Expenditure: Report on NIDS Wave 1

Technical Paper no. 4

Arden Finn
Graduate student, University of Cape Town
ardenfinn@gmail.com

Simon Franklin
Graduate student, University of Cape Town
Franklin.simon@gmail.com

Malcolm Keswell
Department of Economics, University of Stellenbosch
keswell@sun.ac.za

Murray Leibbrandt
Southern African Labour & Development Research Unit
Murray.Leibbrandt@uct.ac.za

Jim Levinsohn
Department of Economics, University of Michigan
jamesl@umich.edu

July 2009

Contents

1. Introduction

This report describes the construction of an aggregate household expenditure variable for the 2008 National Income Dynamics Survey (NIDS.) Section 2 discusses the construction of the food and non-food aggregate expenditure variables. Section 3 explains the construction of the housing expenditure variable. Each of these sections briefly discusses the questionnaire design, the patterns of non-response in the data, and the methods used to impute missing values. In section 4, we compare our constructed aggregate expenditure with the reported aggregate expenditure.[1] In section 5, we conduct some basic analyses using the derived total expenditure variable. .We examine the components of household expenditure. We also analyze differences in expenditures by key descriptors such as race and geotype. The analysis concludes by comparing the derived total expenditure variable to the derived total income variable.

We note from the outset that the expenditure data are collected at the household level. Our imputations correct for item non-response for specific expenditure items for responding households. We do not address issues on non-response at the household level.

The purpose of the imputations is to allow construction of a complete set of household expenditure aggregates. The public release of the data includes five derived expenditure variables (using the prefix "w1_" for wave 1):

> **w1_h_expf**: Total monthly food expenditure (the sum of 32 food items) with full imputations
> **w1_h_expnf**: Total monthly non-food expenditure (the sum of 52 non-food items) with full imputations
> **w1_h_rentexpend**: Actual amount spent on rent per month by non-owning renters with full imputations.
> **w1_h_rentexpend_source**: A flag that indicates whether the w1_h_rentexpend value was imputed for a particular household or not
> . **w1_hhimprent_exp**: True or implied monthly rental expenditure with full imputations[2]
> **w1_h_expenditure**: Total aggregate monthly household expenditure with full imputations

---

[1] We do not use the survey weights in sections 2 through 4, but we do in section 5.

[2] Note that total housing expenditure ought to be made up of both rental expenditure and expenditure on utilities and rates. Utilities expenditure is currently included as a sub-item in the decomposition of non-food expenditure, as the relevant questions were asked in this section. Interested users can use the expenditure do files referred to in this document to recover the utilities sub-aggregate.

The total expenditure variable with full imputations (**w1_h_expenditure)** is simply the aggregation of rental, food, and non-food expenditures. It can be compared to **w1_h_exprough** which is total expenditures with no imputations and no implied rental expenditure. Section 4 does just this.

2. Imputations for the Food and Non-Food Variables

2.1 Layout of the data and general approach to imputation

2.1.1 Sub-items and aggregation

The expenditure section of the NIDS survey lists 32 items of food expenditure and 53 items[3] of non-food expenditure. For each item the household head is asked whether the item was consumed in the last month and, if it was, what the monetary value of that consumption was. In this way there are two sources of item non-response: cases where there are missing values (either refused, unknown or simply not recorded) for monetary values when the household *did* consume that item, and cases where a household simply did not report consuming any goods, or the entire section was not completed in the interview. Any household that reported consumption of any goods and supplied the relevant quantitative values is considered to be a full response.

Note that the food section differs slightly from the non-food section in that consumption expenditure is decomposed into four separate sources of food, with four corresponding questions; how much was *spent* on the item, the value of that item received as *gifts*, the value received as *payment*, and the value from own *production*. Since it is impossible to predict the source of consumption when no value was reported, and since the majority of value responses were for money *spent*[4], these four values were aggregated and imputations performed to arrive at the aggregate consumption figure.

Once all missing item values were imputed, the complete set of values for each item was summed to find the total food and non-food expenditure for each individual. In this way, any households that reported consuming any items will have expenditure aggregates. However, there are cases where no consumption was reported at all (not just non-response for the monetary value of that item's

---

[3] The 53rd and final non-food expenditure item is "Income Tax Payments". Since the entire survey works with income after tax, we do not include this expenditure item in our expenditure aggregates, and it is excluded from the discussion that follows.

[4] At least 95% of total food consumed came from "money spent" on food, for every single item in the survey. In total 97.9% of food consumption was from money spent, 0.95% came from household's own production, 0.41% came in the form of payment and 0.69% came in the form of gifts.

consumption). In these cases, the existing expenditure aggregates are used to impute (in the same way as for each item) the entire missing aggregates.

2.1.2 Rules of thumb

As with the imputations for income figures, in those cases where there were fewer than 100 observations of expenditure of given item, the missing values would not be imputed but rather set to the population median. However, if the number of items to be imputed was more than 40% of the value of observations, no imputation of any type was performed. Fortunately this was not necessary for items in this section as response as response rates were generally very high. Appendix A provides a detailed outline of each consumption item, the rates of missing data and the action taken to deal with item non-response. All food items were fully imputed, while there were 6 non-food items that suffered from too few observations and used medians instead.

In addition, Appendix A provides the median value of consumption for both the original and imputed aggregates. This not only gives us a picture of the relative share of different line items to total household expenditure, but also serves as a rough indicator of the pattern of non-response; as we can see how the imputed aggregates differ from the original data. As expected, the missing items are not missing completely at random; the missing items that have been imputed usually have a considerably lower median than the non-missing observations. It should be noted that in many cases adding imputed values does not alter the medians, but it does decrease the *mean* in almost every case.

2.2 Quality of the data and patterns of non-response

Item non-response is particularly prevalent in the expenditure data because expenditure is decentralized and irregular within the household budget. For instance, there are 22,524 cases of item non-response in the non-food section, and 5,695 in the food section. These are cases, of course, where the respondent has reported that the item was consumed during the last 30 days, but the value of the consumption is not known. All of these missing variables have to be imputed.

Appendix A provides a full breakdown of each of expenditure item and the rates of non-response in each case. Some of the non-food items have observation counts that are too low, in which case the rules of thumb have been used and an asterisk has been used to indicate this. For all other items normal imputation techniques were used. Otherwise it is noted that there are no consumption items for which more than 40% of observations are missing; in fact most items have around 10% of their observations missing. In total 11% of non-food items are missing and 10% of all food items. In addition, this table provides the median values of consumption for the items prior to imputation. This gives us a picture of the relative share of different line items to total household expenditure.

In addition to non-response to particular expenditure items there are households that either report no expenditure whatsoever, or completely omit the expenditure section. These households have to have their entire food or non-food (or both) expenditure values imputed, depending on which is missing.

Table 1 below gives a brief summary of missing expenditure data. “Partially missing” indicates that at least one sub-item was missing (as opposed to the entire aggregate, where every item was missing). As the table indicates, more than 40% of all households in the sample have at least one case of missing expenditure data.

Table 1: Rates of Imputation

	Number Households Imputed For	% Households Imputed For
Entire Aggregate Imputed		
Food imputed	46	0.63%
Non-food imputed	191	2.61%
One imputed	216	2.96%
Both	21	0.29%
Partial Imputation		
Food partially imputed	1,904	26.06%
Non-food partially imputed	2,227	30.49%
Any imputation	3,027	41.44%

2.3 Imputation Techniques

Two different imputation techniques, cell median and regression-based, were used to impute the missing expenditure items. In this section these two techniques are presented.[5] Finally, the results from the two techniques are compared. A close comparison of the imputed values shows remarkably little difference in the results of the two methods. Given this, it was decided that for the sake of consistency with the income section for all imputations that a regression imputation approach should be taken in the key expenditure variables. A brief overview of the two techniques and a comparison of the results are given below:

2.3.1Cell Median Imputations

This technique takes the median of each expenditure value for a certain subgroup of the population and uses that as the imputed value for the missing item. A minimum response rate threshold is used to prevent a large number of missing values within a cell being imputed from just a few observations of nearby households[6]. If less than 60% of a given cell reported the value of consumption, the cell is broadened to incorporate a larger number of observations until the rate of response is high enough. If necessary this process is repeated until the entire province is used, and since there are almost no items with response rate below 60% in an entire province, all values are eventually imputed. The cells that are used, from smallest to broadest are PSU, District and Province.

Once all sub-items are aggregated into total food and non-food variables, any households that didn't report consumption for any items have their entire aggregate food or non-food expenditure imputed using the same system of cell medians for the aggregates.

2.3.2 Regression Imputations

A regression is run with the same independent variables for each imputed item. This regression is then used to predict the missing the values using the known demographic variables. In those cases

---

[5] The relevant do files for these imputations are available at http://www.nids.uct.ac.za/home/index.php/Welcome/datasets.html

[6] For instance there could be PSU with 30 households, of which 10 consume a certain food item, but only 3 actually knew how much they had spent. Using the median in this case would give seven households a median figure from only 3 observations, violating the 40% cutoff rule.

where no expenditure is reported, a similar regression is used to impute the overall food and non-food expenditure. The regressions provide predictable results; incomes, size of household and maximum education consistently have a positive impact on the value of the goods consumed. There are exceptions to this, however, and these include goods such as mealie meal and samp.

The variables used as independent variables in all imputations are as follows:

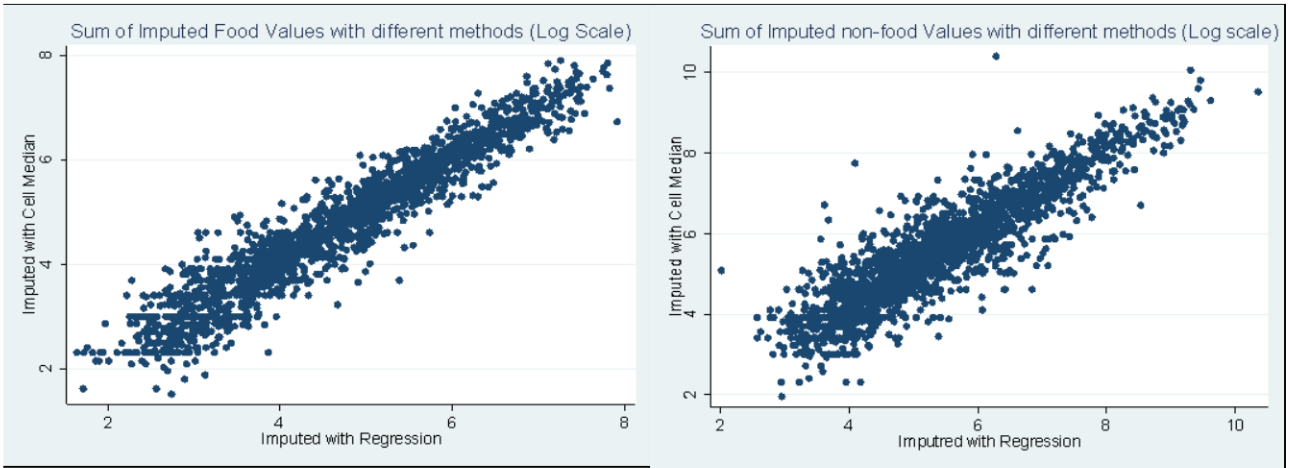
Income (log), number of rooms in the dwelling, race dummy variables, province dummy variables, household size, maximum age of household members, maximum education attainment of household members, dummy for a corrugated roof, dummy for the household being a grant recipient, dummy for the household including any children.

For each regression there are only a few variables that are statistically insignificant, and all variables used in the standardized regression are statistically significant for a majority of the regressions. The adjusted R-square value for these regressions generally fell in the range of 30% to 55%.

2.3.3 Comparison of Cell Median and Regression Techniques

A comparison of the final expenditure aggregates is likely to be misleading, as only around 10% of all observations are missing and thus most of the aggregates will be composed of the same valid observations. Instead we need to compare the imputed values in isolation. This is done on an item by item basis (the results of which are presented in Appendix B) and by comparing the sum (for each household) of only the imputed values for each imputation method. Figure 1 below presents a scatter plot that gives a good first impression of how similar the results are. The data points are simply the sum of just the imputed food/non-food values for each household (such that there 1904 observations for the food sections and 2227 for the non-food section).

Figure 1: Comparison of Cell Mean and Regression Imputations for Food and Non-food Expenditures



The comparison is relevant since the same missing values need to be imputed with both methods for each household. The results are remarkably similar given that the cell median method uses the median of only those households in geographical proximity while the regression method uses only demographic variables that are unrelated to the household's PSU. No doubt this reflects, to a large degree, the homogeneity of demographic variables within PSU's, but it is still a reassuring sign of the robustness of the imputations.

The distributions of the two imputation methods for both food and non-food expenditures is presented in Table 2 below. This exercise is similarly reassuring. The median method gives slightly higher means and medians, and slightly lower standard deviations. In general the two methods give very similar results at every percentile.

Table 2: The Distribution of Food and Non-food Imputed Values By Imputation Method

	Food Expenditure Imputations					
	Regression Imputation Method			Median Imputation Method		
	Percentiles			Percentiles		
1%	159.1858			158		
5%	265.5189			284		
10%	343.5306	Obs	1904	374	Obs	1904
25%	537.5811	Sum of Wgt.	1904	563.25	Sum of Wgt.	1904
50%	885.4087	Mean	1114.506	923	Mean	1145.72
		Std. Dev.	869.7476		Std. Dev.	859.9048
75%	1393.401			1448.5		
90%	2195.578	Variance	756460.9	2203	Variance	739436.3
95%	2718.222	Skewness	2.737124	2790	Skewness	2.515569
99%	4152.809	Kurtosis	18.56118	4236	Kurtosis	15.82855
	Non-Food Expenditure Imputations					
	Percentiles			Percentiles		
1%	40.7132			30		
5%	111.018			102		
10%	160.8835	Obs	2199	167	Obs	2199
25%	322.6566	Sum of Wgt.	2199	350	Sum of Wgt.	2199
50%	1019.775	Mean	3437.663	1110	Mean	3499.585
		Std. Dev.	7608.143		Std. Dev.	7567.497
75%	3445.108			3607.5		
90%	8598.101	Variance	5.79E+07	8425	Variance	5.73E+07
95%	13482.38	Skewness	6.866488	13890	Skewness	6.880103
99%	33712.37	Kurtosis	71.53214	33729	Kurtosis	72.32099

On an individual item basis (as presented in Appendix B) the results are similar. The cell median method gives a higher median for some items and a lower median for others. Recall that these are only the imputed expenditure values, yet the summary statistics (including the range of values) are remarkably similar.

We end this section by comparing the mean and median imputations for the aggregated (food plus non-food) expenditures. The results are presented in Table 3 below.

Given that over 30 000 imputations were performed and over 3 000 of the 7306 household expenditure aggregates have had some kind of imputation performed, one might expect great disparities between any two methods of imputation, such as the median and regression methods outlined in the section above. Yet the summary statistics look remarkably similar; the medians a very small amount. In addition, we find that only 485 households fall in a different expenditure quintile when the quintiles of the two measures are compared. In fact 443 of these households fall in the adjacent expenditure category. Table 4 demonstrates the similarity between the two measures in matrix form. This similarity gives us some reassurance about the accuracy of either measure. However for the rest of this report the **regression** imputation variables have been used, as they provide more reasonable variation within PSU's, while still managing to reflect the differences between PSU's purely due to their demographic make-up.

Table 3: Percentage of Households in Respective Food + Non-food Expenditure Quintiles

Quintile of Regression Imputed Expenditure	Quintile of PSU Median Imputed Expenditure					
	1	2	3	4	5	Total
1	94.8	4.38	0.82	0	0	100
2	4.65	91.03	3.83	0.48	0	100
3	0.55	3.97	91.79	3.63	0.07	100
4	0	0.55	3.22	92.74	3.49	100
5	0	0.14	0.27	3.15	96.44	100

This similarity of all of these results in this section reassures us of two things. Firstly it does not matter greatly which method we use to impute missing expenditure values. This allows us to move ahead with using the regression technique in order to be consistent with the imputation of incomes. Secondly, whichever of techniques we use is not arbitrary in such a way that would be misleading. In other words, we can be fairly confident of the robustness of the imputation results.

3. Imputations of Rent and Implied Rent

3.1 Layout of the data and types of dwellers

Ideally, calculating the rental expenditure would simply involve using the amount of rent paid by each household. However, most houses do not pay rent, either because they own the dwelling or because they receive housing for free and as such do not have a true rental figure that can be used. This is particularly problematic in this survey, where over 73% of households own their dwelling, and 44% of households that do not own do not report paying rent either. Luckily there are questions in the survey to ascertain implied rental value; specifically, "what is the value of rent you could collect per month, if you were to rent this dwelling our?" (D9) and "what is the value of monthly rent you would pay if you had to pay to stay here?" (D12). Hereafter these are abbreviated to "*what would you collect?"* and "*what would you pay?"* respectively.

Four types of households are identified in terms of what sort of housing payments they make. Firstly all households are divided into those that do and do not own the dwelling that they live in. Secondly, those that own the dwelling are divided into those that make mortgage payments and those that have either never had a mortgage or have already paid it off. Finally those that do not own the dwelling are divided into those that pay rent and those that, for whatever reason, do not have to pay rent on the dwelling they live in.

The structure of the questionnaire necessitates this categorisation of households, and the skip patterns ensure that each household answered at least one question about the rent that they pay or would have paid. The four categories are thus defined with the appropriate question specified:

Mortgage Paying Owners: question *"what would you collect?"*
Non mortgage-paying Owners: question *"what would you collect?"*
Rent Paying Non-Owners: question: "what is the amount of rent paid for this dwelling per month?"
Non-Paying Non-Owners: question: *"what would you pay?"*

However, it is also important to note that mortgage payers are systematically different from owners who don't have a mortgage. Perhaps this is because a household that required a mortgage is more likely to be expensive, all else held equal. Regardless, mortgage payments are considerably higher than any other reported housing expenditure questions, and mortgage payers report much higher values for the 'what you would receive' question than other owners, as Table 4 below indicates.

Table 4: Ownership Status and Housing Payments

Ownership Status	Sub-type	Freq.	Percent	Mean value	Question used
	Mortgage Payer	568	7.78	4440.6	what would you collect
Owners	No Mortgage	4770	65.3	1311.3	what would you collect
	Don't specify	49	0.67	1860.0	what would you collect
	Pays Rent	1063	14.55	789.2	how much do you pay
Non-Owners	Doesn't Pay	844	11.55	669.5	how much would you pay
	Don't specify	9	0.12	n/a	combination of do/would
Doesn't Specify	n/a	2	0.03	n/a	combination of do/would

The following discussion provides a comprehensive overview of how rental and implied rental figures were derived for each household. It is largely mirrored a similar section in the NIDS 2008 Income Report.

There are two significant problems to overcome in the pursuit of the estimation of implied rental income. The first is the measurement thereof, and the second is the conceptual difficulty involved in the construction of welfare measures. The measurement problem is made up of two parts being (1) the ability of the relevant questions to measure the parameter of interest, and (2) non-response. Part (1) is a consistent problem that cannot be solved, merely mitigated. With regard to non-response we have followed the same rule as used above in the food and non-food imputations: namely, we have not imputed where non-response exceeds 40%, but in this case it appears that we have no choice. Table 5 below shows the non-response for each of the questions related to the measurement of implied rentals.

Table 5: Missing data for the use of implied rental income

	Non-owners		Owners	
Item	Renters	Don't rent	Mortgage	No mortgage
Amount of bond owing (d7)	n/a	n/a	45%	n/a
Monthly bond payment (d8)	n/a	n/a	32%	n/a
Rent could collect (d9)	n/a	n/a	36%	58%
Rent paid (d11)	7%	n/a	n/a	n/a
Rent would pay (d12)	n/a	61%	n/a	n/a
Market values (d13)	72%	76%	34%	60%
Number in category	1064	845	568	4770

3.1.1 Renters

This is the only completely clear cut case. Here we use the rent these people actually pay and this is added to expenditure. This is not an "implied" rental expenditure, there is actually a flow of payment. Measurement in practice does not really present a problem because NIDS asked for the value of monthly rental payments (if the household claimed to be renting) and non-response was a mere 7%. There is also no obvious reason why measurement in this manner should create a significant bias.

The missing values for renters (and, indeed, all the rent-related categories that follow) were imputed using a regression that contained the following regressors:

Log of income, market value of the dwelling, number of rooms in the dwelling, the square of the number of rooms in the dwelling, the geo-type of the dwelling, the maximum level of education attained by a member of the household, race dummy variables, wall and roof type dummies, house and shack dummy variables and province dummy variables.

3.1.2 Don't own, don't rent

Households in this category are either illegal dwellings, or are houses that do not belong to a person from the household (i.e. a person who does not live there, or a firm). The use of the dwelling at no cost constitutes income for these people. The income amount would be the amount that these people would have to pay in an arm's length rental agreement with the owner, which would be the market rental rate. Clearly this also constitutes an expense since they are making use of the property instead of renting it out. Therefore, they are receiving the benefits of the income stream through use. The same figure should thus be added to both income and expenditure. If we did not add the figure to income we would be underestimating what they are receiving, by not counting the monetary value of the free housing. If we do not add the same amount to expenditure then we are ignoring the fact that these people are making use of the housing.

We measure this implied rental income from the question: "how much rent would you pay, if you had to pay to stay here?" The intention here was to measure the market price of rental at this dwelling. However it is possible that people may have interpreted this question to be asking their willingness to pay. There is essentially nothing that can be done to mitigate this possible bias. We do have the 'reasonable market value for the property' as given by the occupants, but unfortunately the non-response in the case of the market value question is 76% for this group; rendering this data almost useless. Non-response of 61%, in the case of the willingness to pay question, is poor too, but somewhat better. The pragmatic thing to do is to impute for the missing 61% off the data we have and simply accept that this method creates some bias. The regressors used in this imputation are the same as the ones used to impute the missing values for the previous section. We do not have any other useful source of data for this group.

3.1.3 Owners

We separate homeowners into two categories; those with mortgage bonds and those without. While both groups receive the same treatment in terms of the flows of implied income from their housing (see below), they are dealt with separately because the mortgage component is an additional complication.

3.1.4 Owners with mortgage bonds still outstanding

This is a tricky case because there are essentially two linked transactions here that need to be separated out. These are a loan transaction and a purchase transaction. The expense related to the loan is the interest portion of the mortgage repayments. The principal portion of these payments is not an expense but rather an investment (i.e. savings). The purchase of the dwelling is the only part of the transaction that falls into the category of housing. Why should we include the cost of a mortgage in the cost of housing when we do not include the cost of any other means of raising finance to purchase a house?

When a person first purchases the house he then has the right to the economic benefits that flow from ownership: i.e. housing. He may choose to receive the income in the form of rentals, or he may choose to live in it. If he does choose to live in it, then similarly to the case where a dwelling is owned there is an implied rental income as he is receiving housing from his house. In this case he also has an equal expenditure which is the value of consuming the housing, and which is clearly equal to the implied rental income.

Therefore, putting the two transactions together[7]:

---

[7] Note that this means that the difference in expenditure between the homeowner that is fully paid off and the homeowner with the mortgage bond will tend to zero as the mortgage bond tends towards termination. This is because the interest expense will fall as the time to termination decreases.

Income:

Rent* (The value of housing available for consumption)

Expenditure:

Rent* (The value of consumed housing)
Interest expense (The cost of the use of the bank's financial capital)

Certainly one can measure the rent* portion. This can be measured using the question that asks how much rent the respondent thinks he/she could collect if they were to rent it out.[8] While this question is unlikely to provide an unbiased estimate of the true parameter, it is certainly the best way that this information can be obtained, short of making use of other sources of information (e.g. property agents). The non-response of 36% for this question falls beneath our usual threshold of 40% for imputations.

In contrast, measurement of interest is complicated because while the instalments on a mortgage bond are generally equal amounts, the portion of this that is interest (as opposed to principal) will decline over time (and not in a linear manner either!). Thus having no information about where a person is on their repayment timeline and not knowing their interest rate, we cannot guess how much interest they are paying each month on their home loan. In the first month of repayment the interest payment may well be over half the amount of the instalment, but by the final month it will be exactly one month's interest on the final instalment.

However, since NIDS did not collect information on interest expenses (in an effort to avoid questions that may reduce social capital with respondents), the expenditure data does not include any other interest expenses *at all*. It would thus be inconsistent to try and estimate interest expenses from mortgage loans and not from any other type of loan. For this reason, we add rent* to both income and expenditure and acknowledge that there is a general bias in the expenditure data due to the exclusion of interest expenses in general. Households with any debt on which interest is charged will thus have understated expenditures in the NIDS estimations and households with mortgage bonds are just a sub sample of this group.

[8] Note that this is measuring rent* in a different manner to our measurement by those who don't own and don't rent.

3.1.5 Owners of fully paid off dwellings

These people live in a dwelling that is fully owned (no mortgage) by one of the household members. On the income side, there is clearly an income associated with owning the asset, i.e. the benefits of living in the house. These are valued as the market rental price. Since the household is living in this house, they thus use these benefits themselves and so have an expenditure on housing equal to the income. Clearly we thus add the rent* to both the income and expenditure of the household. Unfortunately we must add the caveat here that there is a measurement problem in terms of non-response as 58% of the respondents in this group did not answer the question on the amount of rent they could collect if they rented the dwelling out.

3.1.6 Summary

Table 6 summarises the additions to income and expenditure in the form of rent and implied rent (rent*). The last column summarises the construction of rent/rent*. It does not include interest expenses as these are not housing expenses.

Table 6: Summary of rentals and implied rentals

Category	Expenditure	Income	Notes
Renters	Rent paid	n/a	actual amount paid (D11)
Don't rent / don't own	Rent*	Rent*	amount would pay (D12)
Owners - mortgage	Rent*	Rent*	amount could collect (D9)
Owners - no mortgage	Rent*	Rent*	amount could collect (D9)

It is clear from the table that home owners, with or without mortgage, are treated in the same manner. Those with mortgages would, in a perfect world, also have expenditure equal to the interest payments included under interest expenses. However, as we do not have this data (or any other interest expense data) in NIDS, it is not included and we must simply keep in mind that those with mortgages (or other interest bearing debt) have understated expenses.

Response rates, as previously acknowledged are poor for, in particular, those who don't rent and don't own, and those who own and have no mortgage. In these two cases we are forced to depart from our usual rule of not imputing where missing data exceeds 40% of the total observations. Given the size of the values involved, not imputing would very seriously understate monthly income. It is acknowledged that imputation from such a small base will introduce bias in the data.

However, we maintain that the gains from not substantially understating income outweigh the bias introduced.

4. Comparison of Final Imputed Total Expenditure Figures with the One-shot Expenditure Question

The questionnaire contains a one shot question on total household expenditure for the entire household that asks “How much money did this household spend on all of its expenses in the last 30 days?”[9]. This question has a rather low response rate, which makes it hard to use our expenditure variable, and tends to a rather ill-informed report of expenditure; for instance 1798 households report their total expenditure to be exactly the same as their reported expenditure on foods, despite reporting expenditure on non food items. Summary statistics show this one shot question to have a considerably lower mean and median than the aggregated value of food and non-food expenditures. The following quintile matrix shows how much the imputed expenditure differs from the one shot response. Only the top quintile shows some semblance on consistency between the two measures.

Table 7: Percentage of households in respective expenditure quintiles

	Quintiles of Imputed Total Expenditure					
Quintiles of One-shot Expenditure	1	2	3	4	5	Total
1	19.56	22.13	21.67	19.99	16.66	100
2	46.73	25.53	16.37	8.71	2.66	100
3	21.37	33.97	25.09	15.80	3.77	100
4	6.46	16.90	29.52	33.84	13.28	100
5	0.38	2.15	7.59	22.25	67.63	100

We end the derived variable sections of this report by providing an indication as to how much impact the addition of implied rent and imputed values have made on the expenditure data. Table 8 below reports some descriptive statistics for the one shot income variable, total expenditures without imputations, total food and non-food expenditures with full imputations, total expenditures (food, non-food and rent/implied rent) with full imputations. As expected, both the mean and the median increase as we move from the original raw data to the final imputed expenditure figures.

[9] Question d31 of the Household Questionnaire.

Table 8: Summary of Aggregate Expenditure Measures

Variable	Obs.	Mean	Std. Dev.	Median
One Shot Reported Expenditure*	5802	2124	6157	800
Total Expenditure (Non-imputed)**	7215	2763	5857	1109
Total Food and Non-Food Expenditure (Regression Imputed)***	7305	3041	6087	1254
Total Expenditure (Regression Imputed)****	7305	3869	7170	1622

* Variable w1_h_expnd in the household data file

** Variable w1_h_exprough in the household derived variable data file

*** Sum of variables w1_h_expf + w1_h_expnf in the household derived variable data file

**** Variable w1_h_expenditure in the household derived variable data file

An important health warning is appropriate at this point. It should be noted that the figures presented in Table 8 above and in all previous tables should not be used as the basis for any expenditure analysis because survey weights have not been used their calculation. Their purpose was to describe the process through which we derived a set of expenditure variables for use in analysis. In the next section of the report we go on to present some analysis of expenditures and, therefore, for the first time in the report make use of post-stratified weights in deriving results.

5. Baseline Analysis of Household Expenditure

There are three aims to this section of the report. Section 5.1 aims to present some rudimentary descriptive statistics of the key expenditure aggregates. First we look at expenditure shares of total expenditure before looking at breakdowns by race, geographical type and gender of the household head. Section 5.2 presents an exploratory analysis of food shares in total expenditure as an important example of one of the technical challenges in using expenditure data to make welfare comparisons. Finally, section 5.3 presents a comparison of our expenditure imputations to the imputed income variable.[10]

5.1 Consumption shares

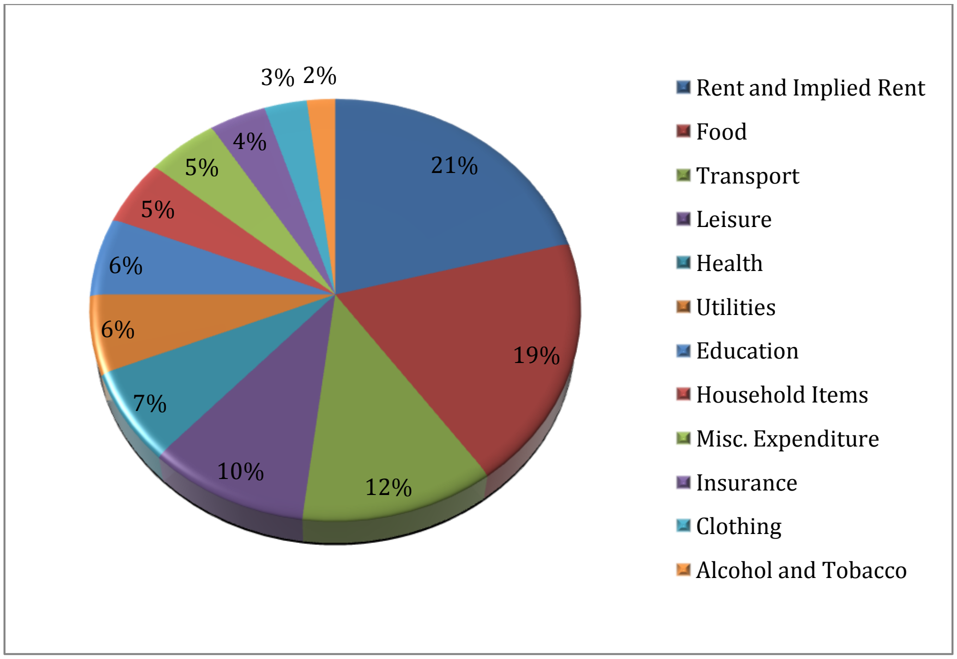
The aggregate imputations for food, non-food and rental expenditure were calculated by imputing all sub-categories of expenditure and then adding these up. For the most part, our analysis of household expenditures is conducted at a fairly aggregated level, using either total household expenditure or the food and non-food sub-aggregates. But before we begin this analysis, it is useful to present a breakdown of total expenditure at a more disaggregated level, especially given the large variety of non-food expenditures in the NIDS data. Figure 2 illustrates this breakdown.

We can see that rental and implied rental expenditure comprise 21%. This is a large share. It illustrates the importance of the work that went into deriving the rent and implied rent variable and the influence of this component of expenditures when total expenditure is used as an indicator of material well-being. Food expenditures accounts for 19% of total expenditures. This is an average share and given the importance of food expenditure, we devote a lot more attention to food expenditures in section 5.2 below. The non-food items make up the remaining 60% of expenditure. It is notable that health, utilities and education collectively constitute 19% of total expenditure; which is the same as food.

---

[10] With the exception of the bivariate densities presented in section 5.2 and 5.3 (i.e, figures 8 and 9) all other reported results are weighted by the post-stratification weights (**w1_wgt**) included in the pubic release of the data.

Figure 1: Decomposition of Total Expenditure into Component Shares



5.2 Race

We begin with a first cut of the derived expenditure aggregates. Table 9 shows descriptive statistics by race for total household expenditure without any imputations (***w1_h_exprough***) and with full imputations (***w1_h_expenditure***) using the methods outlined in sections 2-3. We include the variable ***w1_h_exprough*** in table 9 only for completeness. All results presented here are based on monthly household consumption expenditure with full imputations for item non-responses as outlined in sections 2-3. Thus the key variable in the public release of NIDS WAVE 1 is ***w1_h_expenditure*,** and sub-categories of this variable (such as the share of expenditure devoted to food consumption).

As we discuss below, a key choice in making welfare comparisons across households concerns necessary adjustments that have to be made to account for differences in household composition and size. A key variable in this regard is the share of the total outlay of the household that is devoted to the consumption of food and non-food items. A striking, though somewhat unsurprising, feature of the data is the strong racial dimension of consumption, irrespective of whether one uses the imputed or non-imputed expenditure aggregates. Household consumption expenditure of White and Indian/Asian households (hereafter referred to as "Indian") is about six times as large as that of African households and three times as large as that of Coloured households. Pairwise tests of the equality of means (not reported here) show that these differences are statistically significant for all but one of the six pairwise comparisons, the exception being the difference between the Indian and White distributions.

Some interesting patterns also emerge when testing for equality of means in food shares. Given the similarity in average total consumption between Indian and White households one would expect a similar pattern to hold regarding the difference in the budget share of food. This turns out not to be the case. Similarly African-Coloured food consumption is found to be statistically indistinguishable, yet mean consumption in Coloured households is more than twice that of African households. Thus if total expenditure measures welfare, there appears to be no differences, yet the food consumption patterns of these households appear quite different.

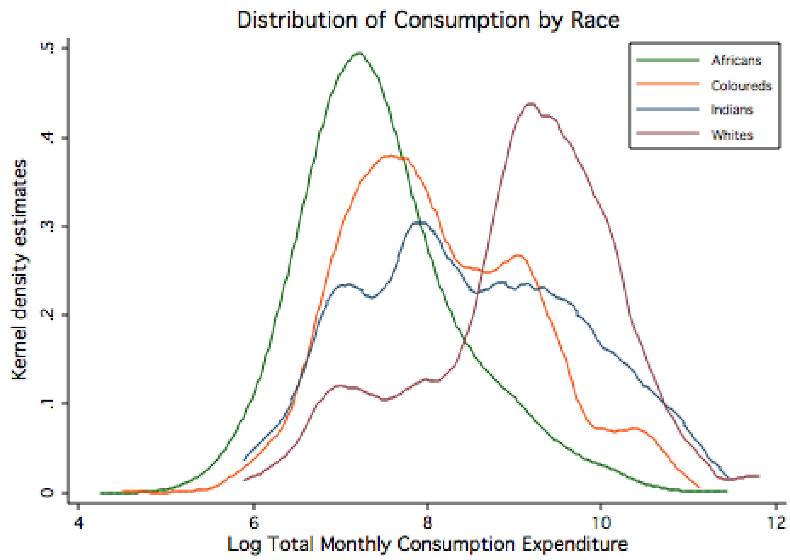
Clearly, the within-race distribution of these consumption aggregates and shares matters, and these distributions in turn reflect other key differences among population groups that might be strongly correlated with consumption patterns. To shed further light on these issues, we look more closely at the empirical distributions underlying these basic welfare rankings by race. Figure 3 shows kernel density estimates of the log of total monthly household expenditure for each of the four race groups.[11]

[11] We fix the bandwidth to 0.5 in all univariate densities reported in this section, and we use the Epanachnikov kernel class in deriving our estimates. Part of our interest here is to investigate the differences in these distributions so it helps not to vary the bandwidth between densities, as would be required if we used any optimizing routines for computing this parameter.

Table 9: Main Expenditure Aggregates by Race

Variable	Mean	Std Error	t	p
Total Expenditure (No Imputations)				
African	2131.53	113.97	18.70	0.00
Coloured	4623.37	891.43	5.19	0.00
Indian	12020.86	3152.62	3.81	0.00
White	12445.34	2145.56	5.80	0.00
Total Expenditure (Full Imputations)				
African	2824.27	146.54	19.27	0.00
Coloured	6370.52	1036.45	6.15	0.00
Indian	16325.08	4194.78	3.89	0.00
White	17392.63	2329.08	7.47	0.00
Food Expenditure				
African	710.91	19.51	36.45	0.00
Coloured	1483.85	127.76	11.61	0.00
Indian	2435.25	249.99	9.74	0.00
White	2024.84	96.38	21.01	0.00
Non-food Expenditure				
African	1577.12	104.81	15.05	0.00
Coloured	3613.47	781.17	4.63	0.00
Indian	10222.80	3184.47	3.21	0.00
White	11353.55	2109.59	5.38	0.00
Food Share				
African	0.38	0.01	51.16	0.00
Coloured	0.35	0.02	20.05	0.00
Indian	0.23	0.03	7.17	0.00
White	0.17	0.01	14.49	0.00
Non-Food Share				
African	0.40	0.01	47.46	0.00
Coloured	0.42	0.02	26.26	0.00
Indian	0.50	0.03	15.47	0.00
White	0.53	0.02	30.05	0.00

Figure 3: Kernel Densities of Households Expenditure by Race



The first thing to note is that these distributions are not merely rightward shifts of each other, as the weighted means of table 9 would suggest. Indeed, a great deal of variance in household consumption is evident within race groups. This is most strikingly illustrated by looking at the Indian and Coloured distributions, which show considerable differences, with the Indian distribution exhibiting a clearly bimodal shape. Moreover, in all cases, the tails of the distributions appear quite pronounced

5.3 Geographical Type

Table 10 presents the same set of descriptive statistics as table 9, but here the conditioning is on geographical type. Some interesting patterns emerge. The data show strong differences between rural and urban areas, with the starkest differences arising between areas formally under the jurisdiction of tribal authorities and formal urban areas which would include major city metropolitan areas as well as sub-urban areas. Strikingly, the mean consumption of households in urban informal areas is not statistically different to that of households in tribal areas or rural formal areas. Interestingly, we note that there does appear to be a statistically significant difference in total expenditure for tribal and rural areas. Yet this is not true for the food share.

Table 10: Main Expenditure Aggregates by Geographical Type

Variable	Mean	Std Error	t	p
Total Expenditure (No Imputations)				
Rural Formal	1908.7	377.1	5.06	0.00
Tribal	1445.8	87.5	16.5	0.00
Urban Formal	5690.3	736.3	7.7	0.00
Urban Informal	1623.9	160.2	10.1	0.00
Total Expenditure (Full Imputations)				
Rural Formal	2705.6	448.2	6.0	0.00
Tribal	1915.2	104.1	18.4	0.00
Urban Formal	7798.7	923.8	8.4	0.00
Urban Informal	2194.4	239.7	9.2	0.00
Food Expenditure				
Rural Formal	817.3	126.0	6.5	0.00
Tribal	632.8	17.9	35.4	0.00
Urban Formal	1233.9	70.7	17.5	0.00
Urban Informal	646.3	53.8	12.0	0.00
Non-food Expenditure				
Rural Formal	1319.5	265.0	5.0	0.00
Tribal	899.5	81.9	11.0	0.00
Urban Formal	4896.2	703.8	7.0	0.00
Urban Informal	1165.6	151.9	7.7	0.00
Food Share				
Rural Formal	0.45	0.02	24.28	0.00
Tribal	0.45	0.01	54.95	0.00
Urban Formal	0.28	0.01	27.86	0.00
Urban Informal	0.38	0.02	16.80	0.00
Non-Food Share				
Rural Formal	0.33	0.02	17.20	0.00
Tribal	0.33	0.01	34.06	0.00
Urban Formal	0.47	0.01	52.36	0.00
Urban Informal	0.43	0.02	17.78	0.00

Figure 4 shows the corresponding empirical distributions for the log of total monthly household expenditure. Again, merely looking at mean consumption hides some stark differences in these distributions. The statistically similar mean incomes of households in rural-formal areas compared to households in urban-informal areas is not entirely borne out by their associated density functions. Indeed, median incomes in urban-informal areas are statistically larger than in both rural-formal and tribal areas and this is reflected in the patterns shown by the density plots, with a clear rightward shift in mass as one moves from rural-formal to tribal to urban-informal areas. Most strikingly, urban-formal consumption differs markedly from consumption in other areas. Indeed, the distribution shows a very sharp pattern of bimodality which is clear evidence of the type of dualism that has characterized major city centres in recent history. While the reasons for such a sharp bifurcation in outcomes are far too complex to address within the scope of this report, part of the explanation undoubtedly has to do with in-migration of poorer (mainly African individuals) into major cities.

Figure 4: Kernel Densities of Household Expenditure by Geographical type

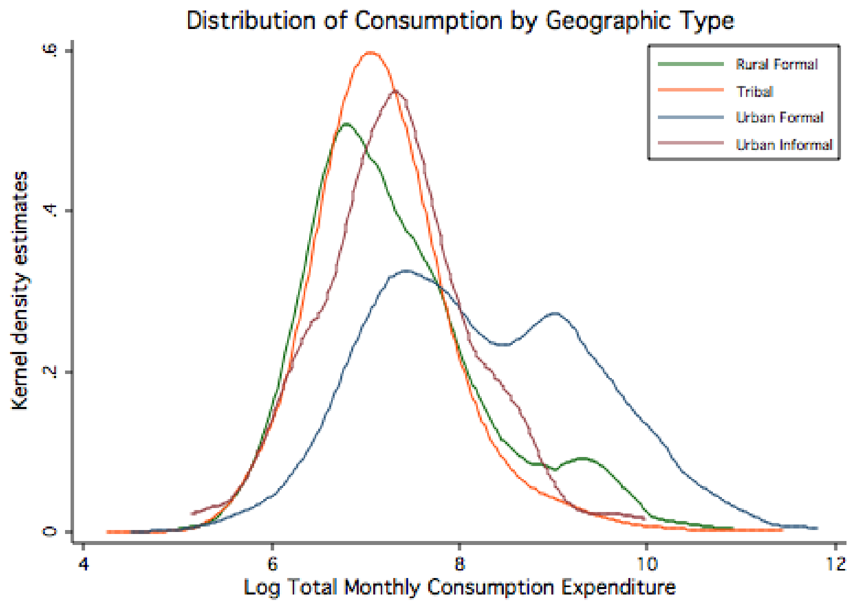


Table 11: Main Expenditure Aggregates by Gender of the Household Head

Variable	Mean	Std Error	t	p
Total Expenditure (No Imputations)				
Male	3895.37	388.37	10.03	0.00
Female	3727.15	601.75	6.19	0.00
Total Expenditure (No Imputations)				
Male	5319.01	517.99	10.27	0.00
Female	5058.91	713.93	7.09	0.00
Food Expenditure				
Male	1004.76	51.75	19.42	0.00
Female	933.88	47.68	19.58	0.00
Non-food Expenditure				
Male	3157.92	352.92	8.95	0.00
Female	3112.63	592.87	5.25	0.00
Food Share				
Male	0.34	0.01	41.36	0.00
Female	0.37	0.01	37.49	0.00
Non-food Share				
Male	0.43	0.01	56.91	0.00
Female	0.40	0.01	41.24	0.00

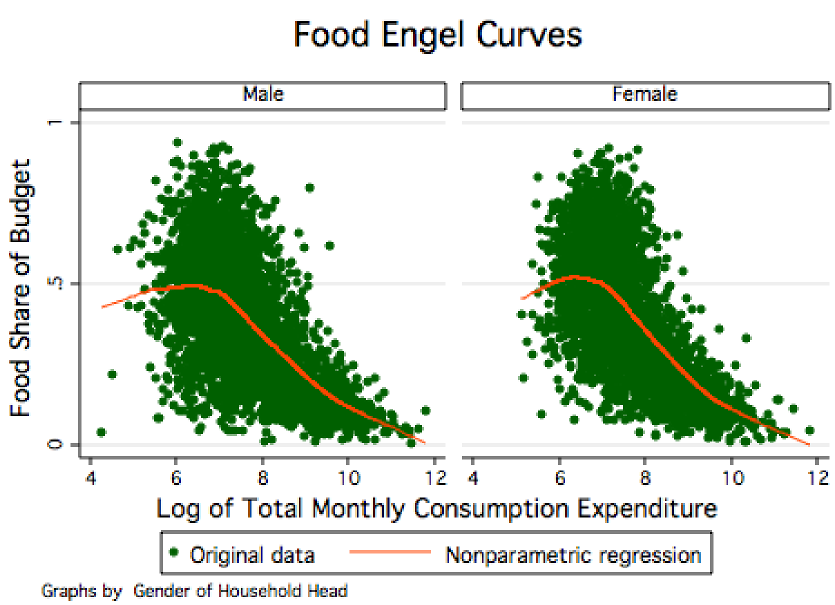
5.4 Gender

Table 11 presents descriptive statistics of the main consumption aggregates by gender of the household head. There are no statistically significant differences in overall consumption expenditure between female-headed households and male-headed households, irrespective of whether we make use of the fully imputed data or the non-imputed data. However, the expenditure devoted to the consumption of food and it's share of the total household outlay do statistically differ by gender of the household head. Female-headed households devote 3% more of total resources to food consumption than their male-headed counterparts even though their mean monthly food expenditure is lower. In order to shed greater light on the relationship between gender, race and food consumption, we need to move beyond univariate statistics. The next sub-section takes up this issue at greater length.

6. Making Welfare Comparisons using Expenditure Data

One of the key policy uses of the NIDS data will be to track household welfare. The previous section of the report began this task by looking at some initial estimates of means and univariate densities of total household consumption expenditures and highlighting important categorical correlates. In addition to these sorts of pairwise comparisons, one usually wants to know if household welfare is sufficient by some or other normative standard. Thus, it is standard practice to compare a variable like total household expenditure to a poverty line. However, total expenditure by itself does not account for the fact that two households with the same level of consumption might actually enjoy vastly differing levels of welfare, both in a relative sense as well as by some absolute criterion because households differ in size and demographic composition.

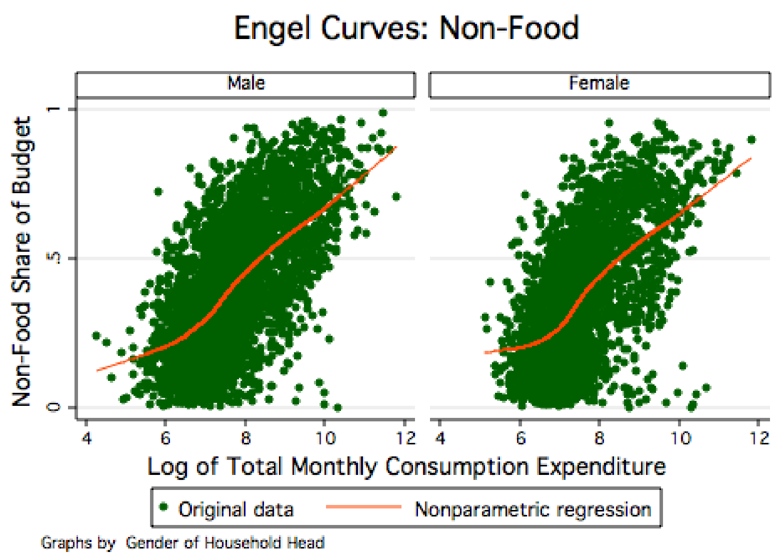
Figure 5: Food Engel Curves by Gender of Household Head



Engel curve analysis – how budget allocations to various categories of consumption changes as the level of living changes – is an integral part of making such welfare comparisons. Figure 5 presents the Engel curve for food consumption by the gender of the household head.[12] Since food consumption is the most basic input into household well-being, a good way of calibrating welfare comparisons between groups, so to speak, is to look at the Engel curve for food. The basic hypothesis here is that we would expect to see a negative relationship (often assumed to be log-linear) – i.e., we would expect poorer households to devote a greater share of their budgets to food consumption. Figure 5 shows this to be the case quite clearly.

Likewise, figure 6 presents an Engel curve for non-food consumption. Here the expectation is that a positive relationship exists between non-food consumption and (the log of) total expenditure.

Figure 6: Non-Food Engel Curves by Gender of Household Head (Full Imputations)



In the case of non-food expenditure, the estimated relationships do appear to be somewhat conditioned on the gender of the household head with noticeable differences in the slope, particularly for poorer female-headed households.

Figure 7 presents Engel curves by race. Here several striking differences are apparent. Indian and White households both appear to devote a smaller share of household expenditure to food consumption at virtually all levels of expenditure. Indeed the shape of their Engel curves for food is not markedly different. Yet, striking differences emerge when we plot the Engel curves by the log of

[12] The fitted curves are based on locally-weighted regressions of the food share on the log of total expenditure.

per capita total expenditure (not shown here), again suggesting that these categorical comparisons of welfare are potentially quite sensitive to differences in household size and composition between race groups.

Figure 7: Food Engel Curves by Race

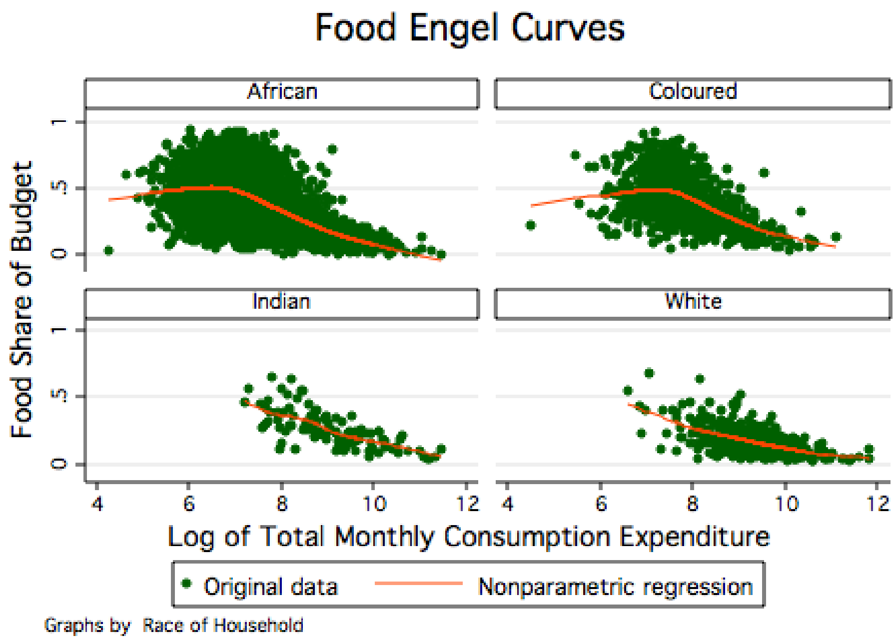
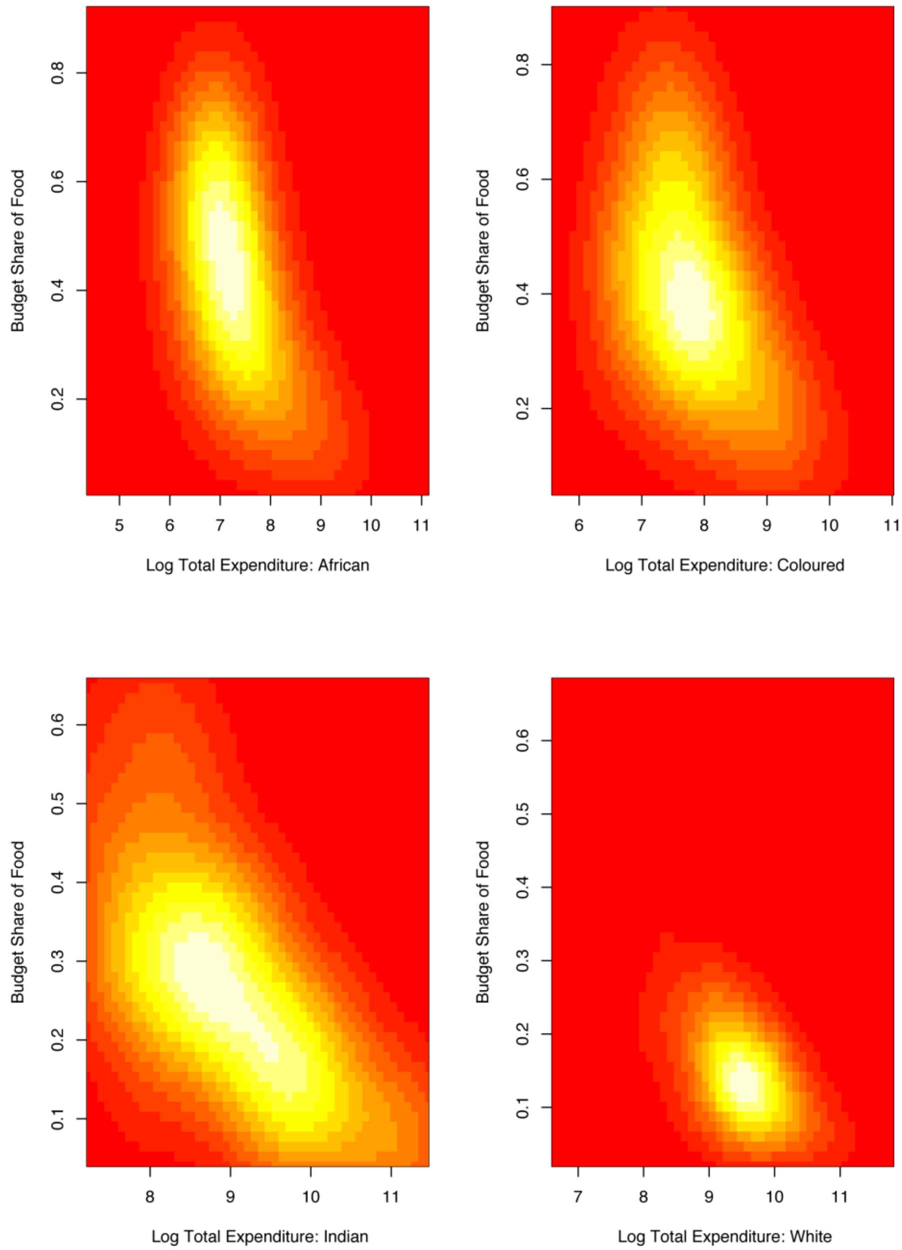


Figure 7 shows significant variation in the budget share across race for any given level of expenditure, as well as substantial variation in household expenditure for any given level of budget share for African and Coloured households. We explore this further by looking at the bivarite densities of the food share and the log of total expenditure. Figure 8 presents a contour plot of the fitted surface. The lighter shaded regions in each panel indicates where the highest concentrations in mass are to be found on the plot. As is clearly shown in the figure, the share of food in the budget is lower for better off households. However, there is considerable dispersion at all levels of living, especially for African (top left) and Coloured (top right) households.

Figure 8: Bivariate Densities of Household Expenditure and Food Share by Race



7. A Comparison of Imputed Expenditures and Incomes

As part of the public release of the data, an imputed household income aggregate is provided in the derived variables folder. It was derived in a very similar way to the expenditure aggregate.[13] In this final section of the report, we compare the imputed income and expenditure variables. All of the categorical analyses of the preceding sections are based on the expenditure aggregates with full imputations. Since income is an alternative metric of welfare, it is important to examine the extent to which the two measures mirror one another.

As a first take at looking at this issue we present a matrix (Table 12) which measures per capita income quintiles along the rows and per capita expenditure quintiles along the columns. The diagonal in Table 12 shows the percentage of household members that are found in the same quintile whether well-being is measured using per capita income or per capita expenditure. Percentages off the main-diagonal then show the extent to which an expenditure view and an income view place household members in different quintiles. The two variables are closely related, particularly for those who are found in the top expenditure quintile. However, there are still a number of people that fall into different expenditure quintiles relative to their income quintile.

Table 12 Percentage of Households in Respective Income and Expenditure Quintiles

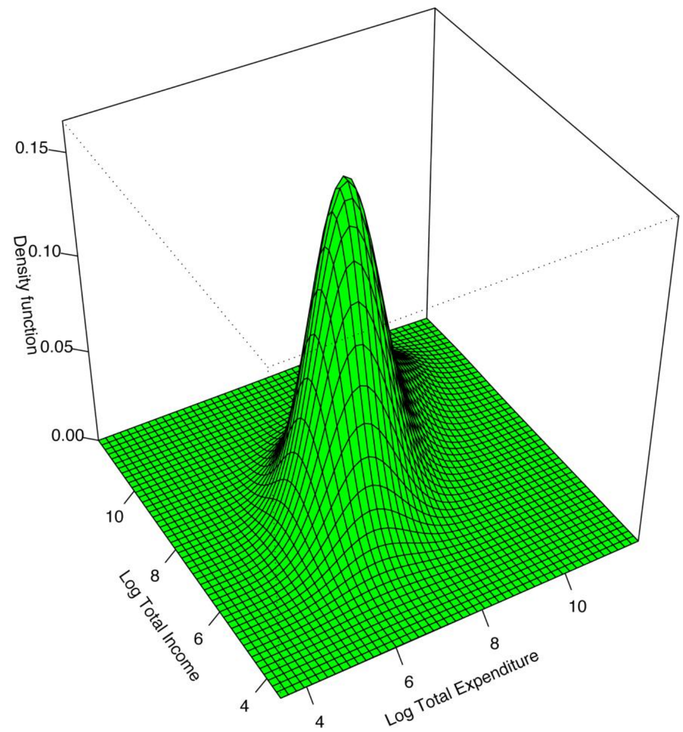
Quintiles of Household Per Capita Income	Quintiles of Household Per Capita Expenditure					
	1	2	3	4	5	Total
1	52.7	27.6	12.5	6.0	1.3	100
2	30.3	35.9	25.1	8.2	0.6	100
3	14.2	26.5	34.4	22.5	2.4	100
4	2.8	9.4	25.6	46.0	16.2	100
5	0.0	0.6	2.5	17.3	79.6	100

Since the concentrations along the main diagonal are somewhat of an artifact of how the matrix is constructed, the matrix does not give an indication of the depth of deviations off the main diagonal, thus obscuring how differences in the underlying distributions of these two variables affects the

[13] See Argent, J. (2009). Income: Report on NIDS Wave 1. *NIDS Technical Paper No. 3*.July. http://www.nids.uct.ac.za/home/index.php/Nids-Documentation/papers.html

classification exercise of table 12. To shed light on this issue, figure 9 shows the fitted surface of the bivariate distributions for the log of total household income and expenditure with full imputations. Somewhat reassuringly, the figure shows that the centre of mass of this surface does not lie too far off the 45-degree line, and shows a pronounced pattern of peakedness. These two factors suggest a much sharper convergence in consistency between the two imputed distributions than that which is indicated by table 12.

Figure 9: Fitted Surface of Bivariate Density of Income and Expenditure



Finally, we note that the log of monthly household income appears higher than the log of monthly household expenditure. However, we caution against interpreting this pattern as evidence of aggregate savings since there are many households in the lower quintiles for whom recorded expenditures are higher than recorded incomes.

Appendices

Appendix A

Rates of non-response and imputation for all expenditure items

	Observations	HH's reporting Consumption	Missing (Item non-response)	Percent Missing	Obs Imputed	Median of original data	Median of imputed observations	Median of Data with imputations
Non-food								
1. Cigarettes and tobacco	1882	2148	266	12%	266	50	54.7	53.2
2. Beer, wine and spirits	1473	1782	309	17%	309	100	72.7	87.6
3. Entertainment	540	607	67	11%	67	350	272.4	305.0
4. Sport	209	244	35	14%	35	150	124.5	150.0
5. Personal Care Items	3601	3950	349	9%	349	60	61.4	60.0
6. Jewellery and watches	119	154	35	23%	35	150	159.0	154.4
7. Newspapers, stationery, etc	1275	1437	162	11%	162	40	34.6	40.0
8. Cell Phone Account	3689	3999	310	8%	310	60	61.4	60.0
9. Telephone Account	742	813	71	9%	71	180	176.1	180.0
10. Lotto, gambling	541	601	60	10%	60	30	31.5	30.0
11. Internet	111	145	34	23%	34	199	122.2	150.0
12. Trips and Holidays	156	186	30	16%	30	475	616.9	500.0
13. Ceremonies	224	253	29	11%	29	200	186.1	200.0
14. Car Payments	413	494	81	16%	81	2000	1491.3	1965.0
15. Petrol, oil, car services	1037	1207	170	14%	170	700	656.4	661.1
16. Buses, taxis train, air tickets	1813	1973	160	8%	160	120	114.1	120.0
17. Water	1979	2286	307	13%	307	100	103.3	100.0
18. Electricity	4454	4817	363	8%	363	100	160.7	100.0
19. Other Energy Sources	2215	2362	147	6%	147	50	49.6	50.0
20. Municipal Rates	877	1231	354	29%	354	180	166.8	175.0
21. Levies	56	88	32	36%	32*	275	275	275
22. Life insurance	950	1149	199	17%	199	200	248.8	218.8
23. Funeral Policies	2738	2992	254	8%	254	80	87.7	80.0
24. Educational Policies	206	253	47	19%	47	220	199.3	213.4
25. Short term insurance	389	491	102	21%	102	500	496.5	500.0

	Observations	HH's reporting Consumption	Missing (Item non-response)	Percent Missing	Obs Imputed	Median of original data	Median of imputed observations	Median of Data with imputations
Non-food (cont)								
26. Kitchen equipment	286	338	52	15%	52	174	154.9	160.0
26. Kitchen equipment	286	338	52	15%	52	174	154.9	160.0
27. Home maintenance	204	235	31	13%	31	400	677.0	450.0
28. Bedding, sheets, blankets	302	335	33	10%	33	200	198.6	200.0
29. Material	66	76	10	13%	10*	295	295.0	295.0
30. Hire purchase on HH items	327	347	20	6%	20	400	366.2	400.0
31. Furniture and appliances	133	155	22	14%	22	500	850.1	600.0
32. Shoes and Clothes cash	845	928	83	9%	83	300	308.7	300.0
33. Shoes and Clothes account	589	648	59	9%	59	320	344.7	337.3
34. Material to make clothing	45	60	15	25%	15*	200	200.0	200.0
35. Medical Aid	752	921	169	18%	169	800	565.3	700.0
36. Dentists, doctors	538	627	89	14%	89	170	225.8	180.0
37. Hospital fees	166	238	72	30%	72*	50	68.4	40.0
38. Medical Supplies	297	353	56	16%	56	100	131.5	100.0
39. Traditional Healers fees	62	67	5	7%	5*	150	150.0	150.0
40. Homeopaths, etc	27	36	9	25%	9*	300	300.0	300.0
41. School fees and tuition	1138	1208	70	6%	70	250	285.5	258.4
42. School books	534	610	76	12%	76	150	136.5	150.0
43. Uniforms	909	989	80	8%	80	300	285.4	300.0
44. Other school expenses	579	658	79	12%	79	100	91.6	100.0
45. Washing and Cleaning goods	5341	5766	425	7%	425	40	43.2	40.0
46. Creche and Childcare	543	561	18	3%	18	100	101.6	100.0
47. Religious dues and donations	1011	1126	115	10%	115	50	90.0	50.0
48. Domestics, Gardeners etc	455	495	40	8%	40	400	368.5	400.0
49. Swimming pool maintenance	77	82	5	6%	5*	200	200.0	200.0
50. Pets	465	517	52	10%	52	103	80.3	100.0
51. Toys	120	136	16	12%	16	80	77.6	80.0
52. Gifts	186	207	21	10%	21	200	137.5	200.0
		53381	5695	11%				

	Observations	HH's reporting Consumption	Missing (Item non-response)	Percent Missing	Obs Imputed	Median of original data	Median of imputed observations	Median of Data with imputations
Food								
1. Mealie Meal	5936	6425	489	8%	489	50	33.4	50.0
2. Samp	2429	2784	355	13%	355	28	18.4	25.0
3. Flour and Bread	5546	6114	568	9%	568	60	56.7	60.0
4. Rice	5573	6116	543	9%	543	40	27.0	40.0
5. Pasta	1705	2009	304	15%	304	20	18.7	20.0
6. Biscuits, Cakes, Rusks	1780	2045	265	13%	265	25	24.2	25.0
7. Red Meat	4122	4456	334	7%	334	100	101.1	100.0
8. Canned red meat	711	849	138	16%	138	28	25.7	27.0
9. Chicken	6103	6519	416	6%	416	60	73.1	64.0
10. Fresh fish and shell fish	1486	1694	208	12%	208	45	47.5	45.7
11. Tinned fish	3081	3450	369	11%	369	20	22.0	20.0
12. Dried peas, lentils, beans	3053	3464	411	12%	411	25	19.6	23.3
13. Potatoes	5916	6375	459	7%	459	30	27.6	30.0
14.Other Vegetables	4755	5219	464	9%	464	26	28.5	28.0
15. Fruits and Nuts	2865	3218	353	11%	353	20	26.5	22.0
16. Oil for Cooking	6263	6794	531	8%	531	32	28.0	30.0
17.Margarine,butter , ghee, other fats	4274	4735	461	10%	461	15	15.6	15.0
18. Peanut butter	2382	2693	311	12%	311	13	14.1	13.7
19. Milk, cheese, yoghurts and dried milk	4286	4702	416	9%	416	30	31.9	30.0
20. Eggs	4859	5256	397	8%	397	30	27.5	30.0
21. Sugar, jam, honey, chocolates and sweets	5617	6211	594	10%	594	35	33.6	35.0
22. Soft drinks and juices	3566	4015	449	11%	449	20	25.5	23.0
23. Tinned fruit and vegetables	951	1167	216	19%	216	25	24.4	25.0
24. Breakfast cereal and porridge	2128	2395	267	11%	267	30	32.1	30.0
25. Baby food and baby formula	746	806	60	7%	60	100	76.4	94.0
26. Salt and spices	5639	6254	615	10%	615	10	11.7	10.0
27. Soya products	1841	2082	241	12%	241	15	15.4	15.0
28. Coffee and tea	5750	6294	544	9%	544	20	24.0	20.0
29. Food Hampers	416	465	49	11%	49	260	207.2	250.0
30.Readymade meals	481	581	100	17%	100	80	74.9	78.6
31. Meals prepared outside the home	828	972	144	15%	144	100	110.6	100.0
32.Other food expenditure	238	301	63	21%	63	70	61.6	64.3
		223222	22524	10%				
*Indicates a replacement to sample median (due to low observation count)								

Appendix B

Comparison of Imputed Food Values for Two different Imputation Methods

Item	Observations	Mean		Sd		Minimum		Maximum	
		Reg	Med	Reg	Med	Reg	Med	Reg	Med
1. Mealie Meal	489	42.984	42.2219	31.9963	32.902	9.5876	7	280	250
2. Samp	355	20.835	22.7578	8.67139	11.3865	9.8076	7	54.14	66
3. Flour and Bread	568	62.612	68.9516	24.7243	33.3417	`	15	309.8	220
4. Rice	543	30.481	33.1716	13.0102	19.6762	12.946	8	120.3	135
5. Pasta	304	20.813	21.5263	7.63859	9.5713	10.999	6	58.56	50
6. Biscuits, Cakes, Rusks	265	28.311	29.1925	13.4543	15.6831	10.586	4.5	92.28	100
7. Red Meat	334	144.1	163.301	103.742	106.858	18.575	20	533.2	700
8. Canned red meat	138	27.282	30.3116	7.2497	20.7734	15.037	8.5	53.17	124
9. Chicken	416	85.679	95.4231	42.0927	46.4216	28.713	22.5	459.9	225
10. Fresh fish and shell fish	208	54.787	56.1154	25.1227	31.3597	16.332	11.5	138	200
11. Tinned fish	369	23.271	22.9431	6.26387	11.1119	12.815	8	43.93	80
12. Dried peas, lentils, beans	411	21.866	23.1192	8.46716	11.7648	11.4	7	74.27	100
13. Potatoes	459	28.542	29.5229	7.71614	8.68807	14.138	7.5	54.18	80
14. Other Vegetables	464	36.696	38.306	22.9397	30.167	11.505	10	139	210
15. Fruits and Nuts	353	33.928	34.4405	22.1013	25.3913	8.3784	8	134.7	150
16. Oil for Cooking	531	30.034	31.2505	9.30196	13.2621	13.04	10.5	108.4	95
17. Margarine, butter, ghee, other fats	461	18.423	19.051	7.99483	8.4826	7.7385	7	55.66	55
18. Peanut butter	311	14.292	14.0241	2.03348	3.93929	9.879	7.5	24.33	30
19. Milk, cheese, yoghurts and dried milk	416	46.53	44.4159	38.0576	40.7748	11.273	7	235.6	275
20. Eggs	397	28.643	29.1839	6.58087	7.57172	16.041	8	61.13	60
21. Sugar, jam, honey, chocolates and sweets	594	35.352	37.3594	11.6799	13.9049	16.878	12	116.5	80
22. Soft drinks and juices	449	32.117	32.1347	18.8647	23.5404	7.2348	8.5	120.3	183
23. Tinned fruit and vegetables	216	26.301	28.7546	8.23669	16.9573	12.515	6.5	57.97	210
24. Breakfast cereal and porridge	267	35.95	35.7116	14.9705	13.8329	12.492	12	96.84	100
25. Baby food and baby formula	60	93.328	99.5083	46.5659	62.2419	33.431	30	233.3	300
26. Salt and spices	615	13.611	14.4358	7.0933	7.30387	4.4792	3	69.9	60
27. Soya products	241	17.376	16.8652	6.87306	6.726	8.8984	4	74.43	67.5
28. Coffee and tea	544	27.664	28.1774	12.5537	12.744	8.9635	8	75.71	85
29. Food Hampers	49	219.31	234.806	69.8582	84.4494	101.23	85	369.9	450
30. Readymade meals	100	81.207	88.235	40.4332	48.3579	23.751	6	206.8	320
31. Meals prepared outside the home	144	134.75	162.024	84.5729	108.543	24.048	25	436.7	675
32. Other food expenditure	63	74.371	79.5238	39.3045	30.988	17.748	10	182.5	200

Comparison of Imputed Non-Food Values for Two different Imputation Methods (cont)

Item	Observations	Mean		Sd		Minimum		Maximum	
		Reg	Med	Reg	Med	Reg	Med	Reg	Med
1. Cigarettes and tobacco	266	75.70	76.42	70.19	78.51	13.5	18.1	600	843
2. Beer, wine and spirits	309	91.36	85.22	48.43	43.05	30	31.27	400	321
3. Entertainment	67	346.81	280.78	111.36	106.67	150	84.99	500	616
4. Sport	35	172.63	160.49	93.11	120.95	65	43.2	525	635
5. Personal Care Items	349	79.41	76.37	38.36	48.66	20	19.73	275	352
6. Jewellery and watches	35	192.97	176.71	103.93	126.18	50	15.59	500	663
7. Newspapers, stationery, etc	162	45.05	41.04	23.88	21.66	8	13.55	150	139
8. Cell Phone Account	310	92.30	93.48	85.40	92.03	12	13.76	600	781
9. Telephone Account	71	205.23	188.45	66.81	90.49	61	53.34	450	564
10. Lotto, gambling	60	32.87	33.41	13.35	14.34	7	13.09	60	113
11. Internet	34	203.43	142.79	101.76	95.90	9	9.662	475	365
12. Trips and Holidays	30	648.67	1157.79	459.07	1564.46	200	182.2	2000	8165
13. Ceremonies	29	433.79	300.67	526.21	273.78	100	63.84	2750	1281
14. Car Payments	81	2057.7	1687.32	811.54	716.88	1200	631.3	5250	4025
15. Petrol, oil, car services	170	672.04	707.56	229.25	292.68	275	278.8	1750	1891
16. Buses, taxis train, air tickets	160	156.20	150.91	92.42	117.35	20	28.81	631	1028
17. Water	307	165.86	137.09	160.61	112.23	20	8.017	1000	768
18. Electricity	363	148.63	243.69	97.15	223.85	20	30.21	650	1279
19. Other Energy Sources	147	51.05	55.09	30.95	21.80	10	27.84	280	145
20. Municipal Rates	354	192.07	198.49	78.62	121.17	30	32.81	1050	762
21. Levies	32	275.00	275.00	0.00	0.00	275	275	275	275
22. Life insurance	199	258.72	295.93	109.00	178.47	74.5	66.5	800	967
23. Funeral Policies	254	81.13	96.74	36.52	34.39	24	41.14	250	254
24. Educational Policies	47	224.22	216.02	54.53	85.23	150	95.89	400	424
25. Short term insurance	102	546.27	535.40	259.85	262.62	254	148.2	2350	1663
26. Kitchen equipment	52	161.59	210.05	72.64	186.96	27.5	43.51	500	1022
27. Home maintenance	31	575.00	925.99	423.70	1061.79	150	101.3	2000	5718
28. Bedding, sheets, blankets	33	203.58	248.85	88.88	151.04	25	74.76	450	745
29. Material	10	295.00	295.00	0.00	0.00	295	295	295	295
30. Hire purchase on HH items	20	371.00	385.08	95.85	98.08	200	259	500	672
31. Furniture and appliances	22	638.39	1220.02	449.03	1198.20	317	170.5	2500	5670
32. Shoes and Clothes cash	83	300.30	324.61	160.83	138.45	60	143.7	1250	822
33. Shoes and Clothes account	59	351.78	389.45	125.62	151.46	175	231.4	950	973
34. Material to make clothing	15	200.00	200.00	0.00	0.00	200	200	200	200

Comparison of Imputed Non-Food Values for Two different Imputation Methods (cont)

Item	Observations	Mean		Sd		Minimum		Maximum	
		Reg	Med	Reg	Med	Reg	Med	Reg	Med
35. Medical Aid	169	813.03	662.73	298.25	406.36	205	146.6	2000	2335
36. Dentists, doctors	89	239.54	260.24	101.45	118.58	91	85.98	400	631
37. Hospital fees	72	84.94	184.28	97.37	256.02	40	14.09	813	1304
38. Medical Supplies	56	122.99	150.38	64.65	102.28	30	31.09	400	534
39. Traditional Healers fees	5	150.00	150.00	0.00	0.00	150	150	150	150
40. Homeopaths, etc	9	300.00	300.00	0.00	0.00	300	300	300	300
41. School fees and tuition	70	391.69	432.19	388.99	412.53	35	47.81	2800	2498
42. School books	76	149.55	177.63	99.06	148.41	40	44.52	850	957
43. Uniforms	80	298.16	290.57	104.02	75.63	125	116.4	700	538
44. Other school expenses	79	122.69	114.56	103.63	60.96	60	37.2	700	306
45. Washing and Cleaning goods	425	49.66	52.31	27.28	30.32	11	16.98	200	260
46. Creche and Childcare	18	125.00	239.02	110.35	315.23	10	45.88	400	1132
47. Religious dues and donations	115	106.07	133.08	113.19	114.23	10	21.52	700	648
48. Domestics, Gardeners etc	40	496.98	453.63	280.52	328.15	40	28.39	1400	1562
49. Swimming pool maintenance	5	200.00	200.00	0.00	0.00	200	200	200	200
50. Pets	52	120.25	92.37	67.05	62.57	40	18.15	350	343
51. Toys	16	79.31	90.99	19.08	50.85	50	18.08	100	218
52. Gifts	21	200.48	186.00	106.96	116.06	70	48.09	475	401