget item to another. All the items are lower than the National Accounts figures, except for consumer durables and other goods.

Non-salary incomes are four times higher in the survey than in the National Accounts (cf. Table 19, Appendix C). Salaries, on the contrary, are lower by half. These differences in the shares of salaries and self-employment incomes in total household income may be due to differing definitions, as the sum of the two is almost identical. Income drawn by entrepreneurs in quasi-corporate undertakings represents over 50% of Madagascan households' available income in the National Accounts, which seems very high in absolute terms and also compared with the survey, which puts it at about 10%. This gap alone accounts for the entire difference between the figures for total income in the National Accounts and in the survey.

Data from the National Accounts is generally used to correct data from household consumption surveys. In this case, the consumption in the surveys is adjusted using an average coefficient taken from the National Accounts, following the example of Bourguignon and Morrisson (2002) and Sala-i-Martin (2002). This method implies, on the one hand, that the National Accounts are considered to be more reliable than the surveys and, on the other, that the gap between the two sources is neutral in distributive terms, i.e. that the underestimation of consumption in the surveys is a constant proportion, at all levels of wealth. As we find these assumptions debatable, this method has not been used in this study.

However, we did think it worthwhile to try to correct the survey data on income, even if the income drawn by formal entrepreneurs in Madagascar is quite probably overestimated in the National Accounts. In order to do this, the first step was to identify these formal entrepreneurs in the survey.

¹⁶ Certain elements of information are missing in the surveys, such as tax paid.

¹⁷ The most frequently underestimated items are consumer durables, equipment and maintenance and other expenditure.

¹⁸ Survey data was therefore used to compile the household account in the National Accounts.

¹⁹ Problems were encountered for matching the two types of information, particularly for aggregating household expenditure on consumption and for harmonising branch and product nomenclatures for surveys and [Input-Output Tables](#). For details on matching the two data sources, see Guénard and Mesplé-Somps (2004).

This was not an easy task, because the section of the questionnaire dealing with employment did not contain the modality “corporate or quasi-corporate entrepreneur”. Twenty-two individuals declared themselves as “employers”, including 14 who owned a company with a “statistical identity card”. In the absence of more precise information to be able to check whether these are corporations or quasi-corporations, these individuals were put in the same category as formal entrepreneurs, 10 of which are in the last two consumption deciles. In addition, 4% of the households surveyed declared that they had received “dividends and other business revenues”, 50% of which were distributed in the three most wealthy deciles.

Four simulations are proposed (cf. Table 6). The first two help equalize the income of formal entrepreneurs in the survey with that of the National Accounts. 90% of the income given in the National Accounts is either allocated exclusively to formal entrepreneurs, or to the latter in addition to households receiving dividends. In this case, the average per capita income increases by around 50% and inequalities increase by over 23 points. The two remaining simulations allocate only 40% of the income given in the National Accounts, in the same way as for the previous two simulations. This re-estimate was made equiproportionally to the income from non-salary, non-farming activities and income received in the form of dividends. In this case, inequalities increase by over 15 points, and the Gini coefficient increases by 12 points when the transfer of dividends concerns formal entrepreneurs and households receiving dividends. Income inequalities are then higher than consumption inequalities.

Table 6: Correction of Madagascar survey data using National Accounts.

Madagascar	Without correction	Correction (1)		Correction (2)	
		(a)	(b)	(a)	(b)
Average per capita income	358.5 [347 – 370]	628.2 [371 – 885]	643.3 [603.3 – 683.3]	478.4 [363.4 – 593.4]	485 [464 - 506]
Gini Index (%)	40.9 [39.8 – 42.3]	66.2 [53.2 – 77.5]	64.0 [62.7 – 65.5]	55.7 [47.1 – 65.8]	53.3 [52.1 – 54.9]

Correction (1): Addition of 90% of income drawn by entrepreneurs in quasi-corporate undertakings given in the National Accounts

Correction (2): Addition of 40% of income drawn by entrepreneurs in quasi-corporate undertakings given in the National Accounts

(a) To « formal » entrepreneurs

(b) To « formal » entrepreneurs and households declaring receipt of dividends

Confidence intervals between brackets, calculated by bootstrap; in thousands of F CFA or millions of current FMG

Sources: EPM93 Madagascar; our own calculations.

In the case of India, Banerjee and Piketty (2003) use a different method to try to make up for the difference between the National Accounts and the household surveys. They re-estimate the high incomes from surveys between 1956 and 1998 using tax data and a method already developed for France and the United States (Piketty, 2003; Piketty and Saez, 2003). Their corrections highlight a 50% growth in income for the last centile and a tripling of average income for the last centile. The gap between the national data and the survey data is only partly reduced and explained by the fact that the latter are not well suited to capturing the income of wealthy individuals. The work using tax statements is extremely interesting, but for want of such information, we were unable to apply it to our case.

4. Assessment and means of correcting sample biases

4.1. Survey cover and potential biases

In order to measure the distributive impact of any sample biases of the surveys, we compared elements from the sample design with population censuses conducted in 1998 in Côte d'Ivoire (RGP98) and 1993 in Madagascar (RGP93) calculating several indicators: rate of urbanisation, structure by nationality, by ethnic group, by socio-professional group, by level of education and by type of housing, depending on the information available in each of the countries. In Côte d'Ivoire, the sample design for the 1998 survey was made on the basis of the 1988 General Population Census, which is problematical. In Madagascar, however, the survey is based on a census in the same year, 1993 (INSTAT, 1997).

A first potential bias concerning the possible under-representation of what are in principle the poorest households, i.e. households with no fixed abode, could not be evaluated. The census in Madagascar specifies that this specific population was interviewed separately and independently from the “domiciled” population, but we do not have the results in question. In Côte d’Ivoire, the definition of a household used in the successive censuses (1975, 1988, 1998) takes into account the place of residence.²⁰ In practice this means that homeless households were neither covered by the census nor by the surveys. On the contrary, other significant sample biases were identified:

- The rate of urbanisation is significantly overestimated in the survey (45%) compared with the census (42%); more precisely, Abidjan accounts for 21% in the survey and only 19% in the census.²¹ This confirms the fact that urban growth is overestimated in many countries, particularly in Africa (Bocquier, 2004) while it seems to be decreasing in Côte d’Ivoire (Beauchemin, 2004);
- The population of foreigners of African origin is under-represented; it accounts for 17% of the total population in the survey compared with 26% in the census.²² This bias points to an underestimation of inequalities given that these foreigners are more likely to belong to the bottom of the distribution, as demonstrated by Grimm et al. (2002);
- Households living in villas or individual houses are significantly underestimated in the survey (22% compared with 39% in the census) and this is also the case for households in makeshift housing such as shacks or huts (6% compared with 11%), particularly in urban areas. This bias points to a potential underestimate of inequalities due to a poor assessment of the distribution extremes.

In the Madagascan case, the corrections concerned geographical strata (urban or rural area) by region and by level of education (cf. correction 1 and 2 of Table 7). The rate of urbanisation for the survey is 18.5%²³ compared with 22.9% in the census. In terms of education, 48.2% of individuals in the survey have received no education compared with 33.7% in the census and 39% have been educated to primary level compared with 47% in the census.²⁴ No clearly identified bias could be extracted from the other indicators available, particularly on housing characteristics (type of lighting, fuel used for cooking, type of floor, etc.). In addition, the breakdown of the population covered by the census by type of housing is not available.

In both countries, another bias is likely to have a strong impact on living standards and their distribution, namely foreigners of non-African origin residing in the country. This small population was deliberately not accounted for in the surveys. In Côte d’Ivoire, this concerned 32,700 individuals (source RGP98), mainly expatriate Europeans (16,028) and persons of Lebanese origin, i.e. 0.2% of the country’s total population, a very marginal percentage in demographical terms with no possible comparison to its economic weight. In Madagascar, the foreign population was not covered by the survey either. It represents 0.2% of total population : over a third are Europeans, 29% Asians, 10% Africans, 17% from the Indian Ocean and the remainder from the United States, the former USSR, Oceania and stateless persons (INSTAT, 1997).

4.2. Impact of correction of biases on inequality levels

The corrections consisted in re-stratifying the survey *a posteriori* by an iterative correction process of cross-tabulation for two criteria (nationality/geographical stratum, for instance). The households’ relative weights could then be corrected by a coefficient restoring their share in the total population before proceeding with the aggregation of incomes.

²⁰ “According to the General Population and Housing Census of 1998, the ordinary household is composed of a group of related or unrelated people who recognize the authority of a single individual called the “head of household”, living under the same roof or in the same concession and whose resources are pooled or partly so.” (Touré, 2001).

²¹ The differences between the survey and the census are significant as confidence intervals for the percentages of urban population and Abidjan population are [44.2 % – 48 %] and [20.3 % – 22.7 %] respectively.

²² This difference is also highly significant since the confidence interval for the percentage of foreigners of African origin in the population surveyed is [16.9 % – 17.9 %];

²³ The confidence interval is [17.98% – 18.91%]

²⁴ Once again, there are significant differences in the totals and by geographical stratum.

Three weighting corrections were made for Côte d'Ivoire: the first for ethnic origin and nationality by geographical stratum (correction 1), the second by type of housing by geographical stratum (correction 2) and the third by nationality by type of housing (correction 3), to correct levels of consumption and income (cf. Table 7). None of these corrections has a significant impact on income or consumption inequalities (cf. Table 7).

For Madagascar, the impact on the Gini coefficient is also fairly low, amounting to roughly one Gini point, and therefore non significant. These small impacts on inequalities no doubt stem from the method itself, as it amounts to replacing the missing values in the sample with average values for the corresponding sub-populations. There is therefore the implicit assumption that the non-respondents in a category (or the non-interviewed, which comes to the same thing) cannot be distinguished on average from the respondents. It also artificially introduces a concentration around the average values, meaning that the variances are calculated on the adjusted sample that underestimates the real differences.

These results suggest that the real biases are not to be found in the sample designs, even though these are not perfect as in the case of Côte d'Ivoire, but from an underestimate of certain types of income, as we mentioned above. We therefore adopted a different method for adding the foreign population.

For Côte d'Ivoire, we could have corrected this bias by correcting the weight of households living in villas or flats in urban areas, as these types of housing are partly occupied by Europeans and Lebanese. This would amount to assuming that this population has the same living standards as the African households living in this type of housing and covered by the survey. However, non-Africans tend to receive income in line with levels in their countries of origin. We therefore drew up two hypotheses. The first, so-called "high" hypothesis,²⁵ assumes that this population, comprising households with four members on average, has a monthly income of 4,500 euros per household and consumes on average 2,290 euros per month per household (in this case, their consumption level is equivalent to that of the ten most wealthy households). A second, so-called "low hypothesis", consists in allocating them with the average income of French households (approximately 2,100 euros) and average consumption of around 1,100 euros.²⁶ Whereas the average consumption and income levels rise by 15% and 30% with the high hypothesis, and 7% and 11% with the low hypothesis, Gini indexes on expenditure rise by 6 points and on income by 9 points with the high hypothesis and by 2 points and 4 points with the low hypothesis (cf. correction 4, Table 7).

In Madagascar, given the make-up of the foreign population, we assumed that one third of them have living standards equivalent to Western expatriates, and that the remaining two-thirds have living standards equivalent to the country's average. As in the case of Côte d'Ivoire, the high hypothesis for the income of the population from Northern countries has a very strong impact (cf. correction 4, Table 7): average consumption and income levels rise by 12 and 15% respectively, whereas the Gini coefficient on consumption increases by 6 points and that of income by 8 points. With the low hypothesis, all these effects are halved. Hence, even though the results of this simulation exercise on inequalities are sensitive to the chosen hypotheses, they show the extent to which inequalities can be underestimated if this category of households is not taken into account.

²⁵ Given that the average non-African salary in the private sector was approximately 2,600 euros in 1996 (Cogneau and Mesplé-Somps, 2002). However, expatriates' salaries in the public sector are far higher than this average private salary (approximately 6,000 euros) and it can also be assumed that their spouses also receive an income.

²⁶ This covers households which declare a positive or nil income to the tax authorities and for which the reference person is neither a student nor doing national service (source INSEE).

Table 7: Adjustment of consumption and income levels by correcting sample design

	Without adjustment	Correction 1	Correction 2	Correction 3	Correction 4	
					(1)	(2)
Côte d'Ivoire						
Ave. per capita consumption	349 [337 – 360]	347 [336 – 359]	347 [332 – 363]	336 [322 – 350]	403 [372 – 434]	373 [356 – 390]
Gini Index (%)	43.6 [42.1 – 45.2]	43.6 [42.3 – 45.5]	44.2 [42.3 – 46.7]	43.7 [41.7 – 45.3]	50.0 [47.4 – 53.3]	46.2 [44.3 – 48.0]
Ave. per capita income	381 [359 – 402]	385 [362 – 408]	387 [361 – 414]	382 [356 – 408]	491 [427 – 554]	426 [393 – 458]
Gini Index (%)	52.2 [50.3 – 55.6]	52.5 [50.3 – 54.3]	52.9 [50.7 – 56.0]	52.8 [50.1 – 55.2]	61.8 [56.6 – 66.1]	56.2 [54.0 – 58.8]
Madagascar						
Ave. per capita consumption	297 [285 – 308]	307 [295 – 320]	311 [298 – 324]		332 [295 – 369]	314 [293 – 335]
Gini Index (%)	45.6 [44.1 – 47.1]	46.4 [44.9 – 48.5]	46.1 [44.6 – 47.9]		51.3 [47.4 – 56.5]	48.5 [46.4 – 51.7]
Ave. per capita income	359 [347 – 370]	362 [350 – 373]	366 [353 – 378]		412 [358 – 465]	385 [357 – 414]
Gini Index (%)	40.9 [39.8 – 42.3]	41.3 [40.1 – 43.0]	41.1 [39.6 – 42.5]		48.5 [40.6 – 53.4]	44.9 [41.2 – 49.4]

In thousands of F CFA or millions of current FMG; confidence intervals for Gini coefficients between brackets, calculated by bootstrap.

Correction 1: on ethnic origin and nationality by geographical strata in Côte d'Ivoire and on area of residence and geographical strata in Madagascar.

Correction 2: on the type of housing by geographical strata in Côte d'Ivoire and on levels of education by geographical strata in Madagascar.

Correction 3: by nationality and type of housing in Côte d'Ivoire.

Correction 4: addition of foreign population of non-African origin in Côte d'Ivoire and the foreign population in Madagascar, with high hypothesis (1) and low hypothesis (2) of living standards for populations of non-African origin.

Sources: EPM93 Madagascar, ENV98 Côte d'Ivoire, our own calculations.

In the following section, we continue our discussion on measurement errors by examining under-declarations of income. Apart from comparisons with the National Accounts and methods for rectifying sample design, with the advantages and limits given above, several methods have been used by the community of researchers and statisticians to rectify biases from under-declarations and non-responses on income. After showing the incoherency of consumption and income declarations using residual savings rates calculations, we will apply different methods of adjustment to the two surveys studied here.

5. Under-declarations and non-responses

5.1. Incoherencies between income and consumption declarations.

It is a well known fact that surveys generally provide negative savings rates and there are discussions as to the origin of this phenomenon. As Deaton noted (1997, p369): “*household surveys from developing countries frequently do record dissaving by substantial fractions of households. There is no doubt that some of this is due to underestimation of income relative to consumption, but the observation may have more truth to it than is often credited*”. However, the fact that large numbers of households had levels of income representing half of the declared levels of consumption attracted our attention (cf. Table 8) and seems to suggest that there are large problems with under-declarations of income, throughout the distribution in Côte d'Ivoire and at the top of the distribution only in Madagascar.

In Côte d'Ivoire, the average rate of residual savings is -86%, and is always negative, whatever the consumption decile, and decreases with the level of consumption. Thus, 61% of households in the Ivorian sample have negative residual savings, which can be normal behaviour in the event of loss of employment, no harvest or other negative shocks. However, over 20% of households had a residual savings rate lower or equal to -100%. These cases are found throughout the distribution, but more frequently in the highest deciles. We tested to see whether there was a link between the economic difficulties experienced by the households and the behaviour adopted to face up to them (using their savings, selling off assets, getting into debt), but found that households with the highest rates of dissaving did not have to face any greater economic difficulties than the other households.

In Madagascar, the average rate of residual savings is only -4% and is distributed in a completely different way: the average savings rates are positive and relatively high for the entire distribution with

the exception of deciles 3, 9 and 10. Only 5% of the households in the total sample consumed twice more than their current income allowed them to. It is also surprising to note such high savings rates at the bottom of the distribution, which seems to suggest that consumption is underestimated.

Table 8: Average residual savings by consumption decile

Consumption decile	Côte d'Ivoire			Madagascar		
	Average savings rate (%)	Stand. Deviation	% of households with savings rates $\leq -100\%$	Average savings rate (en %)	Stand. Deviation	% of households with savings rates $\leq -100\%$
1	-7.7	84.7	7	32.5	30.9	0
2	-50.6	533.5	11	25.6	91.6	2
3	-43.7	244.8	11	-32.5	1199	2
4	-39.5	147.6	17	27.4	46.3	2
5	-60.5	250.3	15	20.5	126.2	2
6	-62.8	214.1	21	11.1	105.4	4
7	-55.9	153.5	22	14.3	467.5	4
8	-77.3	196.8	26	4.9	98.6	4
9	-125.6	366.6	37	-19.9	194.4	9
10	-337.2	1910.2	52	-125.9	1948.6	18
Total	-85.9	662.8	22	-4.2	745.2	5

Sources: EPM93 Madagascar, ENV98 Côte d'Ivoire, our own calculations.

In order to make the income declarations coherent with the consumption declarations, even though these, too, are subject to errors, we then attempted to rectify the income declarations using the different methods currently available.

5.2. Correction of biases for under-declarations of income and for non-responses, and comparison with other data sources

We propose several adjustments: a correction based on the information concerning declared savings (Madagascar); a correction for missing values by Rubin's multiple imputation method (Rubin 2004); and finally, using information given in the Balances of Payments, an assessment of capital flight attributed to certain households in the survey which are likely to hide part of their income.

5.2.1 Adjustment with declared savings

A first method that Loisy (1999) implemented in France helps adjust the income by using information on the amounts of accumulated savings declared by the households. Only the EPM93 survey in Madagascar asked the households about the flow of savings accumulated during the current year, and on credits contracted. In line with Loisy (1999), income can be adjusted as follows: in the cases where the sum of declared consumption and savings ($C + S_{\text{declared}}$) is higher than the declared income (Y_{declared}), and the household has not obtained a consumption loan, the income can be replaced by ($C + S_{\text{declared}}$); in the other cases, i.e. where the income is higher than ($C + S_{\text{declared}}$) and when the households have a lower income but have obtained a loan for the purchase of consumer goods, no replacements are made. This adjustment concerned 23% of the Madagascan households.²⁷

²⁷ 45% in the French « Family Budget 1995 ». Contrary to Loisy, we do not correct the consumption declarations beforehand for the differences observed compared with the National Accounts, for the reasons stated above..

Table 9: Adjustment of Madagascan income with declared savings

	Without adjustment	With adjustment by declared savings
Per capita income	358.5 [347– 370]	407.8 [393 – 422.6]
Gini index (%)	40.9 [39.8 – 42.3]	42.9 [41.34 – 44.36]
Average residual savings rates by consumption decile (%)		
1	32.5	35.9
2	25.6	35.1
3	-32.5	33.2
4	27.4	33.5
5	20.5	30.5
6	11.1	26.6
7	14.3	22.8
8	4.9	21.9
9	-19.9	17.8
10	-125.9	7.8
Total	-4.2	26.5

Confidence intervals between brackets, calculated by bootstrap; in thousands of F CFA or millions of current FMG.
Sources: EPM93 Madagascar, ENV98 Côte d'Ivoire, our own calculations.

This correction increases the average income by 14%, the average residual savings rate by 30 percentage points and the Gini index by 2 points (gap not significant). The total savings rate for Madagascan households is then 27%, a relatively high rate compared with the National Accounts, which give a rate of only 2.3%. This method does not appear to be entirely satisfactory given that it does not obtain a Gini coefficient on income higher than that on consumption.

5.2.2 Accounting for missing values

Basic principles:

A high percentage of households can refuse to answer the part of questionnaires that concerns their income. For example, in the case of the American *Current Population Survey*, the rate of non-response exceeded 25% at the beginning of the 1980s (Lillard *et al.*, 1986). The community of survey statisticians proposes several ways of accounting for missing values. Although this problem concerns less than 1% of the sample in the two cases studied here (cf. note 6),²⁸ it is nonetheless interesting, in our view, to examine the different adjustment methods proposed, to see how they could be used to correct under-evaluations of income, a problem that we identified as important in our previous diagnoses. The following paragraphs outline the basic principles of some of the methods for adjusting missing values.

One relatively simple correction is to allocate the average or median²⁹ of observations belonging to the “same category” as the missing observation, in what is called the *mean or median imputation method*. This method is used by the American Bureau of Statistics to make up for missing values on income: biases for non-responses in income declarations are corrected by allocating to individuals who refused or omitted to declare their income the average response for households with the same characteristics in terms of age, race, gender, type of occupation, level of education and number of hours worked (Census Bureau, 2002). This procedure is not perfect, as the distribution of the new variable is incorrect because average values have been added. This means that variance is underestimated, thereby artificially increasing the significance of the regression estimators. In addition, this matching method does not allow for a large number of variables to be taken into account.

²⁸ General practice in this case is to eliminate observations for which certain variables are missing (*case deletion*).

²⁹ The median is preferred to the mean when the latter is too dependent on the values at the extremes of the distribution..

A second correction consists in estimating an explanatory model for the variable for which part of the observations are missing, using predicted coefficients to estimate the missing values, in what is called the *imputation using a prediction model*. Székely and Hilgert (1999) adopted this method in several Latin American countries, estimating income for each income component and classifying the households according to predicted income. Households which did not provide information on their income were allocated a predicted income, plus a residual term equal to the average of the error terms for the two households just above and just below it in the new distribution.

This method, like the preceding one, does not take into account the selection mechanism which prevails in the total or partial non-response. And yet, for data on income, several elements suggest that not only the fact of being surveyed may not be random (problem of sample bias mentioned above), but also, the fact of answering the enumerator's questions on income correctly or not may not be random either. As we already suspected, missing or underestimated income in the surveys studied are related to income from assets or non-salary activities. In addition, it can be assumed that, from discretion or fear that the information given to the enumerator will be passed on to the tax authorities, certain households refuse to declare their income over a certain amount.

There are several possible mechanisms governing the process of non-response or partial response (Rubin, 2004). A selection mechanism is *Missing Completely at Random* (MCAR) when the probability of a given variable X being missing does not depend on any observed variable. This hypothesis is often quite unrealistic. The *Missing at Random* (MAR) selection is a slightly less restrictive hypothesis. The probability of answering the questionnaire depends on part of or all the values observed (data collected in the survey or data coming from the survey design) but is independent from the behaviour model – in the present case, the model of income formation. In cases where the probability of not being included in the data depends on these same missing observations and that the two models, of selection and of behaviour, are not independent, the process of missing values is said to be “non-ignorable”. In cases where the respondent is informed of the object of the survey, which is the case of household surveys in developing countries, it is possible that the selection process is endogenous and that there is no variable that allows selection to be ignored (Gautier, 2005). A way of checking whether or not the selection process is random is to implement the two-stage selection process proposed by Heckmann (1979) and to test its validity.

Once the selection method identified, the most appropriate imputation method appears to be that proposed by Rubin (2004). It imputes the values of missing observations on the basis of a parametric behaviour model³⁰ and takes into account the uncertainty of the prediction of the values to be estimated by imputing several values to replace the missing values and not single values as in the previous methods. Using a behaviour model (income equation), the missing values are replaced by several values reflecting the uncertainty of the real replacement value. This so-called “multiple imputation” model, based on the Bayesian inference, then produces a series of possible corrections. However, this method can only be used if the selection process is random (MAR). The following box presents the basic principles of this method and the conditions required for it to be implemented.

³⁰ Instead of estimating an income equation, it seems preferable to impute income values using a model explaining the fact of income being missing or underestimated. A propensity score is then estimated (*weighted hotdeck imputation*) with a logit model on the dichotomic variable R taking the value 1 when the income declaration is considered incorrect and 0 in other cases: $\log \text{it}(p) = \beta X$ with $p = \Pr(R = 0 / X)$ and $\log \text{it}(p) = \log(p/(1-p))$. The observations are then grouped depending on their propensity score. This method is more powerful than the mean or median imputation method since it takes into account a greater number of variables in the matching process. However, when the observations are not missing in a totally random way, the propensity score method produces biased results, contrary to the regression-based method which takes into account correlations between variables (Allison, 1999).

Box 1: Basic principles of the Rubin multiple imputation method (2004)

Let Y be the variable in which part, Y_{mis} , of the observations is missing and the other part, Y_{obs} , is observed. In the case where the probability of not being included in the data does not depend on the missing observations Y_{mis} , but may depend on Y_{obs} and the vector X , the process of missing values is random:

$$p(R / X, Y_{obs}, Y_{mis}) = p(R / X, Y_{obs})$$

with a binary variable R taking the value 1 when the variable Y is observed and 0 when the individual interviewed did not give a value for Y .

First, an explanatory model for the variable with missing values is estimated. An *a posteriori* distribution is deduced from the parameters of the model in order to use this distribution to make m random draws of the parameters and the error term in order to create m imputations for each of the missing values using the prediction model.

In the case where the regression model is linear and normal, the procedure is as follows:

a) An explanatory model of income Y is estimated on the set Y_{obs} :

$$Y_{obs} = \beta X \text{ with } u \sim N(0, \sigma^2)$$

The estimated parameters of the model are $\hat{\beta}$ and the associated variance-covariance matrix is $\hat{\sigma}^2 V$ with $V = [X X]^{\text{T}}^{-1}$. The Bayesian inference show that *a posteriori* the distribution of σ is equal to $\hat{\sigma}^2 (n - q) / g$ with n the number of respondents, q the size of the vector X and g a random variable χ_{n-q}^2 .

b) Drawing this random variable χ_{n-q}^2 serves to draw the variance-covariance matrix: $\sigma_*^2 = \sigma^2 (n - q) / g$.

c) The parameters are drawn from the following distribution: $\beta_* = \hat{\beta} + \sigma_* [V]^{\text{T}/2} Z$ where Z is a vector of independent random normal variates.

d) The missing values are then replaced by: $Y_{i*} = X_i \beta_* + z_i \sigma_*$, where z_i is a simulated Normal variable.

e) Operations b, c and d are repeated m times to obtain m missing values for each of the missing observations. The results of the different imputations are combined and analysed, particularly to ensure that there is a sufficient number of imputations, which is the case when the inter-imputation variance is lower than the intra variance ($B < \hat{U}$). Given \hat{Y}_i and \hat{U}_i the mean value of the variable Y and its variance from an imputation i ($i = 1, \dots, m$).

The mean of m imputations \hat{Y}_i is defined as: $\bar{Y} = 1 / m \sum_{i=1}^m \hat{Y}_i$.

$\hat{U} = 1 / m \sum_{i=1}^m \hat{U}_i$ is the intra variance whereas B is the inter variance: $B = \frac{1}{m-1} \sum_{i=1}^m (\hat{Y}_i - \bar{Y})^2$.

Implementation:

In the cases in question, the first step was to identify the households which appeared to have declared their incomes incorrectly. We selected two criteria. The first was used in the Ivorian case, where the income had to be replaced if it was nil because no income of any sort whatsoever was declared, and if the household's savings rate was lower than or equal to -100 %. This first target group represented 23% of the total sample. The second criterion was applied to the Madagascan case, where the income was recalculated when the sum of consumption and declared savings ($C + S_{\text{déclarée}}$) was higher than the declared income ($Y_{\text{déclarée}}$), and that the household had not contracted a consumption loan.

The second step consisted in checking whether or not there was selection bias. A Heckmann procedure has been implemented. In our case where the bias of under-declarations effectively appears to grow with living standards, the exogenous variables chosen in the selection equation are the professional status of the active members of the household (inactive, salaried, independent, farmer or home help), the logarithm of per capita consumption, together with indicators on whether the individuals carry out

several activities concurrently in the household, declare that they do not work, receive unearned income and receive profits in addition to a salary. There is clearly a bias of this sort (cf. column (a) Table 10). As expected, there is a positive correlation between the logarithm of per capita expenditure and the fact of not declaring income correctly, or the presence of independent workers in the household, in Madagascar. However, the other explanatory variables are either non significant or have a negative sign. These unexpected negative effects could show that individuals are coherent in the sense that they cheat about the level of income declared as well as about multi-activity, or the origin of their income by hiding the fact that they receive unearned income, for example.

The data does not seem to satisfy the conditions for correctly implementing the procedure proposed by Rubin as the selection is not ignorable.³¹ Nonetheless, we made two estimates. The first does not include variables to control selection bias, whereas the second tries to take them into account. The vector of the variables serving to calculate the income function is thus made up of two groups of variables: the first comprises the sex and age of the head of household, the average level of education in the household and its level of experience, its demographical breakdown, the number of hectares owned and the region of residence; the second includes the variables concerning the fact of replying correctly or not on the amount of income received.

Both in Madagascar and in Côte d'Ivoire, the results before the second group of variables were introduced are relatively standard (Table 10, columns b): positive, significant impact on level of income from the fact that the head of household is a man, the household's level of education and experience and the number of adults; negative impact from living in a rural area and positive from living in Abidjan (but non significant from living in Antananarivo). In Madagascar, households cultivating farm land have lower incomes than other households, whereas in Côte d'Ivoire this only seems to be the case for households with less than two hectares. The age of the head of household has no incidence in either country.

Columns (c) of Table 10 show the results when the two groups of variables are used to estimate the income equation. As expected, there is a correlation between the logarithm of per capita expenditure and income. Receiving unearned income or profits increases the level of income for Madagascan and Ivorian households. In Côte d'Ivoire, the same is true for multi-activity and declaring to be an independent worker, but also for the presence of inactive members in the household. In Madagascar, the presence of salaried workers in the household, but also of inactives, farmers and home helps has a positive impact on the level of income declared.

Five imputations were made and proved to be sufficient.³² Estimates of missing values using a "standard" income equation do not change either the average values of per capita income or the Gini coefficients. When variables able to explain the declaration bias are taken into account there is an increase in average income of approximately 6% in Côte d'Ivoire and 13% in Madagascar (cf. Table 11). Whereas the Gini coefficient does not change in Côte d'Ivoire, it rises by 2 points in Madagascar. These results are disappointing and show that this procedure is not sufficient for correcting under-declarations of income. The incomes of non-declarants and under-declarants are apparently systematically different from the income declared by the households with the same observed values for the X variables, but we do not have the means of correcting this bias properly. In the end, although there appear to be important problems with under-declarations of income, it is not easy to correct them.

³¹ In this case, Rubin (Chapter 6, 2004) discusses several methods that can be used to take into account this bias, but computerized procedures with standard statistical analysis software have not yet been developed.

³² Depending on the method, the inter-imputation variance being lower than the intra variance (cf. Box 1).

Table 10: Income equation.

Dependent variable: log of total income of household	Côte d'Ivoire			Madagascar		
	(a) (n = 4164, n censored = 3249)	(b) (n = 3249)	(c) (n = 3249)	(a) (n = 4494, n censored = 3480)	(b) (n = 3480)	(c) (n = 3480)
Constant	13.13*	12.04*	2.38*	13.89*	13.50*	3.07
Sex of head of household (1 = masc.)	0.13*	0.28*	0.17*	0.20*	0.29*	0.13*
Age of head of household	-0.01	0.0002	0.004*	0.002	- 0.002	- 0.0004
Ave. level of education of household (Average level of education) ²	1.11*	0.17	- 1.30*	1.15*	0.88*	- 0.07
Average level of experience. (Average level of experience) ²	-0.96	1.57*	1.24*	0.17	0.92*	0.59*
	0.09	3.72*	1.57*	0.004	0.003*	0.002
	0.09	- 3.06*	- 2.36*	-0.00002	- 0.00004	-
						0.00005
Number of children 0 to 5 years (Number of children 0 to 5 years) ²	0.26*	- 0.04	0.13*	0.15	0.08*	0.21*
	-0.05*	0.01*	0.01*	-0.02	- 0.02*	- 0.02*
Number of children 6 to 14 yrs (Number of children 6 to 14 yrs) ²	0.26*	0.04*	0.19*	0.15*	0.08*	0.25*
	-0.03*	0.0009	-0.01*	-0.01	- 0.01*	- 0.03*
Number of adults (Number of adults) ²	0.26*	0.16*	0.26*	0.32*	0.19*	0.34*
	-0.01*	- 0.005*	- 0.01*	-0.02*	- 0.01*	- 0.02*
Ownership of farm land						
Without land	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
[1 to 2 ha]	0.25*	- 0.13*	- 0.08	-0.59*	- 0.57*	- 0.28*
]2 to 5 ha]	0.41*	0.08	0.09*	-0.25	- 0.38*	- 0.17*
over 5 ha	0.22*	0.30*	0.25*	-0.30	- 0.20*	- 0.11*
Living in rural area	-0.18*	- 0.27*	-0.12*	0.10	- 0.12*	0.11*
Abidjan, Antananarivo	0.25*	0.37*	0.08*	0.06	- 0.008	0.009
Log of per capita expenditure	0.40*		0.81*	0.50*		0.77*
Presence of salaried workers	-0.41*		- 0.06	-0.11		0.08*
Presence of independent workers	-0.24*		0.23*	0.31*		- 0.03
Presence of farmers	-0.59*		- 0.01	-0.71*		0.16*
Presence of home helps / apprentices	-0.06*		- 0.03*	-0.01		0.06*
Presence of inactives	-0.07		0.19*	0.12		0.08*
Multi-activities	-0.49*		0.16*	-0.04		- 0.04*
Receives profits	-0.30*		0.22*			
Receives unearned income	-0.23		0.08	-0.10		0.16*
constant	-5.27			-6.85*		
R ²	354.6	0.35	0.63	601.6	0.38	0.70
	(Wald chi2)			(Wald chi2)		
Mills Ratio	-1.08*					

* : variables significant to 5%

(a) income equation with control of selection process (the fact of answering incorrectly about amount of income) by Heckmann procedure;

(b) income equation without taking into account selection process;

(c) income equation without taking into account selection process but introducing variables in the equation to explain the fact of not answering correctly about amount of income.

Sources: EPM93 Madagascar, ENV98 Côte d'Ivoire, our own calculations.

Table 11: Adjustment of income by multiple imputation method

	Without adjustment	Multiple imputation regression (a)	Multiple imputation regression (b)
Côte d'Ivoire			
Per capita income	380.7 [359 – 402]	384.5 [380.8 – 387.7] (c)	402.7 [399 – 407.8] (c)
Gini index (%)	52.2 [50.3 – 55.6] (b)	51.3 [51.1 – 51.5] (c)	52.1 [51.7 – 52.6] (c)
Madagascar			
Per capita income	358.5 [347 – 370]	373 [371 – 376.5] (c)	400.7 [397.7 – 404.6] (c)
Gini index (%)	40.9 [39.8 – 42.3] (b)	41.0 [40.7 – 41.5] (c)	42.8 [42.6 – 43.1] (c)

Confidence intervals in brackets, calculated by bootstrap ; in thousands of F CFA or millions of current FMG.

(c) minimal and maximal values of imputations

Sources: EPM93 Madagascar, ENV98 Côte d'Ivoire, our own calculations.

A final attempt to rectify the survey biases was made by examining capital flight using information given in the Balances of Payments, adding it to the income of certain categories of households.

5.2.3. Assessing capital flight and its impact on income inequalities

Using the different balances in the Balance of Payments, it is possible to estimate capital flight, usually defined as the accumulation of unrecorded foreign assets. There are many different ways of evaluating capital flight, which can give very different results.³³ In view of the advantages and disadvantages of each of the methods and data availability, we decided to apply a variation of the one developed by Eggerstedt, Brideau Hall and Van Wijnbergen (1995). Capital flight is defined as the remainder of four components of the Balance of Payments (foreign debt flows, foreign direct investments, variations in exchange reserves and the balance of the current balance) to which are added interest received on the capital stocks accumulated from capital flight in previous years.³⁴

Whereas in Madagascar there seems to have been an influx of foreign capital as of 1993, in a context of political instability and economic liberalisation, in Côte d'Ivoire, the trend was for income to be transferred abroad. These secret income flows represent around 1.5% of gross available income for households in Côte d'Ivoire and between 3 and 4% of gross income for Madagascar households. It is difficult to know what percentage of this money served to finance household consumption or investment transactions. However, in Côte d'Ivoire, it is likely that part of the income generated in the country but illegally invested abroad came from households living in the country, with another part coming from formal enterprises.

Table 12: Estimate of capital flight

	Madagascar	Côte d'Ivoire
% of gross available income		
National Accounts (a)	- 4.1	1.5
Surveys (b)	- 3.4	1.7
% of GDP (c)	-1.7	1.0

(a) 1995 in Madagascar and 1998 in Côte d'Ivoire.

(b) EPM93 in Madagascar and ENV98 in Côte d'Ivoire.

(c) 1993 in Madagascar and 1998 in Côte d'Ivoire

The correction made following this evaluation consisted in simulating the addition of income from capital flight in the income distribution. In Côte d'Ivoire, it was assumed that only 80% of these investments were made by households belonging to the last decile and this income was imputed to the households in the last decile which declared that they had received dividends. This simulation had very little impact on inequalities (cf. Table 13).

³³ See Appendix D, Guénard and Mesplé-Somps (2004) for details of the various calculation methods. Our assessments of the Ivorian and Madagascan data from 1986 and 1999 are given in the same document.

³⁴ To calculate interest on foreign investments, it was assumed that 1/3 of investments were made in France and 2/3 in Switzerland.

Table 13: Capital flight and income inequalities in Côte d'Ivoire.

Côte d'Ivoire	Without adjustment	With adjustment
Per capita income	380.7 [359 – 402]	384.8 [363.5 – 407]
Gini Index (%)	52.2 [50.3 – 55.6]	52.67 [50.74 - 54.78]

Adjustment : addition of 80% of income illegally invested abroad.

Confidence intervals in brackets, calculated by bootstrap; in thousands of F CFA or millions of current FMG.

Sources: ENV98 Côte d'Ivoire, our own calculations.

6. Summary of results and conclusions

What is to be learned from this examination of the different possible sources of biases and our various attempts to correct the survey data?

In the two cases studied, we detected an impact of lapses of memory, seasonality of consumption and income, and made choices for the means of annualising living standards aggregates. Nonetheless, our diagnosis shows that all these phenomena have relatively little impact on the assessment of living standards and their distribution. On the other hand, data entry errors, calculating imputed rents for households which own the housing they live in, and regional price variations have a large impact. The latter two factors increase average living standards and reduce inequalities in a significant manner. Errors in measurement and/or data entry and the corresponding corrections strongly bias the assessment of living standards and inequalities, particularly with respect to income. This observation strengthens the argument in favour of measuring living standards by levels of consumption rather than levels of income. However, an approach using consumption is not satisfactory, even if it is considered better, given that the inequalities measured in many countries (Latin America and western countries) come from data on income, and also because a study of the origin of inequalities requires an analysis of the formation of income.

Overall, it can be noted that the top-end of the distribution is not covered by the surveys, for the reasons mentioned above, i.e. problems with sample design, under-declarations and the fact that formal income distributed to the population of non-African origin is deliberately not accounted for. This is particularly highlighted by the evident absence of people with relatively high non-salary incomes. Székely and Hilgert (1999) noted the same phenomenon with a sample of Latin American countries.

Our work therefore consisted in proposing different scenarios to correct the data, in a view to making the income declarations more coherent with the declarations of expenditure and also to come closer to aggregates from other sources of national data. However, there are clearly problems with the overall quality of data both in the household surveys and in the National Accounts, making it difficult to compare the two and somewhat unrealistic to use one of the two sources of information to correct the other.

Although it is relatively easy to spot problems of internal coherency between declarations of income, savings and consumption, correction can be somewhat delicate. There are a large number of possible hypotheses and potential corrections to choose from in order to make the aggregates match better. We used the method consisting in replacing what were considered abnormal observations by multiple imputations, which we believe gives the most satisfactory results compared with other relevant methods currently in use. However, this method does not sufficiently correct certain declaration biases relating to living standards. In addition, it was impossible to assess or rectify certain biases, such as those relating to the absence of homeless people in the surveys. Whatever the case may be, the multiple imputation method shows that, at the very least, the average income in Madagascar must be adjusted by 14%, whereas matching with the National Accounts results in an adjustment of around 50%! In Côte d'Ivoire, adjusting under-declarations of income for the households surveyed increases the average income level by 17% and adding the population of foreign origin, by 34%. It seems necessary to add these corrections.

What is the impact of these different adjustments on inequalities in the two countries? Inequalities would be much higher in both cases (cf. Table 14).

In Madagascar, the multiple imputation method for correcting income results in a 2 point increase in the Gini index (43 compared with 41) whereas the addition of the population of foreign origin leads to increases, depending on the hypothesis retained, of 4 to 7 points in the Gini index. In Côte d'Ivoire, the first adjustment has no impact on the levels of inequality, whereas the second leads to an increase of the Gini index on incomes from 52 to 56, or even 62, depending on the scenario retained. Finally, if more credibility is given to the data from the Madagascar National Accounts than to the household survey and to the origin of the differences between the two data sources, the impact on inequalities can be very significant, up to 23 Gini points. These different corrections bring the inequality levels in the two countries nearer to those in the most inegalitarian countries, such as Brazil.

Are inequalities underestimated to the same extent in the other countries? Does this change the difference in levels of inequalities between countries? These questions can only be answered by wide-scale research, which statisticians and others who use these databases would be wise to undertake. At the same time, this work could help in drafting a frame of reference to assess the quality of the data available on living standards distribution.

Table 14: Summary of results

		Consumption aggregate		Income aggregate	
		Average per capita	Gini index	Average per capita	Gini index
Côte d'Ivoire (current F CFA)					
World Bank data ^a		319 136	45.15		
(1)	Our calculations ^b	348 989	43.6	380 713	52.2
<i>Impact of data entry errors and methodological choices</i>					
(2)=	(1) without corr. entry errors	349 070	43.6	422 468	57.0
(3)=	(1) without regional price deflators	315 023	45.4		
<i>Adjustment for sample design bias and under-declarations</i>					
(4)=	(1) plus foreign pop. of non-African origin (high hyp.) ^c	403 169	50.0	490 671	61.8
(5)=	(1) plus foreign pop. of non-African origin (low hyp.) ^c	373 199	46.2	425 519	56.2
(6)=	(1) adjusted for under-declarations of income by multiple imputation method ^d			402 695	52.1
Madagascar (current F MG)					
World Bank data ^a		237 099	43.4		
(1)	Our own calculations ^b	296 630	45.6	358 569	40.9
<i>Impact of data entry errors and methodological choices</i>					
(2)=	(1) without corr. entry errors	301 372	46.4	640 712	69.1
(3)=	(1) without regional price deflators	264 760	46.8		
<i>Adjustment for sample design bias and under-declarations</i>					
(4)=	(1) plus foreign population (high hypothesis) ^c	332 163	51.3	411 925	48.5
(5)=	(1) plus foreign population (low hypothesis) ^c	314 294	48.5	385 122	44.9
(6)=	(1) adjusted for under-declarations of income by multiple imputation method ^d			400 718	42.8
(7)	(1) adjusted for under-declarations of income by declared savings ^e			407 863	42.9
<i>Adjustment on basis of National Accounts data</i>					
(8)=	(1) plus 90% of income drawn by entrepreneurs in quasi-corporate undertakings in NAs ^f			643 329	64.0
(9)=	(1) plus 40% of income drawn by entrepreneurs in quasi-corporate undertakings in NAs ^f			485 129	53.3

In thousands of F CFA or millions of current FMG.

- a) Sources: EPM93, ENV98 Côte d'Ivoire, our own calculations.
 - Average consumption levels: <http://www.worldbank.org/research/povmonitor/>
 - Gini indexes: *World Development Indicators*, World Bank, 1998 and 2004.
- b) The consumption and income aggregates are calculated after eliminating data entry errors from the survey files and are deflated by regional prices.
Definition of consumption aggregate: current expenditure, including imputed rent, transfers and durable goods;
Definition of income aggregate: current income from productive salaried or non-salaried activities carried out during the year, from principle and secondary activities of all the members, from income generated by the households' assets (dividends, rent received and imputed rent for owners), and private and public net transfers.
- c) Addition of the foreign population of non-African origin in Côte d'Ivoire and the foreign population in Madagascar, on a high hypothesis (5) and a low hypothesis (6) of the living standards of non-African origin populations. High hypothesis: available income per 4-member household, per month of 4,500 euros, average consumption per household, per month of 2,290 euros. Low hypothesis: available income per 4-member household, per month of 2,100 euros (average French income, INSEE), average consumption per household, per month of 1,100 euros.
- d) Using the multiple imputation method for missing data (Rubin, 2004).
- e) Adjustment of income by comparing levels of consumption and declared savings (vs. residual savings).
- f) Addition of 90% of income drawn by entrepreneurs in quasi-corporate undertakings given in the National Accounts to households which declare themselves to be "formal" entrepreneurs or to receive dividends.

Bibliography

- Allison (1999), « Multiple Imputation for Missing Data: A Cautionary Tale ». Mimeo University of Pennsylvania, 14p.
- Appleton, S. (2003), « Regional or National Poverty Lines? The Case of Uganda in the 1990s », *Journal of African Economies*, vol. 12, n°4, December, pp. 598-624.
- Atkinson A.B., Brandolini A. (2001), « Promise and Pitfalls in the Use of “Secondary” Data-Sets: Income Inequality in OECD Countries as a Case Study », *Journal of Economic Literature*, vol.39, September, pp771-799.
- Banerjee A., Piketty T. (2003), « Top Indian Incomes, 1922-2000 », *Bread Working Paper* 46, December.
- Beauchemin C. (2004), « Pour une relecture des tendances migratoires internes entre villes et campagnes : une étude comparée Burkina Faso-Côte-d’Ivoire », *Cahiers Québécois de démographie*, vol.33, n°2, automne, pp167-179.
- Bhalla S.S. (2002), « Imagine there’s no country: Poverty, Inequality and Growth in the era of Globalization », Washington D.C., *Institute for International Economics* (IIE), October, 288 p.
- Bocquier P. (2004), « World Urbanization Prospects: an alternative to the UN model of projection compatible with urban transition theory », Document de travail DIAL N° DT/2004/08, Octobre.
- Bourguignon, F. (2002), « The growth elasticity of poverty reduction: explaining heterogeneity across countries and time periods », Working paper DELTA n°2002-03.
- Bourguignon F., Morrisson C. (2002), « Inequality Among World Citizens: 1820-1992 », *The American Economic Review*, vol.92, n°4, pp727-744.
- Census Bureau (2003), « Current Population Survey Design and Methodology », Technical Paper 63, Washington D. C., US Department of Commerce.
- Chen S., Ravallion M. (2004), « How the World’s Poorest have Fared Since the Early 1980s ». *The World Bank Research Observer*, vol. 19, n°2, fall, pp. 141-170.
- Cling, J.-P., De Vreyer P., Razafindrakoto M., Roubaud F. (2004), « La croissance ne suffit pas pour réduire la pauvreté ». *Revue française d’économie*, vol. XVIII, n°3, janvier, pp. 137-188.
- Cogneau D., Mesplé-Somps S. (2002), « L’économie ivoirienne, la fin du mirage ? », *Document de Travail DIAL* 2002/18, Paris.
- Deaton A. (1997), *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*, Johns Hopkins University Press, World Bank, August, 480p.
- Deaton A. (2001), « Counting the World’s Poor: Problems and Possible Solutions », *World Bank Research Observer*, Fall 2001, v. 16, iss. 2, pp. 125-147.
- Deaton A (2004), « Measuring Poverty in a Growing World (or Measuring Growth in a Poor World) », NBER Working Paper n°9822, February.
- Deining K., Squire L. (1996), « A New Data Set Measuring Income Inequality », *The World Bank Economic Review*, vol.10,n°3, pp565-591.
- Eggerstedt, H., Brideau Hall R., Van Wijnbergen S. (1995), « Measuring Capital Flight, a Case Study of Mexico », *World Development*, 23(2): 221-232
- Gibson J., Huang J., Rozelle S. (2003), « Improving Estimates of Inequality and Poverty from Urban China’s Household Income and Expenditure Survey », *Review of Income and Wealth*, series 49, number 1, March, pp. 53-68.
- Gautier E. (2005), *Eléments sur les mécanismes de sélection dans les enquêtes et sur la non-réponse non-ignorable*, mimeo, INSEE, Unités méthodes Statistiques, Paris.
- Guénard C., S. Mesplé-Somps (2004), « Mesurer les inégalités : que captent réellement les enquêtes ? Analyse de deux enquêtes ivoirienne et malgache », *Document de Travail DIAL* 2004/13, Paris, 47p.

- Grimm M., Guénard C., Mesplé-Somps S. (2002), « What has happened to the urban population in Côte d'Ivoire since the eighties ? An analysis of monetary poverty and deprivation over 15 years of household data », *World development*, June vol 30 n°6, pp. 1073-1096.
- Heckmann J. (1979), « Sample Selection Bias as a Specification Error », *Econometrica*, vol.47, n°1, pp 153-161.
- INSTAT (1997), Recensement général de la population et de l'habitat, 1993, vol. 2 Tome 1 Etat de la population; Tome 7, Ménages et habitat; Tome 2, Education, *Institut National de la Statistique Malgache*, Janvier.
- Jones C., Ye X. (1997), « Issues in Comparing Poverty Trends Over Time in Côte d'Ivoire », World Bank Policy Research Working paper n°1711, January, 77 p.
- Lerman R. I., Lerman D. L. (1989), « Income Sources and Income Inequality: Measurements from Three U.S. Income Surveys », *Journal of Economic and Social Measurement*, 15, pp. 167-179.
- Lillard L., Smith J.P., Welch F. (1986), « What Do Really Know about Wages ? The Importance of Nonreporting and Census Imputation », *Journal of Political Economy*, vol.94, n°3, pp489-506.
- Loisy C. (1999), « L'épargne des ménages de 1984 à 1995: disparité et diversité », *Economie et Statistique*, n°324-325, INSEE, pp113-133.
- Mistiaen J.A., Ravallion M. (2003), « Survey Compliance and the Distribution of Income », World Bank, Working Paper n°2956, January.
- Piketty T. (2003), « Income Inequality in France, 1901-1998 », *Journal of Political Economy*, vol. 111, pp.1004-1042.
- Piketty T., Saez E. (2003), « Income Inequality in the United States, 1913-1998 », *The Quarterly Journal of Economics*, vol.118, Issue 1, February, pp1-41.
- Pyatt G. (2003), « Development and the Distribution of Living Standards: A Critique of the Evolving Data Base », *Review of Income and Wealth*, n°3, Series 49, September, pp 333-358.
- Ravallion M. (2000), « Should Poverty Measures Be Anchored to the National Accounts? », *Special Articles, Economic and Political Weekly*, August 26-September 2, pp 3245-3252.
- Ravallion M. (2001), « Measuring welfare in developing countries : how well do national accounts and surveys agree? », World Bank Policy Research Working Paper, n°2665, August.
- Rubin D. B. (2004), *Multiple Imputation for Nonresponse in Surveys*. Wiley Classics Library, 287 p. (first edition 1987)
- Sala-i-Martin X. (2002), The World Distribution of Income (estimated from Individual Country Distributions), *NBER Working Paper No. 8933*, May.
- Székely M., M.Hilgert (1999), « What's Behind The Inequality We Measure : An Investigation Using Latin American Data », *Inter-American Development Bank Working Paper n° 409*, December.
- Székely M., Lustig N., Cumpa M., Mejia J-A. (2000), « Do we know how much poverty there is? » Inter-American Development Bank Research Working Paper no. 437, December, 40 p.
- Visaria P. (2000), « Poverty in India during 1994-98: alternative estimates, Institute for Economic Growth », New Delhi, processed, June 9.
- WIDER (2000), « World Income Inequality Database », UNU/WIDER-UNDP », version 1.0, 12 septembre 2000, téléchargeable sur www.wider.unu.edu/wiid/wiid.htm.

APPENDIX A: Presentation of ENV98 survey in Côte d'Ivoire and EPM93 survey in Madagascar

The sample size was quite similar in the two countries: 4,303 households in the Madagascan survey (22,710 individuals) and 4,200 households in the Ivorian survey (24,211 individuals). This represents relatively large samples for developing countries but quite small ones to ensure “correct” precision of indicators. Both surveys were built to be representative on a national level; breakdown by geographical strata was as follows: 2 urban (Abidjan and Other Towns) and 3 rural (Forest East, Forest West and Savannah) in Côte d'Ivoire; 6 regions (faritany) in Madagascar (Antananarivo and neighbouring area, Tuléar, Antsiranana, Fianarantsoa, Toamasina and Mahajanga) broken down into urban and rural areas for each region. In the Ivorian survey, the sample was made up of a total of 210 clusters of 20 households each (43 clusters in Abidjan, 53 in the other towns and 114 in rural areas). EPM93 was based on repeat visits and data collection was organized in different ways depending on the area: in urban areas, a cluster of 12 households was surveyed for a reference period of one month and in rural areas, clusters of 16 households for a reference period of 7 days. The survey took place with 10 cycles of surveys, each cycle corresponding to about one month of survey during which two clusters were interviewed. In principle, this procedure minimises problems relating to seasonality.

Consumption:

We used a relatively wide definition of the consumption aggregate, which includes all food and non-food current consumption goods, self-consumption (including non-food goods on the part of informal entrepreneurs in the case of Madagascar), purchases of durable goods during the year⁶⁹, net transfers and a imputed rent for house owners.⁷⁰

Income:

Methodological choices must also be made for the income calculations, which depend to a great extent on the type of survey and questionnaires.

As in the case of consumption, the differences between the two questionnaires prevent us from calculating the two countries' income in the same way, which makes it difficult to compare their standards of living.⁷¹ For both surveys, the chosen definition for current income is the sum of:

- income from productive salaried or non-salaried activities carried out during the year, from principle and secondary activities of all the members,
- income generated by the households' assets (dividends, rent received from rented housing or land and imputed rent for owners),
- private and public net transfers.

Amounts received from sales of assets owned by the household (house, land, animals, etc), loans contracted during the year, inheritance or dowries received and savings cumulated in the previous year are not considered as current income.

There are three possible methods for estimating the value of sales of farm produce: the first is based on declarations of income from sales of farm produce; the second is based on the quantities sold and the declared unit price and finally the third applies an average sales price to the quantities sold. Similarly, self-consumption can be estimated by applying either individual sales prices or an average price per product to the quantities given for self-consumption. There is a good case for using average (or median) prices per product in order to control errors in declarations which proved to be fairly frequent (problem of matching the unit of measurement of the

⁶⁹ Unfortunately, we could not evaluate the value of durable goods already owned by the household.

⁷⁰ Housing produces a service that can be sold (owner renter) or consumed by the holder (owner occupier). It is not the value of the housing that is consumed but the value of the service provided by it. Thus, whether or not imputed rents are taken into account for people owning their housing is a methodological choice. Before estimating the model for imputed rents, a selection equation (probit on being tenant or not) was estimated to take into account any effects of selection that may lead to a bias in the estimators, using Heckmann's two-stage process (1979).

⁷¹ The household questionnaire for EPM93 goes into far more detail on the components of income than for ENV98: for example, it includes information on the origin of farming revenues and details on intermediate consumption. The EPM93 provides data on the individuals' income from the four potential activities in the 12 previous months, the main activity being the one to which the individual devotes the most time (and not the most lucrative), which is a quite unusual amount of detail. On the other hand, in the ENV98, multi-activity is assessed by a single questionnaire on secondary activities.

quantity of product and the unit price, data entry errors, etc.). On the other hand, choosing an average price, whether by region or not, is problematical and there is a risk of artificially reducing income dispersion. In a view to harmonising the calculation method with that of the Madagascan survey, we decided to estimate farming income using overall sales amounts and to assess self-consumption by the declared unit prices (line 1, Table 15). It can be noted that, in the end, the choices concerning farming income do not have a real impact either on the averages or on the levels of income inequality.

Table 15: Impact of methodological choices on income aggregate, Côte d'Ivoire

	Average per capita (Current F CFA)*	Gini Index
1	380.7 [359.2 – 402]	52.2 [50.3 – 55.6]
2	381 [359.4 – 402.7]	52.3 [50.5 – 55.4]
3	376.5 [355 – 398]	52.2 [49.7 – 53.9]

* Aggregates deflated for regional price variations.

1. value of sales of farm produce calculated using declarations of income from sales, and self-consumption calculated using declared unit prices..
2. value of sales of farm produce calculated using quantities sold and declared unit prices and self-consumption calculated using declared unit prices..
3. value of sales of farm produce calculated by applying an average sales price to the declared quantities sold and self-consumption calculated using average sales prices for the same products at the markets.

Confidence intervals in brackets, calculated by bootstrap ; in thousands of F CFA or millions of current FMG.

Sources: ENV98 Côte d'Ivoire, our own calculations.

APPENDIX B: Structure of consumption and income by quartiles, deciles and centiles.

Table 16: Budget shares by extreme quartiles, deciles and centiles of per capita consumption.

	total	q1	q2	q3	q4	d1	d2	d3	d8	d9	d10	c98	c99	c100
Madagascar														
Food	61	72	71	69	52	70	73	73	63	63	46	49	51	27
Including self-consumption (%)	40,4	58	55	45	29	55	58	56	40	38	21	23	15	18
Clothing	5	7	6	6	5	7	6	7	6	5	5	5	4	4
Housing	10	10	8	8	11	12	9	8	9	10	11	12	10	12
Of which rent (inc. imputed rent) (%)	50	72	64	57	43	76	71	67	56	50	38	38	38	21
Equipment, maintenance	4	3	4	4	4	3	4	3	4	4	4	4	3	3
Transport	3	1	1	1	4	0	1	1	2	2	6	4	4	11
Leisure	1	0	1	1	2	0	0	0	1	1	2	2	3	3
Education, Health	5	3,3	3,4	5	6	3	3,3	3,3	6	6	7	6	5	5
Durable goods	6	1	2	2	10	0	1	1	4	5	14	10	8	30
Other goods	5	3	4	5	6	4	3	3	5	5	7	7	12	5
Including ceremonies (%)	57	55	62	57	55	59	51	70	47	57	57	72	87	87
Total	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Côte d'Ivoire														
Food	55	57	56	55	55	59	56	57	54	54	57	63	55	51
Including self-consumption (%)	23	45	33	24	11	50	44	38	18	13	10	0	1	4
Clothing	7	8	8	7	6	9	9	8	7	6	5	5	6	5
Housing	14	18	16	14	11	17	17	17	13	13	10	10	9	8
Of which rent (inc. imputed rent)(%)	62	69	65	59	58	70	70	65	57	57	60	63	68	50
Equipment, maintenance	3	3	3	3	3	3	3	3	3	4	3	3	3	4
Transport	7	4	5	7	9	3	4	4	8	9	9	6	9	8
Leisure	1	1	1	1	2	1	1	1	1	2	1	1	1	3
Education, Health	9	9	10	10	9	8	9	8	11	10	7	8	8	7
Durable goods	2	1	1	2	3	0	1	1	1	1	4	1	5	8
Other goods	2	1	1	2	3	1	1	1	2	2	3	2	3	6
Total	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Sources: EPM93 Madagascar, ENV98 Côte d'Ivoire, our own calculations.

Table 17: Structure of aggregate per capita income, by extreme quartiles, deciles and centiles (%)

	Total	(a)	q1	(a)	q2	(a)	q3	(a)	q4	(a)	d1	d2	d3	d8	d9	d10	d10/d1	c98	c99	c100	10 + riches
Madagascar																					
Wages & Salaries	22		17		20		20		25		17	19	17	24	25	25	13	29	17	25	38
formal	15	68	6	36	10	50	13	63	21	81	4	7	8	17	20	20	41	24	16	22	33
informal	5	21	6	33	5	27	6	29	3	14	6	6	4	5	4	3	5	2	1	3	0
farming	3	11	5	31	5	24	4	8	6	6	2	3	3	2	1	3	2	1	0	0	0
Self-employ. incomes	66		71		70		70		61		69	70	71	66	64	60	8	49	66	59	23
farming	59	88	62	88	64	91	63	90	53	86	59	61	65	60	55	50	8	41	56	52	16
non farming	8	12	9	12	6	9	7	10	8	14	10	8	5	6	9	9	8	9	10	7	7
Wealth	5		9		5		5		5		8	5	5	5	5	5	5	4	5	6	20
dividends	1	21	0	5	1	16	1	24	1	26	0.5	0.5	0.6	1	2	1	24	2	2	1	2
Unearned incomes (b)	4	79	9	95	4	84	4	76	4	74	7	4	4	4	3	4	89	2	3	5	17
Pensions, insurance	0		0,0		0,1		0,1		0,1		0,0	0,1	0,1	0,0	0,0	0,1	20	0,0	0,1	0,1	0,8
Transfers	3		3		2		2		3		2	4	3	2	3	4	14	6	3	4	9
public	1.1	39	0.6	24	0.7	30	0.7	56	1.4	39	0.6	0.2	1.3	1.3	1.1	1.4	21	2.1	0.5	1.3	9
private	2	61	3	76	2	70	1	44	2	61	2	4	2	1	2	3	12	3	3	3	0
Other income	4		3		3		3		5		3	2	4	3	3	7	23	12	10	6	10
TOTAL	100		100	100	100	100	100	100	10	100	100	100	100								
Côte d'Ivoire																					
Wages & Salaries	28		17		27		29		29		16	19	20	34	27	28	57	13	33	33	47
Public	8	29	2	11	7	27	10	32	8	30	0	3	2	13	11	7	42	2	8	5	0
Private	20	71	15	89	20	73	20	68	20	70	16	16	18	21	16	22	44	11	25	28	47
Self-employ. incomes	60		64		58		57		61		64	63	62	53	62	62	31	69	59	60	43
farming	25	42	41	64	33	57	30	53	19	31	40	41	39	24	25	16	13	10	8	4	4
non farming	35	58	23	36	25	43	27	47	42	69	25	22	23	29	37	47	60	56	51	57	39
Wealth (b) (c)	6		10		8		6		5		11	12	11	6	5	4	12	13	3	3	6
Pensions, Insurance	2		1		1		1		2		1	1	1	2	2	2	56	3	2	2	1
Transfers	4		6		5		4		3		7	5	5	3	4	2	9	1	1	1	2
public	1	22	1.1	17	1.0	17	1.0	25	0.7	24	1,3	1,1	0,9	1,0	0,8	0,5	8	0,3	0,6	0,3	0,1
private	3	78	5	83	4	83	3	75	2	76	7	5	5	2	3	2	8	1	1	1	2
Other income	1		1		1		2		1		1	1	1	2	1	1	33	2	2	1	1
Total	100		100	100	100	100	100	100	35	100	100	100	100								

ENV98 expressed in regional prices for 1998; EPM93 expressed in regional prices for 1993

(a) % intra category; (b) including imputed rent for housing occupied by the owners; (c) exclusively income from the property.

Sources: EPM93 Madagascar, ENV98 Côte d'Ivoire, our own calculations.

APPENDIX C: Comparison of income from surveys with the National Accounts

Table 18: Comparison of consumption levels: Surveys / National Accounts

Madagascar		CN95		EPM93 at 1995 values	
Items	(a)	Budget shares	(a)	Budget shares	EPM93 at 95 values as % of CN95
Food	6 311	51,1	4 174.8	60.6 40,4% (c)	66.1
Clothing	1 114	9,0	376.3	5.5	33.8
Housing (b)	624.6	5,1	405.3	6.6 50,5% (d)	64.9
Equipment, maintenance	1 251.4	10,1	254.2	3.7	20.3
Leisure	914	7,4	90.7	1.3	9.9
Durable goods	311.6	2,5	426.5	6.2	136.9
Transport	445.6	3,6	193	2.8	43.3
Healthcare and domestic services	820.4	6,6	63.1	0.9	7.7
Other goods	164.7	1,3	347.2	5.0	210.8
Education and health	389.7	3,2	285.8	4.2 57% (e)	73.4
TOTAL	12 347	100	5 806.5	100	53,5%
Côte d'Ivoire		CN98		ENV 98	
Postes	(a)	Coefficients budgétaires	(a)	Coefficients budgétaires	ENV98 en % de CN98
Food	169	52	165	56	97,8
Clothing	17	5	20	7	117,3
Housing	31	10	41	14	130,8
Equipment, maintenance	25	8	9	3	37,2
Health	2	0	17	6	1088,7
Transport	28	9	20	7	72,9
Leisure	7	2	4	1	53,2
Education	1	0	10	3	1289,8
Durable goods	22	7	4	2	20,9
Other goods	21	7	5	2	25,4
Total	323	100	297	100	92%
Level of consumption (millions current FCFA)	4.637.616		4.305.210		92,8%

(a) Per capita consumption in millions of FMG or thousands of current F CFA; (b) not including imputed rents; (c) including self-consumption; (d) of which rent, including imputed rents; (e) including ceremonies
Sources: EPM93 Madagascar, ENV98 Côte d'Ivoire, our own calculations.

Table 19: Comparison of incomes from National Accounts and Household Accounts, Madagascar

Millions of current FMG	CN95 at 95 values	EPM93 at 95 values	EPM93 at 95 value (% of CN95)
Production account			
Resources			
Production	766.4	7 738.4	1 009.7
Uses			
Intermediate consumption	43.6	4 197.5	9 617.8
Gross value added	722.7	3 540.9	489.9
Generation of income account			
Gross Value Added (RESOURCE)			
	722.7	3 540.9	489.9
Uses			
Compensation of employees	22.9	246.3	1 075.2
Gross wages and salaries	20.1		
Employers' social contributions	2.8		
Taxes on production and imports	1.1	56.9	4 991.8
Gross Operating Surplus (GOS) (a)	698.7	2 682.5	383.9
Allocation of primary income account (millions of current FMG)			
Resources			
GOS	698.7	2 682.5	383.9
Compensation of employees	3 923.9	1 906.9	48.6
Gross wages and salaries	3 773.6	1 906.9	50.5
Employers' social contributions	150 228		
Property income	7 434.4	803.3	10.8
Distributed income of corporations	7 319.2	719.3	9.8
Property income attributed to insurance policyholders	8.2		
Rent	106.9	83.9	78.5
Primary disposable income (USES)	12 057	5 947.9	49.3
Secondary distribution of income account			
Resources			
Primary disposable income	12 057	5 948	49.3
Social benefits other than current transfers in kind	128.4		
Other current transfers	577.5	400.5	69.4
Uses			
Current taxes on income, wealth	59.8	147	0.2
Social contributions and benefits	161.2	81.3	50.4
Other current transfers	232.5	144	61.9
Gross disposable income	12 309	6 123	49.7

(a) In the National Accounts, the balance of this account is called Mixed Income and the GOS, the normal balance of this account, is calculated and not residual. Due to lack of information on the meaning of Mixed Income and the means of obtaining the GOS, the choice was made to group these two aggregates, thus making a comparison with the GOS calculated from EPM93 more coherent.

In millions of current FMG.

Sources: EPM93 Madagascar, ENV98 Côte d'Ivoire, our own calculations

Table 20 : Comparison of incomes from National Accounts and Household Accounts - Côte d'Ivoire

INCOME	CN98 (a)	ENV 98 (a)	ENV98 as % of CN97
Resources			
GOS	239	203	85.2
GOS farming		85	
GOS non farming		118	
Wages and salaries	115	90	78.7
Public		27	
Private		63	
Income from wealth	12	19	
Interest and dividends	1		
Other property income	11	19	
Compensation from damage insurance (b)	2	5	
Social benefits	7	3	
International private transfers	11		
Other current transfers		10	
Other income		4	
Total	386	321	83.2
Uses			
Interest and dividends	4		
Other property income	4		
Net insurance premiums	1		
Current taxes on income and wealth	7		
Social contributions	6		
International private transfers	17		
Other current transfers	1	19	
Gross disposable income	350	316	90.3

(a) In thousands of current F CFA per capita, without taking into account regional price differences.

(b) For the survey data, this concerns income from retirement pensions and insurance.

Sources: EPM93 Madagascar, ENV98 Côte d'Ivoire, our own calculations.