MOZAMBIQUE EDUCATION OUTCOMES NATIONAL PANEL SURVEY (NPS)

1. BACKGROUND

In 2004, the Bank, jointly with other donors and the Government of Mozambique, prepared a Poverty and Social Impact Analysis on the issue of fee reform in primary school. Partly as a result of the study findings, the Government took the step of abolishing tuition fees in primary education. In 2006, Ministry of Education and Culture (MEC) requested a repeat of this analysis, as well as a similar baseline study on barriers to enrollment for the poor in secondary education. In particular the MEC sought World Bank assistance in (a) evaluating the success of the reforms in primary education financing to date, and (b) formulating new policies and initiatives to reduce the barriers the poorest households face in accessing primary and secondary education. This panel survey is part of the Bank's response to this request.

2. DATA COLLECTION

The Education Outcomes Panel Survey (NPS) was designed as a panel survey based on a subsample of households interviewed in the 2002/03 *Inquérito aos Agregados Familiares* (IAF), a national household income and expenditure survey conducted in all provinces of Mozambique from July 2002 to June 2003. The NPS data collection took place from September 2008 to February 2009 and it was performed by a contractor in Mozambique (KPMG), with World Bank and UNICEF field supervision.

The NPS sampling frame consists of enumeration areas (EA) that were drawn to correspond to a particular set of months of the 2002/03 IAF, namely March to May 2003, since it is expected that the IAF has a nationally representative subsample of EAs assigned each quarter. It is important to highlight that the NPS data is nationally representative at the rural and urban areas, but not representative below this level. The main reason is that the IAF sample was clustered to maximize efficiency in the data collection process across a 12 month period, while the NPS sample, due to costs constrains, includes only 3 months. Therefore, the NPS sample does not have enough geographic dispersion to be representative at the province level or below.

All IAF households in the enumeration areas during the months of March-May were included in the NPS sample, resulting in 221 EAs and 2,234 households. This sampling strategy was chosen to reduce the effect of seasonality in the panel analysis when comparing the 2002/03 IAF data to the 2008 NPS data for the same sample households. Originally it was planned to interview all the IAF sample households in these EAs during the same month in which they had been interviewed for the 2002/03 IAF. However, because of delays in the survey planning process, the data collection for the NPS was postponed took place from September 2008 to February 2009.

The survey was designed to target eligible children/student (i.e. children aged 0-17 y.o. in 2003 or members enrolled in school in 2003) from the IAF sample. The households in the NPS sample were divided into 2 categories based on their status in 2003:

A. Target 2003 households. These are households that meet at least one of the following criteria:
    - Households that had at least one child 0-17 years-old in 2003 (see question a13 in the questionnaire)
    - Households that had someone in primary or secondary school in 2003, in spite of age (see question a14 in the questionnaire)
B. Alternate 2003 Households (14% of original NPS sample)

For the households that did not have any children or student in 2003 but were part of the IAF sample and were in the NPS enumeration area, the following two questions were asked to the first person who was found in the alternate household in 2008:

- Does this person's 2008 household currently have anyone who is between 5 and 17 years of age? (see question a15 in the questionnaire)
- Does this person's 2008 household currently have anyone who attending primary or secondary school? (see question a16 in the questionnaire)

If the answer was YES to either question (a15 or a16), the interviewer proceeded with the entire questionnaire. If the answer was NO to both questions, the interviewer stopped the interview.

In sum, target households are the source for the panel of children, while alternate households were included to supplement sample size.

There were two types of tracking in the NPS, that of households and that of children/students who split from the original 2003 household and joined new households in 2008. If the entire 2003 household moved in 2008, the field team would gather their new contact information with local leaders, neighbors, friend, etc and follow and interview the household at their new location, provided the household moved within the district (the survey only followed households/children that moved within the district level). New members of the household were also included in the interview.

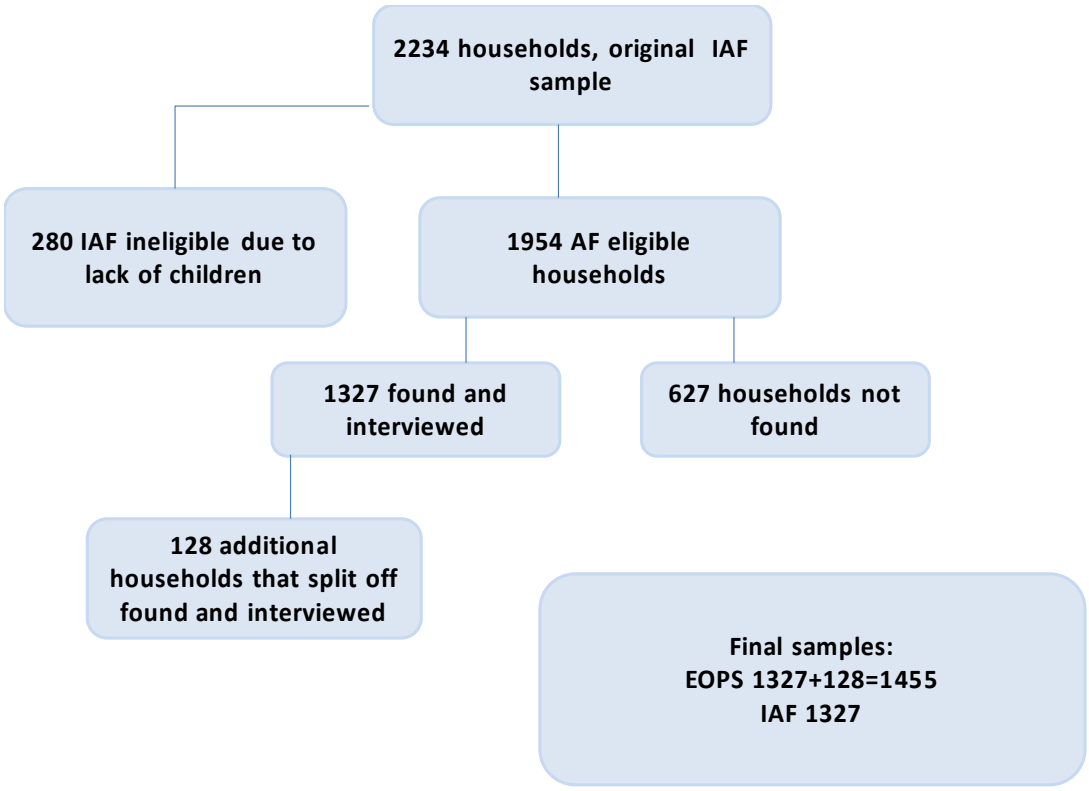
If the 2003 household was split in 2008 and the members who moved out were a target member (children/student in 2003) who had moved within the district, then the team followed the individuals and interviewed both the original household (if a target member still lived there) and the split household.

The screening for tracking are in section B1 of the questionnaire. A member would be tracked if b100a =1 (this variable is an indicator of whether the member was target member, i.e. less than 17 y.o. in 2003 or attending school in 2003), **and** b108=2 (the member no longer lives in the household), **and** b111 <=2 (the member moved to the same village or district). If all these conditions were met, questions B112 (should the member be tracked?) should be 1 (YES) and the household should be followed. The variable "sp" indicates whether the household was the original (sp=0) or a split household (sp>=1).

In case all target members (b100a=1) moved out of the household, the interviewer should end the interview with the original household at question B114.

The chart below illustrates the final outcome of the field work in terms of interview status of the households in the original sample. The final count of households interviewed is 1455 of which 128 are additional households that split between off the original IAF sample.

Figure 1 Final NPS sample



3. DATA DESCRIPTION

The main instrument of NPS is a household questionnaire that consists of the following modules:

- General Household Questionnaire: modules A, B0, B1, B2 (demographics), C0 (education), D0 D1, D2, D3 (employment), E (household characteristics), H (education quality perception), I (transfers)
- Consumption module: modules F, GA, GB, GC, GD
- Education Event History Module*:* module C1
- Education Expenditure Modules*:* module C2

The NPS data can be found in the folder "raw data/NPS 2008" with a stata data file for each section of the questionnaire. Minor cleaning of values out of range, miscoded member numbers and miscodes in the education variables was done by the World Bank after receiving it from KPMG.

In addition to NPS the package also include the original IAF data that used as original sample frame (2234 observations) with recoded identification codes so they match NPS. See below.

3.1 IDENTIFICATION VARIABLES AND MERGE WITH IAF

Households in NPS identified by enumeration area (ea2) and household number (af). These two variables is also the link back to the IAF survey. Both the IAF and NPS datasets included in this package have these two variables to facilitate and easy merge between the two survey years. NPS identify the households that split by the variable sp, while the individual ID numbers in are found in the variable "pind". Table 1 shows the original coding used in IAF in case researches have an interest in using the complete IAF sample.

Table 1: Identifying variables in in AIF 2002/03 and NPS 2008

Variable	IAF 2002/03	NPS 2008
Enumeration area	a1	ea2
Household number	a2	af
Split household		sp

3.2 POST SAMPLING WEIGHTS

The variable psia_wt is the basic household weight for NPS. This is the final household weight produced by the team's sampling specialist and should be used to make the sample estimates from the NPS representative of the population. For details regarding the construction of the psia_wt, please refer to the "Sample Design and Estimation Procedures" documentation included as additional documentation with the data.

Section 4 below analysis attrition in the sample with more detail provided in appendix 3. Based on this work a second set of inverse probability weights have also been included in the auxiliary data. These are name “hhweights” for household weights and “indweights”, for individual weights.

3.3 CONSUMPTION AGGREGATE

NPS contains a several sections on food and non food consumed by households. In the stata file auxiliary data this information has been combined into a consumption aggregate that can be used to rank households consumption per capita. Appendix 1 has additional information on how this was done.

Spatial price differences are often important when evaluating consumption levels. To correct for this a second version of the consumption aggregate corrected for spatial price differences is also included in the auxiliary data set. The spatial price index is from the 2008/09 IOF household survey that is richer in price data while implemented in the same year as NPS .

3.1 DATA PACKAGE

The package that comes along with this documentation includes the following data files:

- The 2002/03 IAF data prepared for merge with the NPS 2008 round of interviews. There 2 files – one at individual level and one at household level
- The NPS 2008 round of the survey. This folder contains 27 data files matching the questionnaire.
- The asset and consumption aggregate is combined into one data set. The do-files that created the asset index and consumption aggregate are included in the folder.
- Weights. A total of four different sets of weights exist for these data. They are all included in the Weights folder. The data set hhweights include 3 sets of household weights. One set from the original IAF 2002/03 survey. One from the new set of NPS weights as documented in “Sample Design and Estimation Procedures and finally the inverted probability weights as documented in section 4.3 and appendix 3. The final set of weights are also inverted probability weights, but calculated at individual level. These are included as a separate data file.

4 ATTRITION ANALYSIS

This part presents the attrition patterns in the NPS sample and a few examples of regression designs correcting for the selection bias. It first describes the differences between the observations in the final sample and those originally selected for sampling.

4.1 DESCRIPTION OF NPS ATTRITION

Appendix table 2 compares final attrition rates for households, children and adults between provinces. The attrition rates appear to be very variable between provinces. The interpretation is nevertheless difficult as the survey is not representative by province. Indeed, those differences can be explained things as sampling, differences in quality of field teams[1] and by unobservable differences between provinces. The share of non-respondents is higher in urban areas, both for households and children. The share of females and males resurveyed by NPS is the same for children, but adult women are more likely to be resurveyed than men.

Appendix 2 table 2 and 3 compares the demographic characteristics of respondent and non-respondent households in urban and rural areas respectively. The non-respondent households are smaller in rural areas, which is intuitive as large households may move less often (and may be easier to track). Similarly, the heads are elder in urban areas. They also include more children in rural areas, which could be due to the same reasons. The non-respondent households are wealthier in rural areas, as they are more frequently in the top quintile of the national distribution of consumption per capita, and less frequently in the top quintile. The head of non-respondent households has on average less education in urban areas. Surprisingly, the school enrollment rate is higher in respondent households. The occupation of the head does not appear to be much related with household attrition.

Table 4 and 5 compares the demographic characteristics of respondent and non-respondent individuals in urban and rural areas respectively. Respondent adults work more often in agriculture, especially in rural areas. Respondents are less often students in rural areas, which seem logical as a student aged 18 or more in rural Mozambique may have to leave his village to study. In urban areas, the respondents are more often government workers and less often private wage workers. Children are less likely to be working among respondents than non-respondents in both urban and rural areas. This is credible as elder children may be more likely to work and to leave the household. Among children, students are more likely to be respondents in urban areas. Overall elder people are less likely to be respondents. As for the household attrition, respondent individuals are more likely to be poor than non-respondents in rural areas.

4.2 CORRELATION COEFFICIENTS OF NPS ATTRITION

Tables 1, 2, 3, 4 and 5 in appendix 2 present the differences between the samples of respondents and non-respondents. This is nevertheless not sufficient to understand the links between the observable characteristics of the household and attrition. Indeed, many of the differences between attriters and the NPS sample can be explained by the correlation between the explanatory variables. This section looks at this in table 6 and 7 found in the appendix 2.

Appendix table 6 shows probit models predicting the inclusion of the household in the NPS sample. They shows that the respondents to the NPS survey are more likely to be larger

[1] There were 11 field teams in NPS, 1 by province.

households, older households heads, and have more children than non-respondents This effect is nevertheless non-linear in urban areas, as children with at least one child of school age are less likely to be respondents. The poverty variables do not appear to be a strong determinant of the response in NPS. In urban areas however, the respondents have higher school enrollment rates. In urban areas again, the households whose head reported a disease in 2003 are less likely to be respondents.

Appendix table 7 shows inclusion of a child in NPS sample. The household determinants of children attrition are not the same than those of household attrition. In Table 7, the occupation of the head appears to be unrelated with attrition. The rural areas appear to have more respondents. The coefficient of the other geographic characteristics on response is not significant. The field team quality index is based on the observations of the Bank staff supervising the contractor.[2] Its coefficient is positive and significant, as expected, in all specifications. Column 2 and 3 control for other geographic characteristics: Central and South regions interacted with rural in column 2, dummies for Maputo city and Province in column. The coefficient for field team quality remains positive and significant. Big households are more likely to be respondents; children from big households are more likely to be respondents in urban areas only. The households with old heads are more likely to be respondent for NPS, whereas the children from these households are less likely to be in the NPS sample. Households with a high share of children are more likely to be respondents. The children from these households are more likely to be respondents in urban areas only. Apart from education, the poverty indicators do not predict the inclusion in the NPS sample for children. Education seems to be a deterrent of dropout in this panel for children in urban areas. Indeed, children whose household-level enrollment rate is high are more likely to be included in the NPS sample in urban areas. Similarly, children who were enrolled in 2003 in urban areas were more likely to be respondents. The children aged 10-17 are more likely to be surveyed by NPS whenever the head is a farmer. This is logical as some teenagers might leave the household. In rural areas, provided they work in agriculture, they may not have left their village between 2003 and 2008. Concerning the individual characteristics, elder children are more likely to be non-respondents. This seems logical: teenagers are more likely to leave the household. This is especially true in urban areas. Children who are enrolled at school or too young to be active are more likely to be respondents. Children are more likely to be included in the NPS sample in rural areas, which is not surprising. The attrition varies between regions in Table 7: inclusion in the NPS sample is more likely for children in the Central and Southern regions, and especially so in rural areas.

[2]The teams were ranked teams as (i) good, (ii) bad, and (iii) no information.

4.3 Control for selection bias in the NPS survey with Inverse Probability Weighting (IPW)

The literature suggests two approaches to control for the selection bias observed and described above is panel surveys like NPS: (a) a 2-stage approach following Heckman, and (b) and inverse probability approach following Wooldridge (denoted IPW hereafter). Appendix 3 analyses the second approach in details, and compares its results with the first in two regressions on the household and individual sample.

The results appendix 3 are quite optimistic: the NPS sample seems reasonably representative of the Mozambican population concerning child education, and the control for selection hardly changes a few coefficients in urban areas in our regressions predicting child education. In addition, the implementation of the controls for selection is not very costly. The controls would nevertheless need to be adapted for more complicated regression designs. However, this does not prove that regression coefficients would never be biased by the control for the selection bias. In particular, shocks between 2003 and 2008 may have caused both selection and demographic variables in 2008. Researchers may still wish to control for selection in their own analysis.

REFERENCES

Megill, D.J. (2009), Sample Design and Estimation Procedures for Poverty and Social Impact Analysis (PSIA) Follow-up study, World Bank

Wooldridge, J. (2002), Econometric Analysis of Cross Section and Panel Data. XXIII, 752 pp. MIT Press, Cambridge, Mass., 2002

Filmer, D.; Lant H. Pritchett (2001), Estimating Wealth Effects without Expenditure Data-or Tears: An Application to Educational Enrollments in States of India, Demography, Vol. 38, No. 1 (Feb., 2001), pp. 115-132

APPENDIX 1: CONSUMPTION AGGREGATE

Food consumption in the last 7 days:

The goal is to calculate the value of consumed food for all the items consumed in the last 7 days. The problem is that the value of consumed food is not observed in the data. Instead, we observe the following:

- Item i,
- Quantity consumed, Unit consumed,
- Quantity produced, Unit produced, Value produced,
- Quantity gifts, Unit gifts, Value gifts,
- Quantity purchased, Unit purchased, Value purchased, Number of days in which this purchase will be consumed.

The "Value consumed" for item *i,* needs to be imputed according to the formula:

Value consumed = Quantity consumed * Price per unit consumed.

However, the "Price per unit consumed" is never directly observed and must be also imputed.

Total number of values to be imputed: 18,862

NB: For codes 165, 166, 167 (meals consumed outside home), the consumed value is estimated as sum of value of purchased meals and those received as gift.

Imputation of "Price per unit consumed":

1. When possible, household reported unit values were used to impute price of goods consumed in the following order:

- unit values from purchases

- unit values from home production

- unit values from gifts.

Unit values from purchases are most reliable, because they correspond to the actual purchases made by household and actual local prices, whereas unit values from home production and gifts involve a respondent's guess.

2. If a household reports consuming an item, but does not report any unit values, then the estimated local prices are used in the following order:

- median price in the district

- median price in the province

- median price in the country

For each item/unit a median price is calculated over all recorded unit values (i.e. purchased prices, produced prices, gifts prices) in the corresponding area\region.

To improve the reliability of the estimates, median prices for the district are used only if there are 5 or more observations of price of an item in the district. Otherwise, province level prices and then country level prices are used.

3. When prices cannot be estimated, two more imputation methods are used:

First, if household reports the purchased value and the number of days in which it will be consumed, the consumed value is estimated to be equal to:

Value Consumed = Purchased Value, if Number of days <= 7;

Value Consumed = Purchased Value*7/Number of days, if Number of days > 7;

For example, if household reported purchasing 50 MT worth of rice that will be used in 14 days, then the imputed value of consumption will be 25 MT.

For the remaining cases, the value consumed is estimated as the national median per capita value of an item, multiplied by the household size. Table 1 shows the counts of the different imputation rules that were used.

Table 1

Method of imputation		
1	HH unit value from purchase	11,619
2	HH unit value from home production	3,781
3	HH unit value from gifts	468

4	District level prices	2,640
5	Provice level prices	66
6	National level prices	187
7	Adjusted purchased value	20
8	Median consumption value	49
9	Meals consumed outside	30
Total		18,862

Conversion of the units:

All the quantities were recorded in one of the 31 units. It is important to convert different types of units into few standard ones, for two reasons:

1) To be able to apply household unit values when unit values and quantity consumed are recorded in different units. For example, a household may report consumption in kilograms, but purchase in "sacks of 100 kg". Therefore, for this household, we do not formally observe the price per 1 kg, but only the price per "sack of 100kg". To apply the purchase unit value, the quantity of "sack of 100kg" must be converted to kg.

2) To have a larger number of observations of price per unit, when median regional values are calculated.

The 31 available units were converted into 3 standard ones (kg, liters and "units") and 5 non-standard ones. Table 2 summarizes the conversion adjustments that were made.

Table 2

Code	Unit	Number of observations (units for quantity consumed)	Standard unit	Multiplier to convert unit to Std. unit
1	kg	5,648	KG	1
2	Unit	3,978	UNIT	1
11	Bag of 100kg	2	KG	100

12	Bag of 90kg	3	KG	90
13	Bag of 70kg	0	KG	70
14	Bag of 60kg	2	KG	60
15	Bag of 50kg	25	KG	50
16	Bag of 25kg	37	KG	25
21	Can of 25lts	63	LITER	25
22	Can of 20lts	225	LITER	20
23	Can of 10lts	99	LITER	10
24	Can of 5lts	148	LITER	5
25	Can of 1lt	141	LITER	1
31	large bunch	28	UNIT	10
32	average bunch	33	UNIT	5
33	small bunch	25	UNIT	3
34	Canteiro	16	CANTEIRHO	1
35	molho (bundle)	2,219	MOLHO	1
36	Montinho (heap, mount of spices, etc)	2,117	MONTINHO	1
37	Box	30	BOX	1
40	300 ml (bottle for refreshments)	111	LITER	0.3
41	500ml	179	LITER	0.5
42	750 ml	38	LITER	0.75
43	1 Liter	795	LITER	1

45	Gallon, 5 Liters	29	LITER	5
50	540 ml (medium beer)	65	LITER	0.54
51	Bottle (small beer)	47	LITER	0.2
52	A cup, if item is not "sal grosso"	1,267	LITER	0.25
52	A cup, if item is "sal grosso"	771	KG	0.180
53	Bottle cap	145	LITER	0.005
54	Small spoon	173	LITER	0.005
55	Grams	313	KG	1000
56	Other ____	90	OTHER	1
	Total	18,862		

Table 3 shows the distribution of Standardized units for four variables.

Table 3

	Standard units			
	Cons	Prod	Gift	Purch
Kg	6,801	1,275	216	4,676
Units	4,064	786	135	2,686
Liters	3,525	662	111	2,108
Canteirho	16	14	0	1
Molho	2,219	647	71	1,379
Montinho	2,117	685	75	1,330
Box	30	0	1	30

Other	90	11	1	80
Total	18,862	4,080	610	12,290

Comments on unit conversion:

1) Non-standard units.

The non-standard units that were left as is: Canteirrho, Molho, Montinho, Box, Other.

2) Exception for salt.

For the unit 52, "a cup", different conversion was used, depending on the code of the reported item. By definition, a cup is a measure of volume and should be converted to Liters, a cup is equal to 250 ml. However, if the item was "sal grosso" (salt)", "a cup" was converted into kg (one cup equal to 180 mg of salt) since purchases of salt were recorded more often in kg's rather in liters.

3) Trade-offs in converting units.

Converting all units to Kg's:

It is theoretically possible to convert units even further. For example, one may try to convert quantities measured in Liters to Kg's by using the estimated densities of each product. Let's consider the trade-offs of such conversion.

Table 4 shows the cross tabulation of standardized units for consumption and purchases.

	Units: purchases									
Units: consumption	Kg	Units	Litres	Canteirho	Molho	Montinho	Box	Other	MISSING	Total
Kg	4,367	44	96	0	7	4	1	3	1,013	5,535
Units	70	2,597	12	0	21	44	11	8	451	3,214
Litres	212	23	1,983	0	11	2	0	1	593	2,825
Canteirho	0	0	1	0	0	0	0	0	2	3
Molho	13	12	2	1	1,323	13	1	0	169	1,534

Montinho	8	7	10	0	15	1,265	0	1	88	1,394
Box	0	0	0	0	1	0	17	0	11	29
Other	1	0	1	0	0	0	0	67	10	79
Total	4,671	2,683	2,105	1	1,378	1,328	30	80	2,337	14,613

Note: To illustrate the point, this table includes only those observations, for which there are no observed unit values for produced food or gifts. Therefore, purchases are the only source of unit values.

Converting Liters to Kg's for all the items would allow using household level prices rather than regional prices. At the same time, the MISSING column shows that imputation using regional prices has to be made for 2,337 cases anyway! These are the cases in which households report consumption of an item, but do not report any values for purchases, gifts, or home production.

Conversion of Liters to Kg's is difficult, because one needs to know densities of the product, which can depend on the type of the product, quality of the product, the way it is packed, etc. For example, fine rice will have a different density than coarse rice. In case of grains, it depends on the "estado", or type of the grain, i.e. whether the grain comes with "espiga" (stalk), or is it already processed – the information that we do not observe in the data. It rules out the possibility to use the conversion file ("Rui's" file) which gives different conversion rates for several types of grains, depending on the "estado" type of the grain. Furthermore, the conversion files we have do not contain all the products and measurements from the PSIA.

Converting other units to Kg is even harder than converting Liters to Kg and involves a lot of assumptions. For example, one would need to assume how much kilos are in one banana or one "molho" of tomatoes. Overall, the difficulty of converting Liters and other units to Kg does not seem to be worth the benefit that it can provide.

Converting mass to Kg's and volume to Liters:

A comment on conversion rules used in Table 2: Table 2 shows that a "sack of 100 kg" is converted to 100 kg. A question arises whether other conversion rules from Rui's file need to be used here?

Rui's file gives several conversion rules for each grain, depending on the "estado", or type of the grain. For example, possible rules for conversion of "sack of 100 kg" of maize (milho) are:

pesokg1 pesokg2 estado 58.098 29.62998 1

57.546	38.55582	2
100.05	100.05	3

(here pesokg1 is the weight of raw grain and pesokg2 is the weight of "standard grain").

Unfortunately, using Rui's file is not possible with the Mozambique data: we do not observe type of the estado, nor do we know whether pesokg1 or pesokg2 is used in the recording. In the absence of this information, I use the default conversion rules for estado = 3. These are the most straightforward rules, that convert a "sack of 100kg" to a 100 kg, and pesokg1= pesokg2.

Dealing with outliers:

It is important to examine outliers and typos when dealing with the price imputation. This is done in the following way:

1. Household level price is replaced by national median price if it is more than 5 times larger or less than 5 times smaller than the median price.

2. There were several cases when the interviewer made the following mistake:

When recording consumption of beverages, the unit was 500 ml, and the quantity was also 500. Clearly, the household did not consume 500 bottles in a week, and in such cases quantity was replaced to 1.

3. If the per capita quantity of consumed item is less than 1 gram, more than 100 kg, more than 100 units, or more than 100 liters, these cases were marked as outliers. There were two cases like this:

case_id	code	quantity	unit	hhsize	per capita quant
42304601	SeaSalt	15	Bag of 90kg	7	192 kg
59802501	Fresh fish	10	Bag of 90kg	6	150 kg

5 households did not report consuming any food in the last 7 days:

Non-Food consumption:

Non-food expenditure consists of several components: monthly expenditure, annual expenditure, expenditure on education, durable goods.

1. Monthly expenditures reported in section gc were added together and converted to annual values.

Only 25 households reported paying housing rent. Since such small number of observations does not allow to impute the value of rent for the households that own their houses and do not pay rent, actual rent expenditures were not included in consumption aggregate.

There were two suspected outliers with very high values of expenditures:

case_id	gc01	gc02	gc03	gc04
23803200	DESPESAS DE PAGAMENTO DE ENERG	despesas	sim	22278
68003200	DESPESAS DE PAGAMENTO DE ÁGUA	despesas	sim	18000

These values were replaced by average expenditure on the corresponding item.

2. Annual expenditures reported in section gd were added together.

There was one household (case_id = 75611500) which had expenditure values = 999999 for all items.

These values were replaced to missing.

3. Education expenditures.

All education expenditures reported in section C2 were added together. When the trimester values were reported instead of annual values (c204==2), they were added together and used for the annual value.

4. Durables – done by Melissa Gaal in consultation with Kenneth Simler and Rose Mungai.

Assumptions used during the calculation of durable goods consumption

1. For most of the goods, the depreciation rate was estimated in the data, by using 1/(expected lifetime), where expected lifetime= 2*avg. age of good in the data. This rate was compared with the depreciation rate used for the IAF 2003 survey. If there was a difference of more than 4 years in the expected lifetime of the goods, the depreciation rate of the IAF was used.

2. Median purchase price was used for purchase value of goods that did not have a purchase price (if f07==. , but the household owned the good, f03==1)

3. Current value of the good was done as following:

a) for new goods (less than 12 months), use purchase price (f07)
b) for older goods, use the formula straight depreciation formula:
current value= median_pricevalue - (median_pricevalue *dep_rt*age) if f07=.

4. Old assets, for which we don't have the age, were given ½ the value of new assets (this was done in IAF), and assumed to be twice as old as new assets.

5. "Use value" of the good was calculated using Deaton and Zaide methodology (formula 2.12):

use_value=currentvalue*(real interet rate + depreciation)

6. When use_value is negative because "age of good" is too old (35, 50, years), replace the use_value with the minimum use_value calculate for that good that is not less than zero. This happened in 1.6% of cases.

APPENDIX 2: TABLES DESCRIBING ATTRITION

Table 1: Share of non-respondents in the NPS enumeration areas of the IAF sample for households and individuals

	Share of non-respondents among households			Share of non-respondents among children (0-17)							Share of non-respondents among adults (18+)		
				Total	Females			Males			Total	Females	Males
Province	Total	Rural	Urban		Total	Rural	Urban	Total	Rural	Urban			
Cabo Delgado	28%	23%	62%	39%	41%	38%	56%	36%	32%	75%	53%	52%	54%
Gaza	17%	14%	20%	38%	40%	39%	41%	36%	37%	35%	52%	49%	55%
Inhambane	14%	15%	8%	32%	30%	30%	35%	34%	33%	47%	56%	53%	59%
Manica	33%	32%	33%	42%	44%	35%	49%	40%	40%	39%	53%	52%	55%
Maputo Cidade	39%		39%	48%	50%		50%	47%		47%	62%	63%	62%
Maputo Province	49%	44%	52%	61%	62%	64%	61%	59%	59%	60%	68%	66%	69%
Nampula	23%	21%	29%	42%	44%	39%	61%	40%	38%	44%	51%	47%	55%
Niassa	29%	21%	34%	47%	49%	44%	53%	45%	37%	51%	51%	50%	52%
Sofala	40%	37%	43%	49%	50%	53%	46%	48%	45%	51%	56%	53%	60%
Tete	25%	24%	24%	36%	34%	35%	32%	37%	38%	35%	42%	39%	45%
Zambezia	16%	13%	28%	27%	26%	24%	37%	28%	20%	51%	48%	43%	54%
Total	29%	22%	36%	42%	43%	38%	49%	42%	37%	47%	55%	53%	57%

Notes: Unweighted. The share of non-respondents among households is the share of households that were not found at the screening stage (i.e. they did not answer section b0, and a20 = No). The respondent individuals are the individuals who responded to the full NPS questionnaire.

Table 2: Demographic characteristics in 2003 of respondent and non-respondent households: Rural sample

	Non-respondents	Respondents	Obs.	Difference between samples		
				F-test	p-value	
Demographic characteristics						
Number of household members	3.61 (0.19)	4.80 (0.18)	1211	35.54	0.00	***
Age of the household head	43.64 (1.69)	43.35 (1.17)	1209	0.0218	0.89	
Share of children (0-17) in the household	0.36 (0.02)	0.49 (0.02)	1211	25.73	0.00	***
Share of primary school age children (7-14) in the household	0.12 (0.01)	0.19 (0.01)	1211	20.38	0.00	***
Household had a child or an member enrolled at school in 2003	0.66 (0.03)	0.90 (0.02)	1211	41.61	0.00	***
Wealth and education						
Asset index	-0.74 (0.14)	-0.73 (0.08)	1211	0.0202	0.89	
Mean education of the household members aged more than 25	0.71 (0.04)	0.74 (0.06)	1211	0.615	0.45	
School enrollment rate of children aged 7-14	0.67 (0.03)	0.69 (0.05)	695	0.160	0.70	
Occupation of the head						
Household head agricultural worker	0.87 (0.03)	0.88 (0.01)	1211	0.162	0.70	
Household head non-agricultural self-employed	0.04 (0.02)	0.04 (0.01)	1211	0.0135	0.91	
Household head government worker	0.03 (0.01)	0.03 (0.01)	1211	0.00819	0.93	
Household head private wage worker	0.04 (0.02)	0.03 (0.01)	1211	0.386	0.55	
Household head: Other (retired, jobless …)	0.02 (0.01)	0.02 (0.00)	1211	0.0339	0.86	
Health						
The household head reported a recent disease	0.25 (0.02)	0.21 (0.03)	1211	2.240	0.17	
Bottom quintile	0.13 (0.02)	0.24 (0.05)	1211	8.296	0.02	**
Second quintile	0.19 (0.03)	0.26 (0.04)	1211	2.328	0.16	
Third quintile	0.18 (0.03)	0.15 (0.03)	1211	0.814	0.39	
Fourth quintile	0.23 (0.04)	0.22 (0.02)	1211	0.0687	0.80	
Top quintile	0.27 (0.04)	0.13 (0.02)	1211	21.90	0.00	***

Notes: The respondents households are the households found at the screening stage (i.e. they did answer section b0, and a20 = Yes). These estimations use the "basic NPS weights". The standard errors of the estimators are corrected for the clustering by enumeration area. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Demographic characteristics in 2003 of respondent and non-respondent households: Urban sample

	Non-respondants	Respondants	Obs.	Difference between samples		
				F-test	p-value	
Demographic characteristics						
Number of household members	4.17 (0.40)	4.60 (0.87)	1023	0.751	0.41	
Age of the household head	38.91 (0.96)	45.18 (0.96)	1020	13.14	0.00	***
Share of children (0-17) in the household	0.39 (0.01)	0.41 (0.03)	1022	0.331	0.58	
Share of primary school age children (7-14) in the household	0.15 (0.01)	0.16 (0.01)	1022	0.571	0.47	
Household had a child or an member enrolled at school in 2003	0.78 (0.03)	0.75 (0.10)	1023	0.172	0.69	
Wealth and education						
Asset index	1.29 (0.93)	1.11 (1.07)	1023	0.536	0.48	
Mean education of the household members aged more than 25	1.46 (0.33)	1.33 (0.31)	1023	5.255	0.04	**
School enrollment rate of children aged 7-14	0.70 (0.09)	0.79 (0.10)	629	25.12	0.00	***
Occupation of the head						
Household head agricultural worker	0.60 (0.20)	0.58 (0.16)	1023	0.164	0.69	
Household head non-agricultural self-employed	0.11 (0.05)	0.08 (0.04)	1023	1.580	0.24	
Household head government worker	0.11 (0.05)	0.12 (0.06)	1023	0.884	0.37	
Household head private wage worker	0.14 (0.08)	0.18 (0.05)	1023	0.754	0.41	
Household head: Other (retired, jobless …)	0.04 (0.02)	0.05 (0.03)	1023	0.320	0.58	
Health						
The household head reported a recent disease	0.12 (0.05)	0.09 (0.03)	1023	2.140	0.17	
Share of households in each quintile of national distribution of household consumption per capita						
Bottom quintile	0.21 (0.12)	0.15 (0.05)	1023	0.917	0.36	
Second quintile	0.18 (0.04)	0.10 (0.02)	1023	3.461	0.09	*
Third quintile	0.17 (0.05)	0.32 (0.13)	1023	3.311	0.10	*
Fourth quintile	0.15 (0.06)	0.17 (0.03)	1023	0.393	0.55	
Top quintile	0.29 (0.16)	0.26 (0.15)	1023	0.752	0.41	

Notes: The respondents households are the households found at the screening stage (i.e. they did answer section b0, and a20 = Yes). These estimations use the "basic NPS weights". The standard errors of the estimators are corrected for the clustering by enumeration area. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Demographic characteristics in 2003 of respondent and non-respondent individuals: Rural sample

	Non-respondants	Respondants	Obs.	Difference between samples		
				F-test	p-value	
Adults occupation and education (Age > 18)						
Education	0.78 (0.04)	0.79 (0.04)	2668	0.177	0.67	
Agricultural workers	0.85 (0.01)	0.90 (0.01)	2669	8.701	0.00	***
Non-agricultural self-employed	0.04 (0.01)	0.03 (0.01)	2669	0.171	0.68	
Government workers	0.02 (0.00)	0.02 (0.01)	2669	0.0551	0.82	
Prrivate wage workers	0.02 (0.01)	0.01 (0.00)	2669	2.415	0.12	
Students	0.03 (0.01)	0.01 (0.00)	2669	7.069	0.01	**
Others (Retired, ill, …)	0.04 (0.01)	0.02 (0.01)	2669	3.621	0.06	*
Children's occupation (age < 18)						
Working	0.15 (0.01)	0.08 (0.01)	3020	22.12	0.00	***
Students	0.38 (0.02)	0.35 (0.02)	3020	1.528	0.22	
Too young to have an occupation	0.09 (0.01)	0.11 (0.01)	3020	2.888	0.09	*
Others (ill, jobless)	0.01 (0.00)	0.01 (0.00)	3020	0.616	0.43	
Other individual characteristics						
Age	24.94 (0.60)	19.38 (0.40)	5683	68.62	0.00	***
Reported disease in 2003	0.18 (0.01)	0.16 (0.01)	5689	0.877	0.35	
Share of households in each quintile of national distribution of household consumption per capita						
Bottom quintile	0.21 (0.05)	0.27 (0.05)	5689	5.229	0.05	**
Second quintile	0.22 (0.02)	0.26 (0.03)	5689	2.617	0.14	
Third quintile	0.17 (0.02)	0.18 (0.03)	5689	0.376	0.56	
Fourth quintile	0.23 (0.03)	0.20 (0.03)	5689	1.307	0.28	
Top quintile	0.17 (0.03)	0.09 (0.02)	5689	18.63	0.00	***

Notes: The respondents are the individuals who responded to the full NPS questionnaire. These estimations use the "basic NPS weights". The standard errors of the estimators are corrected for the clustering by enumeration area. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Demographic characteristics in 2003 of respondent and non-respondent individuals: Urban sample

	Non-respondants	Respondants	Obs.	Difference between samples F-test	p-value	
Adults occupation and education (Age > 18)						
Education	1.63 (0.24)	1.61 (0.27)	2865	0.148	0.70	
Agricultural workers	0.55 (0.13)	0.55 (0.10)	2869	0.0278	0.87	
Non-agricultural self-employed	0.09 (0.03)	0.11 (0.02)	2869	1.435	0.23	
Government workers	0.06 (0.02)	0.09 (0.02)	2869	11.39	0.00	***
Prrivate wage workers	0.13 (0.02)	0.10 (0.02)	2869	9.519	0.00	***
Students	0.07 (0.02)	0.08 (0.02)	2869	0.223	0.64	
Others (Retired, ill, …)	0.10 (0.03)	0.08 (0.02)	2869	0.780	0.38	
Children's occupation (age < 18)						
Working	0.10 (0.03)	0.02 (0.01)	2839	4.613	0.03	**
Students	0.37 (0.08)	0.48 (0.05)	2839	6.991	0.01	**
Too young to have an occupation	0.13 (0.05)	0.07 (0.02)	2839	3.048	0.08	*
Others (ill, jobless)	0.02 (0.01)	0.01 (0.00)	2839	2.844	0.10	*
Other individual characteristics						
Age	24.33 (1.50)	20.20 (0.45)	5692	10.35	0.00	***
Reported disease in 2003	0.10 (0.02)	0.13 (0.01)	5708	1.618	0.21	
Share of households in each quintile of national distribution of household consumption per capita						
Bottom quintile	0.19 (0.11)	0.05 (0.02)	5708	1.377	0.27	
Second quintile	0.09 (0.03)	0.18 (0.06)	5708	1.063	0.33	
Third quintile	0.22 (0.09)	0.26 (0.13)	5708	1.111	0.32	
Fourth quintile	0.18 (0.04)	0.17 (0.06)	5708	0.457	0.51	
Top quintile	0.32 (0.15)	0.35 (0.14)	5708	0.789	0.40	

Notes: The respondents are the individuals who responded to the full NPS questionnaire.

These estimations use the "basic NPS weights". The standard errors of the estimators are corrected for the clustering by enumeration area. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Probit models predicting household response in the NPS survey

	All households			urban	rural
Demographic characteristics of the household					
Number of household members (2003)	0.061**	0.065**	0.065**	0.032	0.074**
	(0.027)	(0.026)	(0.026)	(0.052)	(0.037)
Age of the head (2003)	0.009**	0.009**	0.009**	0.026***	0.001
	(0.004)	(0.004)	(0.004)	(0.009)	(0.005)
Share of children (0-17) in the household (2003)	0.517***	0.515***	0.516***	0.800**	0.538*
	(0.175)	(0.170)	(0.171)	(0.390)	(0.320)
There is a child of school age (7-14) in the household (2003)	-0.124	-0.155	-0.154	-0.937***	0.166
	(0.108)	(0.121)	(0.120)	(0.220)	(0.252)
Poverty indicators					
Asset index (2003)	-0.035	-0.018	-0.014	-0.019	0.052
	(0.035)	(0.026)	(0.026)	(0.031)	(0.095)
Mean education of the household members aged more than 25 (2003)	-0.012	-0.019	-0.019	-0.082	0.005
	(0.042)	(0.040)	(0.041)	(0.068)	(0.057)
School enrollment rate of children aged 7-14 (2003)	0.242***	0.271***	0.269***	0.978***	0.057
	(0.091)	(0.098)	(0.097)	(0.262)	(0.194)
The household head reported a recent disease (2003)	-0.174**	-0.176**	-0.177**	-0.367**	-0.086
	(0.086)	(0.088)	(0.088)	(0.152)	(0.108)
Occupation of the head (ref: agricultural worker) (2003)					
Household head non-agricultural self-employed	-0.113	-0.076	-0.078	0.032	-0.134
	(0.164)	(0.187)	(0.188)	(0.287)	(0.263)
Household head government worker	0.039	0.039	0.033	0.340	-0.258
	(0.259)	(0.273)	(0.275)	(0.378)	(0.297)
Household head private wage worker	0.205	0.274	0.280	0.491	-0.148
	(0.335)	(0.355)	(0.354)	(0.452)	(0.165)
Household head: Other (retired, jobless …)	0.097	0.136	0.145	0.257	0.098
	(0.131)	(0.149)	(0.149)	(0.299)	(0.245)
Geography					
Maputo City Province			-0.310**	-0.443**	
			(0.157)	(0.204)	
Maputo Province			-0.404*	-0.526*	-0.310
			(0.229)	(0.301)	(0.276)
Rural enumeration area	0.386***	0.462***	0.464***		
	(0.078)	(0.069)	(0.069)		
Central Region (ref: Northern)		-0.054	-0.098	-0.124	-0.067
		(0.170)	(0.190)	(0.236)	(0.180)
Southern Region (ref: Northern)		-0.167	0.076	0.152	-0.187
		(0.142)	(0.180)	(0.253)	(0.168)
Rural * Central Region (ref: Northern)		-0.012	0.015		
		(0.130)	(0.129)		
Rural * Southern Region (ref: Northern)		-0.194*	-0.281**		
		(0.106)	(0.110)		
Exclusion restriction					
Field team quality index (2008)	0.393***	0.466***	0.337***	0.312**	0.390**
	(0.087)	(0.067)	(0.123)	(0.146)	(0.164)
Observations	2229	2229	2229	1020	1209
Pseudo R-squared	0.0798	0.0835	0.0841	0.0933	0.0768

Notes: The respondent households are the households found at the screening stage (i.e. they did answer section b0, and a20 = Yes). The standard errors of the estimators are corrected for the correlation between different observations of the same province which

occurs with cluster sampling. These estimations use the “basic NPS weights”. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Probit models predicting child inclusion in the NPS sample

Sample		0-17y.o.		0-17y.o. urban	0-17y.o. rural	10-17 y.o.
Demographic characteristics of the household						
Number of household members (2003)	0.018**	0.017*	0.015	0.026*	0.011	0.022*
	(0.009)	(0.010)	(0.010)	(0.015)	(0.013)	(0.011)
Age of the head (2003)	-0.009**	-0.009**	-0.009**	-0.004	-0.008	-0.006
	(0.004)	(0.004)	(0.004)	(0.007)	(0.005)	(0.006)
Share of children (0-17) in the household (2003)	0.264	0.270	0.281	0.778**	-0.107	0.280
	(0.194)	(0.187)	(0.187)	(0.315)	(0.271)	(0.251)
There is a child of school age (7-14) in the household (2003)	0.123	0.137	0.140	-0.264	0.266*	-0.105
	(0.104)	(0.119)	(0.117)	(0.326)	(0.140)	(0.202)
Poverty indicators						
Asset index (2003)	0.023*	0.016	0.016	0.004	0.057	0.018
	(0.013)	(0.010)	(0.012)	(0.014)	(0.063)	(0.021)
Mean education of the household members aged more than 25 (2003)	-0.087	-0.095*	-0.093	-0.163***	-0.011	-0.0425
	(0.061)	(0.058)	(0.058)	(0.061)	(0.061)	(0.0609)
School enrollment rate of children aged 7-14 (2003)	0.212**	0.182*	0.180*	0.648***	0.044	0.283
	(0.100)	(0.096)	(0.096)	(0.146)	(0.111)	(0.221)
The household head reported a recent disease (2003)	0.040	0.033	0.035	-0.184	0.056	0.004
	(0.055)	(0.059)	(0.058)	(0.151)	(0.068)	(0.06)
Occupation of the head (ref: agricultural worker) (2003)						
Household head non-agricultural self-employed	-0.189	-0.215*	-0.221*	-0.040	-0.341*	-0.529***
	(0.131)	(0.123)	(0.124)	(0.190)	(0.197)	(0.139)
Household head government worker	-0.038	-0.070	-0.073	0.214	-0.475	-0.314**
	(0.161)	(0.146)	(0.147)	(0.181)	(0.316)	(0.132)
Household head private wage worker	-0.035	-0.058	-0.056	0.109	-0.114	-0.295***
	(0.073)	(0.069)	(0.066)	(0.134)	(0.113)	(0.110)
Household head: Other (retired, jobless …)	0.106	0.022	0.021	-0.053	0.370*	-0.00472
	(0.108)	(0.108)	(0.110)	(0.120)	(0.196)	(0.162)
Individual demographic characteristics						
Sex (1 for females)	-0.003	-0.006	-0.007	0.058	-0.041	-0.029
	(0.031)	(0.030)	(0.030)	(0.097)	(0.040)	(0.094)
Age (2003)	-0.035**	-0.034**	-0.034**	-0.038***	-0.025	-0.089***
	(0.014)	(0.015)	(0.015)	(0.008)	(0.020)	(0.028)
Reported disease (2003)	0.037	0.026	0.025	0.303	-0.087	-0.072
	(0.062)	(0.068)	(0.068)	(0.196)	(0.061)	(0.079)
Own occupational choice (2003)						
Own education in 2003	0.006	0.002	0.001	0.026	-0.044	0.036
	(0.106)	(0.105)	(0.105)	(0.068)	(0.116)	(0.097)
Student (ref: agricultural worker)	0.328***	0.339***	0.338***	0.762***	0.241	0.390**
	(0.090)	(0.090)	(0.091)	(0.239)	(0.162)	(0.166)
Considered too young to work (ref: agricultural worker)	0.271**	0.276***	0.275***	0.492**	0.297**	-0.136
	(0.110)	(0.104)	(0.104)	(0.244)	(0.149)	(0.185)
Sick, jobless, or missing info (ref: agricultural worker)	0.355	0.356	0.355	0.180	0.519**	0.513
	(0.252)	(0.267)	(0.267)	(0.426)	(0.237)	(0.374)
Less than 5 years old: no information (ref: agricultural worker)	0.184	0.203	0.202	0.453	0.205	
	(0.128)	(0.132)	(0.131)	(0.278)	(0.187)	
Geography						
Maputo City Province			0.015	0.227**		0.231
			(0.175)	(0.107)		(0.155)
Maputo Province			-0.195	0.173	-0.502*	-0.129
			(0.197)	(0.147)	(0.303)	(0.176)
Rural enumeration area	0.384***	0.446***	0.447***			0.396***
	(0.100)	(0.038)	(0.038)			(0.0314)
Central Region (ref: Northern)		0.331***	0.315***	0.119	0.203	0.255***
		(0.067)	(0.080)	(0.084)	(0.153)	(0.0818)
Southern Region (ref: Northern)		0.274***	0.318**	-0.084	-0.019	0.239*
		(0.088)	(0.156)	(0.103)	(0.183)	(0.134)
Rural * Central Region (ref: Northern)		-0.144	-0.134			0.022
		(0.123)	(0.116)			(0.243)
Rural * Southern Region (ref: Northern)		-0.454***	-0.419***			-0.505***
		(0.122)	(0.134)			(0.147)
Exclusion restriction						
Field team quality index (2008)	0.140**	0.266***	0.200**	0.238***	0.179	0.190**
	(0.068)	(0.058)	(0.081)	(0.068)	(0.152)	(0.0760)
Observations	5846	5846	5846	2837	3009	1913
Pseudo R-squared	0.0547	0.0615	0.0619	0.0925	0.0416	0.0898

Notes: The respondents are the individuals who responded to the full NPS questionnaire. The standard errors of the estimators are corrected for the correlation between different observations of the same province which occurs with cluster sampling. These estimations use the "basic NPS weights". *** p<0.01, ** p<0.05, * p<0.1

APPENDIX 3: CONTROL FOR THE SELECTION AND REWEIGHTING

Description of the IPW reweighting technique

The IAF sample was representative of the Mozambican population in 2003. Hence the NPS enumeration areas of the IAF sample are representative as well, provided the selection of NPS enumeration areas is not biased. The children aged 10-17 in 2003 in the NPS enumeration areas of IAF survey are therefore representative of the Mozambican children aged 10-17 in 2003, if the "basic NPS weights" are used.

The potential selection bias is therefore due to the impossibility to resurvey all household and individuals from the NPS enumeration areas. The probability to be resurveyed is predicted by the probit specifications of Tables 6 and 7 in appendix 2. So the basic idea of reweighting is to predict the probability of being resurveyed with probit models such as these in Tables 6 and 7, and to construct new weights. There is nevertheless one difference with these Tables: we do not need to make any interpretation of the coefficients for selection. We just need to predict as closely as possible the selection. There is one consequence to that: the probit specifications include dummies for (nearly) each IAF stratum. The coefficients for these regressions are given in Table 14.

Wooldridge (2002) describes the hypotheses under which the estimators using this technique are unbiased. IPW is valid whenever the unobservable characteristics underlying the selection process are uncorrelated with the unobservable characteristics explaining the output variable. If this hypothesis is violated, it is necessary to treat attrition as a selection bias in regressions similar to the 2-stage Heckman approach. These approaches are a strategy to test the validity of the IPW techniques. This is done below with child education, and the results seemingly confirm the validity of IPW (and the absence of severe selection bias in this case).

CONTROL FOR THE SELECTION WITH THE SAMPLE OF HOUSEHOLDS

To analyze the possible bias caused by household attrition, we estimated an equation predicting household school enrollment in 2008. We run this estimation for the school enrollment of children aged 7-14. The baseline equation is:

$$E_{it} = \alpha E_{i,t-1} + X_{i,t-1}\beta + \varepsilon_{it} \qquad (1)$$

In this equation, E_{it} denotes the school enrollment rate in household *i* at date *t*, X_{it-1} are the covariates. The panel is therefore made of the households including at least one child aged 7-14 in 2003 and in 2008. We fear that households in the NPS sample at date *t* are non-randomly

selected. We also fear that households including children aged 7-14 in 2008 are non-randomly selected.[3] Therefore, $\square$ and $\square$ may be biased due to selection. For convenience, I write the equation for selection (s_i = 1 whenever a household is in the NPS sample in 2008 and includes a child, 0 otherwise):

$$s_i = 1\left(\alpha_s E_{i,t-1} + X_{i,t-1}\beta_s + \delta_s Z_{it} + u_{it}\right) \quad (2)$$

Table 11 gives the results of the estimation of equation (1) when controlling for selection. Column 1 gives the results of the estimation by OLS, with no control for selection. Column 2 gives the results with OLS corrected with reweighting. Column 3 gives the results with the Heckman procedure. As the selection is not exactly the same than household selection in Table 4, the first stage for the Heckman procedure is not exactly the same and is given in Table 15 . The main result in this section is that none of the control for selection affected any of the coefficients of the estimation of equation (1). This note does not aim at finding the determinants of the school trajectory of the children between 2003 and 2008, so no comments are given on the coefficients. The differences between columns in Table 11 are negligible. The Heckman procedure includes a test for the correlation between the error terms u_{it} and $\square_{it}$. All the tests fail to reject the null. In this case, the correction for selection seems unnecessary. Therefore, reweighting the samples with probabilities would be suggested in this case to improve the representativity of the population.

Table 8: Prediction of household enrollment rate and controls for selection

[3] We do not ask whether households including a child aged 7-14 in 2003 are selected. Indeed, they are supposed representative of the national sample of household including a child aged 7-14, and it does not make sense to compute school enrollment rate in other households.

	OLS (1)	reweighted (2)	Heckman (3)
Lagged variable			
School enrollment rate of children aged 7-14 (2003)	0.230*** (0.048)	0.246*** (0.045)	0.225*** (0.047)
Demographic characteristics of the household			
Number of household members (2003)	0.006 (0.004)	0.006 (0.004)	-0.001 (0.011)
Age of the head (2003)	0.000 (0.002)	0.000 (0.002)	0.001 (0.002)
Share of children (0-17) in the household (2003)	-0.079 (0.130)	-0.131 (0.117)	-0.190 (0.216)
Poverty indicators			
Asset index (2003)	-0.014** (0.005)	-0.012** (0.005)	-0.018* (0.009)
Mean education of the household members aged more than 25 (2003)	0.078*** (0.022)	0.071*** (0.018)	0.093** (0.038)
The household head reported a recent disease (2003)	0.061 (0.037)	0.077* (0.037)	0.062 (0.036)
Occupation of the head (ref: agricultural worker) (2003)			
Household head non-agricultural self-employed	0.091*** (0.022)	0.077*** (0.021)	0.113** (0.040)
Household head government worker	0.006 (0.039)	0.008 (0.036)	-0.006 (0.044)
Household head private wage worker	0.016 (0.020)	0.016 (0.021)	0.013 (0.019)
Household head: Other (retired, jobless …)	0.072* (0.039)	0.066** (0.024)	0.070* (0.039)
Geography			
Maputo City Province	-0.004 (0.019)	0.004 (0.023)	0.035 (0.064)
Maputo Province	0.023 (0.020)	0.019 (0.022)	0.067 (0.065)
Rural enumeration area	0.028 (0.029)	0.023 (0.033)	-0.005 (0.073)
Central Region (ref: Northern)	0.085* (0.044)	0.076* (0.041)	0.101* (0.051)
Southern Region (ref: Northern)	0.128*** (0.037)	0.118** (0.037)	0.115** (0.047)
Rural * Central Region (ref: Northern)	-0.071 (0.065)	-0.060 (0.064)	-0.077 (0.064)
Rural * Southern Region (ref: Northern)	0.003 (0.043)	0.009 (0.048)	0.014 (0.053)
Other			
Mills ratio			-0.138 (0.229)
Observations	788	788	788
R-squared	0.209	0.233	0.210

Notes: The household enrollment rate is the enrollment rate among children aged 7-14. The standard errors of the estimators are corrected for the correlation between different observations of the same province which occurs with cluster sampling. These estimations use the "basic NPS weights", unless reweighted. *** p<0.01, ** p<0.05, * p<0.1.

Comparison of descriptive statistics with a nationally representative survey

This section compares descriptive statistics from a nationally representative survey of Mozambican households (MICS 2008) with the statistics from the NPS sample in three cases. IPW can easily correct for the selection bias in this for these descriptive statistics.

Table 8 compares primary school enrollment rates between MICS and NPS. Primary school enrollment rate is the enrollment rate for children aged 6 – 12 at the beginning of 2008. Age at the beginning of 2008 is not available in NPS, so this Table uses the best approximation available: the age 1 year before the interview. This may nevertheless limit the comparability between both Tables. Overall, the statistics issued from NPS are rarely statistically different from the statistics from MICS. The school enrollment rate at age 6 is nevertheless lower in NPS, especially for girls. It appears that most of the estimations are rather imprecise, as the sample size from NPS is quite small for such approximation. Indeed, the primary school enrollment rate in urban areas is approximately 89% in MICS for boys, and 80% in NPS. The difference between the two is not significant at the 10% level, although 9 percentage points are a lot.

Table 9 compares secondary school enrollment rates between MICS and NPS. Overall, the coefficients are again imprecise and the only significant difference is in the comparison between age groups. It could be that the age is less precisely measured in NPS, as Table 8 underestimates primary school enrollment at age 6, and Table 9 does not detect the peak in secondary school enrollment at age 16 in MICS.

Table 10 compares the primary school completion rate between MICS and NPS. The primary school completion rate is the primary completion rate of children being aged 12 at the beginning of the current school year. It neglects delayed completion rate, and is therefore very sensitive to the measure of age. Indeed, delays in school enrollment in Mozambique may be frequent, because of grade repetition and because of delayed school entries. The primary school completion rate of children aged 13 at the date of the interview is 11% after reweighting in NPS. It would be 33.8% among children aged 14. In Table 10, there isn't any statistically significant difference between primary school completion rates between MICS and NPS at the 5% level. One reason for that is that the sample sizes are very small. For example, the 14.7% of primary school completion rate in urban areas from NPS (with IPW) is much smaller than the 30.5% from MICS; but the NPS statistic is computed from 159 individuals.

The secondary school transition rate is the share of pupils enrolled in secondary school among those being in the last year of primary school the last school year. There is no statistically significant difference between NPS and MICS for this statistic. This is nevertheless not surprising, as sample sizes are even smaller.

Table 9: Comparison between MICS and NPS, primary school enrollment rates in 2008

	Boys					Girls					Total				
	MICS		EOPS			MICS		EOPS			MICS		EOPS		
	Share	Obs.	Share	Share IPW	Obs.	Share	Obs.	Share	Share IPW	Obs.	Share	Obs.	Share	Share IPW	Obs.
Total	82.3	6478	79.8	79.9	805	80.2	6686	75.1*	79.4	834	81.3	13164	77.4	79.6	1632
Zone															
Urban	89.3	1950	80.7	80.3	335	88.4	2114	77	85.8	325	88.8	4064	79	83.7	660
Rural	79.3	4532	79.4	79.7	470	76.5	4578	74.5	74.9	509	77.9	9110	76.8	77.1	979
Age															
6	73.1	998	62.7	64.9	103	68.2	1059	53.3**	58.6	118	70.6	2058	57.2**	61.1**	221
7	76.7	1079	73	70.4	145	76.7	1133	70.9	72.5	131	76.7	2212	72.2	71.2	276
8	82.8	920	75.6	78.1	100	82.9	933	87.1	88.4	110	82.8	1853	82.9	84.2	210
9	84.1	1014	87.4	88.8	136	81.0	1092	83	81	142	82.5	2106	85	84.7	278
10	90.2	806	88	88.6	94	86.9	796	91.5	92.1*	112	88.6	1602	89.8	90.4	206
11	86.0	951	86	85.6	137	84.6	974	62.4	80.1	128	85.3	1925	73.7	81.6	265
12	87.1	713	87.7	89.7	90	85.9	705	84.4	85.3	93	86.5	1417	86.1	87.6	183

Notes: This Table compares the primary school enrollment rate in 2008 in MICS and NPS data for children aged 6 – 12 at the beginning of 2008. Age at the beginning of 2008 is not available in NPS. This Table uses the best approximation available, which is the age 1 year before the interview. The first column in NPS (denoted Share) is weighted with the "basic NPS weights". The second column in NPS (denoted Share IPW) is weighted with inverse probability weighting so as to control for the selection bias. * significantly different from MICS at 10% ** at 5% *** at 1%. The standard errors of the MICS estimations are not given, so the Table takes into account the imprecision of NPS only.

Table 10: Comparison between MICS and NPS, secondary school enrollment rates in 2008

	Boys					Girls					Total				
	MICS		EOPS			MICS		EOPS			MICS		EOPS		
	Share	Obs.	Share	Share IPW	Obs.	Share	Obs.	Share	Share IPW	Obs.	Share	Obs.	Share	Share IPW	Obs.
Total	20.7	3247	20.2	22.9	456	20.2	3097	16.2	16.1	448	20.4	6,344	18.2	19.3	902
Zone															
Urban	37.6	1207	31.5	33.2	206	37.6	1269	37.7	30.7	206	37.6	2,476	34.3	31.9	412
Rural	10.7	2043	15	17.7	250	8.0	1828	7.8	8.6	242	9.5	3,870	11.4	12.9	492
Age															
13	10.5	865	8.6	12.8	106	9.5	861	13.8	15.3	104	10.0	1,727	10.8	13.9	210
14	12.9	657	12.3	13.9	96	23.1	584	16.3	12.6*	82	17.7	1,241	14	13.2	178
15	22.9	615	29.8	33.3*	78	23.1	596	18.5	20.8	81	23.0	1,211	23.7	26.5	159
16	35.5	461	23.5**	27	88	33.4	442	15.3***	15.7***	81	34.5	903	19.4***	21***	169
17	29.6	651	40	39.8	88	20.0	613	17.5	18.9	100	24.9	1,264	26.8	26.5	188

Notes: This Table compares the secondary school enrollment rate in 2008 in MICS and NPS data for children aged 13 – 17 at the beginning of 2008. Age at the beginning of 2008 is not available in NPS. This Table uses the best approximation available, which is the age 1 year before the interview. The first column in NPS (denoted Share) is weighted with the "basic NPS weights". The second column in NPS (denoted Share IPW) is weighted with inverse probability weighting so as to control for the selection bias. * significantly different from MICS at 10% ** at 5% *** at 1%. The standard errors of the MICS estimations are not given, so the Table takes into account the imprecision of NPS only.

Table 11: Comparison between MICS and NPS: Primary school completion and transition to secondary school in 2008

	Primary school completion rate					Secondary school transition rate				
	MICS		EOPS			MICS		EOPS		
	Share	Obs	Share	Share IPW	Obs	Share	Obs	Share	Share IPW	Obs
Total	15.3	1418	14.1	11	290	72.8	1,005	67.7	68.4	164
Zone										
Urbano	30.5	506	25.4	14.7*	159	74.5	650	80.8	77.3	109
Rural	7.0	913	6.4	6.2	131	69.6	355	56.8	58.2	55
Sex										
Male	14.1	713	12.5	8.7	154	74.9	576	70.1	71.2	88
Female	16.7	705	16.5	15.6	136	70.0	429	63.9	65	76

Notes: This Table compares the primary school completion rate and secondary school transition rate in 2008. The primary school completion rate is the share of children aged XXX who have finished primary school. This is consequently the primary school completion rate "on time" and delayed completions of primary school are neglected. The secondary school transition rate is the share of pupils enrolled in secondary school among those being in the last year of primary school the last school year. Age at the beginning of 2008 is not available in NPS. This Table uses the best approximation available, which is the age 1 year before the interview. The first column in NPS (denoted Share) is weighted with the "basic NPS weights". The second column in NPS (denoted Share IPW) is weighted with inverse probability weighting so as to control for the selection bias. * significantly different from MICS at 10% ** at 5% *** at 1%. The standard errors of the MICS estimations are not given, so the Table takes into account the imprecision of NPS only.

Overall, Tables 8, 9 and 10 show relatively moderate differences between the MICS statistics and the NPS sample. This section checks that the coefficients of regressions predicting school enrollment are the same once controlled for the selection, either with IPW or with a control for selection à la Heckman.

CONTROL FOR SELECTION USING THE NPS PANEL OF CHILDREN

To show the possible options to control for attrition in the panel of children, this section estimates an equation predicting school achievement in 2008. We run this estimation on children aged 11-17 in 2003, as young children may start their schooling after 2003. The NPS survey intended to track all the children aged 0-17 in 2003 from the NPS enumeration areas of the IAF sample (provided they were in the district in 2008). The baseline equation is:

$$A_{it} = \alpha A_{i,t-1} + X_{i,t-1}\beta + \varepsilon_{it} \quad (3)$$

In this equation, A_{it} denotes the school achievement for child *i* at date *t*, X_{it-1} are the covariates. We fear that individuals in the NPS sample at date *t* are non-randomly selected. Therefore, α and β may be biased due to selection. For convenience, I write the equation for selection (s_i = 1 whenever child *i* is in the NPS sample, 0 otherwise):

$$s_i = 1\left(\alpha_s A_{i,t-1} + X_{i,t-1}\beta_s + \delta_s Z_{it} + u_{it}\right) \quad (4)$$

This section uses the same techniques to control for section than section 2.2.4, namely IPW and Heckman's.

Table 12 gives the results of the estimation of equation (3) when controlling for selection. Column 1 gives the results of the estimation by OLS, with no control for selection. Column 2 gives the results with OLS corrected with reweighting. Column 3 gives the results with the Heckman procedure. Table 13 gives the same specifications than columns 1, 2, 3, split between the urban and rural samples.

The main result in this section is that none of the control for selection affected any of the coefficients of the estimation of equation (3). This note does not aim at finding the determinants of the school trajectory of the children between 2003 and 2008, so I don't give any comments to the coefficients. There are very few differences between columns in Tables 12 and 13. The Heckman procedure includes a test for the correlation between the error terms u_{it} and ε_{it}. All the tests but one fail to reject the null. In this case, the correction for selection seems unnecessary. Reweighting the samples with probabilities might help to me representative of the population but does not change the regression coefficients.

A few results in Table 13, column 6 with Heckman's procedure in urban areas are nevertheless different from columns 4 and 5. In particular, the coefficients for education of the household,

own education and school enrollment of the children in the household are affected by the control for selection.

Table 12: Prediction of children educational attainment and controls for selection

	OLS	reweighted	Heckman
Lagged variable			
Own education (2003)	0.589***	0.609***	0.592***
	(0.087)	(0.082)	(0.091)
Demographic characteristics of the household			
Number of household members (2003)	0.012	0.004	0.014
	(0.018)	(0.018)	(0.018)
Age of the head (2003)	0.001	0.002	0.001
	(0.003)	(0.003)	(0.003)
Share of children (0-17) in the household (2003)	-0.278	-0.225	-0.251
	(0.412)	(0.356)	(0.395)
There is a child of school age (7-14) in the household (2003)	0.355	0.328*	0.351
	(0.234)	(0.172)	(0.229)
Poverty indicators			
Asset index (2003)	0.050**	0.058***	0.052**
	(0.019)	(0.016)	(0.021)
Mean education of the household members aged more than 25 (2003)	0.130**	0.075	0.127***
	(0.043)	(0.046)	(0.042)
School enrollment rate of children aged 7-14 (2003)	-0.017	0.099	0.007
	(0.247)	(0.222)	(0.253)
The household head reported a recent disease (2003)	0.199	0.214**	0.200
	(0.125)	(0.094)	(0.124)
Occupation of the head (ref: agricultural worker) (2003)			
Household head non-agricultural self-employed	0.161	0.107	0.113
	(0.120)	(0.124)	(0.153)
Household head government worker	0.182	0.238	0.153
	(0.130)	(0.184)	(0.133)
Household head private wage worker	0.298***	0.314**	0.271***
	(0.071)	(0.102)	(0.089)
Household head: Other (retired, jobless …)	0.290**	0.257**	0.291***
	(0.105)	(0.096)	(0.109)
Individual demographic characteristics			
Sex (1 for females)	-0.272***	-0.217***	-0.273***
	(0.061)	(0.063)	(0.054)
Age (2003)	0.022	0.021	0.014
	(0.025)	(0.022)	(0.023)
Reported disease (2003)	0.051	0.073	0.044
	(0.126)	(0.143)	(0.120)
Own occupational choice (2003)			
Student (ref: agricultural worker)	0.577***	0.487***	0.614***
	(0.146)	(0.101)	(0.130)
Considered too young to work (ref: agricultural worker)	-0.198	-0.210	-0.207
	(0.255)	(0.190)	(0.240)
Sick, jobless, or missing info (ref: agricultural worker)	0.224	0.169	0.272
	(0.211)	(0.275)	(0.254)
Geography			
Maputo City Province	-0.241**	-0.266**	-0.240***
	(0.088)	(0.097)	(0.091)
Maputo Province	-0.108	0.072	-0.158
	(0.099)	(0.097)	(0.115)
Rural enumeration area	0.579***	0.311**	0.615***
	(0.127)	(0.132)	(0.132)
Central Region (ref: Northern)	0.839***	0.574**	0.857***
	(0.214)	(0.182)	(0.209)
Southern Region (ref: Northern)	1.002***	0.788***	1.041***
	(0.225)	(0.200)	(0.222)
Rural * Central Region (ref: Northern)	-0.810***	-0.531***	-0.807***
	(0.174)	(0.131)	(0.176)
Rural * Southern Region (ref: Northern)	-0.587***	-0.385**	-0.632***
	(0.163)	(0.157)	(0.175)
Arcth(rho)			0.176
			(0.196)
Observations	951	951	1913
R-squared	0.505	0.543	

Notes: The standard errors of the estimators are corrected for the correlation between different observations of the same province which occurs with cluster sampling. These estimations use the "basic NPS weights", unless reweighted. *** p<0.01, ** p<0.05, * p<0.1

Table 13: Prediction of children educational attainment and controls for selection: split Urban - Rural

	Rural			Urban		
	OLS	reweighted	Heckman	OLS	reweighted	Heckman
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged variable						
Own education (2003)	0.617***	0.639***	0.618***	0.484***	0.509***	0.461***
	(0.138)	(0.000)	(0.132)	(0.000)	(0.000)	(0.115)
Demographic characteristics of the household						
Number of household members (2003)	0.006	0.001	0.006	0.007	-0.010	-0.027
	(0.015)	(0.015)	(0.015)	(0.017)	(0.019)	(0.023)
Age of the head (2003)	-0.005	-0.005	-0.005	0.017***	0.015*	0.019**
	(0.005)	(0.004)	(0.005)	(0.005)	(0.007)	(0.009)
Share of children (0-17) in the household (2003)	-0.305	-0.377	-0.305	0.309	0.411	0.086
	(0.397)	(0.400)	(0.389)	(0.339)	(0.290)	(0.452)
There is a child of school age (7-14) in the household (2003)	0.454	0.349	0.455*	0.090	0.290	0.308
	(0.262)	(0.226)	(0.238)	(0.345)	(0.296)	(0.242)
Poverty indicators						
Asset index (2003)	0.049	0.011	0.047	0.061***	0.082***	0.061***
	(0.064)	(0.051)	(0.093)	(0.019)	(0.020)	(0.019)
Mean education of the household members aged more than 25 (2003)	0.140	0.118	0.139*	0.152*	0.084	0.219***
	(0.091)	(0.091)	(0.072)	(0.072)	(0.072)	(0.074)
School enrollment rate of children aged 7-14 (2003)	-0.016	0.018	-0.018	0.044	0.180	-0.392**
	(0.262)	(0.250)	(0.231)	(0.209)	(0.230)	(0.168)
The household head reported a recent disease (2003)	0.019	0.031	0.020	0.476**	0.424**	0.514***
	(0.136)	(0.110)	(0.133)	(0.159)	(0.138)	(0.146)
Occupation of the head (ref: agricultural worker) (2003)						
Household head non-agricultural self-employed	-0.021	-0.002	-0.010	0.391*	0.271	0.481**
	(0.220)	(0.216)	(0.369)	(0.214)	(0.207)	(0.226)
Household head government worker	0.011	0.101	0.023	0.269	0.321	0.241
	(0.163)	(0.168)	(0.408)	(0.182)	(0.225)	(0.233)
Household head private wage worker	0.526*	0.625**	0.535	0.309*	0.293*	0.290
	(0.241)	(0.252)	(0.350)	(0.169)	(0.161)	(0.203)
Household head: Other (retired, jobless …)	0.516**	0.420**	0.511***	0.220	0.246	0.208
	(0.159)	(0.182)	(0.165)	(0.133)	(0.155)	(0.231)
Individual demographic characteristics						
Sex (1 for females)	-0.149	-0.113	-0.148	-0.439**	-0.390**	-0.412**
	(0.125)	(0.131)	(0.111)	(0.176)	(0.167)	(0.210)
Age (2003)	0.000	-0.012	0.001	0.080	0.085	0.132**
	(0.027)	(0.026)	(0.045)	(0.071)	(0.059)	(0.052)
Reported disease (2003)	0.005	-0.070	0.006	0.117	0.328	0.286
	(0.127)	(0.124)	(0.133)	(0.267)	(0.188)	(0.279)
Own occupational choice (2003)						
Student (ref: agricultural worker)	0.587**	0.541**	0.583**	0.576***	0.518***	0.115
	(0.192)	(0.189)	(0.266)	(0.099)	(0.088)	(0.239)
Considered too young to work (ref: agricultural worker)	-0.186	-0.208	-0.184	-0.415	-0.476	-0.628
	(0.423)	(0.435)	(0.434)	(0.524)	(0.520)	(0.672)
Sick, jobless, or missing info (ref: agricultural worker)	0.357	0.369	0.345	0.100	-0.031	0.176
	(0.212)	(0.202)	(0.389)	(0.601)	(0.534)	(0.837)
Geography						
Maputo City Province				-0.165***	-0.232***	-0.184**
				(0.049)	(0.056)	(0.089)
Maputo Province	-0.220	0.066	-0.207	-0.069	-0.011	0.139
	(0.197)	(0.189)	(0.400)	(0.045)	(0.073)	(0.129)
Central Region (ref: Northern)	0.013	0.025	0.009	0.592**	0.414**	0.628**
	(0.092)	(0.085)	(0.170)	(0.230)	(0.169)	(0.245)
Southern Region (ref: Northern)	0.462**	0.476**	0.461***	0.668**	0.583**	0.627***
	(0.172)	(0.175)	(0.167)	(0.249)	(0.203)	(0.243)
Misc.						
Arcth(rho)			-0.029			-1.166**
			(0.708)			(0.579)
Observations	437	437	832	514	514	1081
R-squared	0.382	0.430		0.603	0.620	

Notes: The standard errors of the estimators are corrected for the correlation between different observations of the same province which occurs with cluster sampling. These estimations use the "basic NPS weights", unless reweighted. *** p<0.01, ** p<0.05, * p<0.1

Table 14: First stage regressions for the basic NPS Weights

age group	0-5 y.o.	6-10 y.o.	11-17 y.o.	18-25 y.o.	26-34 y.o.	35-50 y.o.	51+ y.o.	Households
Number of household members in 2003	0.011	0.058***	0.024	0.019	0.049*	0.110***	0.053**	0.070**
	(0.016)	(0.020)	(0.015)	(0.022)	(0.026)	(0.028)	(0.024)	(0.027)
Age of the head in 2003	0.004	0.001	-0.009*	0.000	0.009	-0.001	-0.023**	0.010
	(0.003)	(0.006)	(0.004)	(0.005)	(0.008)	(0.009)	(0.010)	(0.006)
Share of children (0-17) in the household	-0.023	-0.422	0.206	1.494***	1.836***	0.896**	2.098***	0.097
	(0.438)	(0.445)	(0.339)	(0.301)	(0.287)	(0.389)	(0.442)	(0.389)
There is a child of school age (7-14) in the household	-0.051	0.522	-0.416*	-0.363	0.278	0.056	0.195	-0.184
	(0.134)	(0.355)	(0.234)	(0.226)	(0.298)	(0.199)	(0.365)	(0.209)
Asset index	0.036	-0.019	0.033	0.009	0.021	0.012	0.049	0.004
	(0.025)	(0.025)	(0.024)	(0.023)	(0.028)	(0.026)	(0.044)	(0.027)
Mean education of the household members aged more than 25	-0.082	-0.089	-0.037	0.009	-0.189*	0.036	-0.125	-0.038
	(0.059)	(0.084)	(0.055)	(0.053)	(0.099)	(0.119)	(0.127)	(0.038)
School enrollment rate of children aged 7-14	0.509**	-0.499	0.515***	0.078	-0.009	0.167	-0.016	0.282
	(0.259)	(0.406)	(0.189)	(0.216)	(0.211)	(0.148)	(0.305)	(0.190)
The household head reported a recent disease	0.146	0.025	-0.009	-0.191	0.272	-0.029	0.062	-0.149
	(0.148)	(0.132)	(0.132)	(0.132)	(0.191)	(0.168)	(0.189)	(0.099)
Household head non-agricultural self-employed	-0.012	-0.421*	-0.603***	-0.399**	0.134	-0.052	-0.328	-0.074
	(0.198)	(0.245)	(0.148)	(0.179)	(0.255)	(0.263)	(0.338)	(0.172)
Household head government worker	0.050	-0.231	-0.370**	-0.121	-0.241	-0.016	-0.607	0.061
	(0.192)	(0.203)	(0.151)	(0.172)	(0.267)	(0.235)	(0.403)	(0.190)
Household head private wage worker	0.031	-0.192	-0.473***	-0.365**	0.248	-0.237	-0.147	0.341
	(0.156)	(0.187)	(0.157)	(0.161)	(0.227)	(0.247)	(0.332)	(0.312)
Household head: Other (retired, jobless …)	-0.036	-0.073	-0.043	-0.366*	-0.172	-0.036	-0.220	0.067
	(0.235)	(0.266)	(0.177)	(0.222)	(0.280)	(0.289)	(0.320)	(0.188)
Sex (1 for females)	0.104	-0.026	-0.072	-0.119	0.279	-0.124	-0.118	
	(0.066)	(0.088)	(0.086)	(0.116)	(0.212)	(0.170)	(0.171)	
Reported disease in 2003	0.254	-0.235*	-0.014	-0.130	0.022	0.129	-0.310*	
	(0.156)	(0.123)	(0.138)	(0.133)	(0.164)	(0.143)	(0.188)	
Own education		0.455	-0.027	-0.001	0.177**	-0.085	-0.023	
		(0.318)	(0.081)	(0.055)	(0.082)	(0.079)	(0.105)	
ref : Working in any sector		Yes	Yes					
ref : Agricultural worker				Yes	Yes	Yes	Yes	
Non-agricultural self-employed				0.004	-0.088	-0.391	0.274	
				(0.209)	(0.248)	(0.240)	(0.405)	
Government worker				0.322	-0.093	0.032	0.430	
				(0.350)	(0.320)	(0.247)	(0.418)	
Private employee				-0.180	-0.439	-0.170	-0.251	
				(0.206)	(0.299)	(0.345)	(0.387)	
Student		3.422***	0.686*	0.630***	0.333			
		(1.252)	(0.388)	(0.187)	(0.367)			
Considered too young or too old to work		2.934***	-0.030				-0.309	
		(1.137)	(0.482)				(0.416)	
Sick				0.920*	-1.084***	-0.007	-0.130	
				(0.539)	(0.387)	(0.445)	(0.442)	
Jobless, or missing info		2.253*	-0.064	0.330	0.038	-0.486	-0.186	
		(1.217)	(0.427)	(0.225)	(0.257)	(0.332)	(0.510)	
Rural * Non-agricultural self-employed				0.394	0.290	-0.210	-0.083	
				(0.524)	(0.403)	(0.473)	(0.528)	
Rural * Government worker					-0.751	-0.190	0.244	
					(0.841)	(0.317)	(0.532)	
Rural * Private employee				-0.575	-0.392	-0.049	0.251	
				(0.350)	(0.418)	(0.565)	(0.675)	
Rural * Student		-3.642***	-0.626	-0.771**	-0.054			
		(1.194)	(0.421)	(0.305)	(0.576)			
Rural * Considered too young or too old to work		-2.859**	-0.220				1.463**	
		(1.175)	(0.572)				(0.618)	
Rural * Sick				-0.547	1.024	-0.270	-0.279	
				(0.692)	(0.770)	(0.630)	(0.610)	
Rural * Jobless, or missing info		-2.341*	1.028*	-0.857**	-0.176		-0.524	
		(1.318)	(0.539)	(0.411)	(0.354)		(0.796)	
Head of the household			-0.956	0.400**	Ref.	Ref.	Ref.	
			(0.687)	(0.197)				
Spouse of the household's head			-0.426	0.799***	-0.110	0.117	0.310	
			(0.333)	(0.203)	(0.163)	(0.154)	(0.196)	
Son/daughter of the household head	Ref.	Ref.	Ref.	Ref.	-0.736**	-1.128**	-0.076	
					(0.296)	(0.460)	(0.716)	
Father/Mother of the household head						-0.295	-0.620*	
						(0.569)	(0.359)	
Brother/Sister of the household head	-0.710***	-0.727***	-0.475***	-0.459***	-1.161***	-1.309***	-0.894***	
	(0.168)	(0.247)	(0.109)	(0.106)	(0.280)	(0.310)	(0.278)	
Son/Daughter in law of the household head		-1.478*	-1.012*	0.035	-1.778**	-1.778**		
		(0.876)	(0.523)	(0.426)	(0.745)	(0.745)		
Basic EOPS Weights	-0.000***	-0.000***	-0.000**	-0.000***	-0.000***	0.000***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
EOPS Stratum dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2257	1673	1908	1714	1308	1417	1071	2229

Notes: The standard errors of the estimators are corrected for the correlation between different observations of the same enumeration area. These estimations use the “basic NPS weights”, unless reweighted. *** p<0.01, ** p<0.05, * p<0.1

Table 15: First stage predicting the inclusion in the regressions of Table 6

Lagged variable	
School enrollment rate of children aged 7-14	0.059
	(0.190)
Demographic characteristics of the household	
Number of household members in 2003	0.144***
	(0.023)
Age of the head in 2003	-0.012**
	(0.006)
Share of children (0-17) in the household	1.573***
	(0.389)
Poverty indicators	
Asset index	0.063***
	(0.019)
Mean education of the household members aged more than 25	-0.231
	(0.160)
The household head reported a recent disease	-0.018
	(0.116)
Occupation of the head (ref: agricultural worker)	
Household head non-agricultural self-employed	-0.292*
	(0.167)
Household head government worker	0.141
	(0.412)
Household head private wage worker	0.031
	(0.236)
Household head: Other (retired, jobless …)	-0.047
	(0.272)
Geography	
Maputo City Province	-0.238
	(0.270)
Maputo Province	-0.081
	(0.301)
Rural enumeration area	0.467***
	(0.098)
Central Region (ref: Northern)	-0.202
	(0.156)
Southern Region (ref: Northern)	-0.146
	(0.196)
Rural * Central Region (ref: Northern)	0.078
	(0.255)
Rural * Southern Region (ref: Northern)	-0.073
	(0.227)
Other	
Field team quality index	0.283**
	(0.138)
Observations	1199
R-squared	0.131

Notes: The standard errors of the estimators are corrected for the correlation between different observations of the same province which occurs with cluster sampling. These estimations use the “basic NPS weights”, unless reweighted. *** p<0.01, ** p<0.05, * p<0.1

APPENDIX 4: DESCRIPTION OF EXPLANATORY VARIABLES

Age: The age at the date of the interview is self declared.

Age of the household head:

Self-declared age of the household member reported as being the head.

Arcth(rho):

This variable has the same sign than the correlation between u_{it} and ε_{it} in equations (3) and (4).

Asset index:

This variable is an asset index à la Filmer and Pritchett (2001). The variables included in the principal component analysis include dwelling amenities (walls, roofs, access to tap water …) and the ownership of long term assets (car, television …).

Education of a person:

0: the person has never been to school

1: he/she has attended read/write classes

2: he/she has attended primary school 1

3: he/she has attended primary school 2

4: he/she has attended secondary school 1

5: he/she has attended secondary school 2

6: he/she has been to university

Household eligible for an interview during NPS:

Dummy taking value 1 if the household included for IAF a household member reportedly aged 0-17 in 2003 or enrolled at school at that date. It takes value 0 otherwise.

Field team quality index:

A subjective ranking of quality of field teams by World Bank staff based on field visits.

Mills ratio:

It is a first stage variable of Heckman correction for selection. It is the estimation of $E\left(u_{it} \mid X_{i,t-1}, s_i = 1\right)$ from equation (2) estimated with a probit model. The sign for its coefficient is therefore the sign of the correlation between u_{it} and ε_{it}.

Mean education of the household members aged more than 25:

Mean of the "Education of a person" variable among household members aged more than 25. Replaced by the education of the 2 eldest household members whenever there is no household member aged more than 25.

Number of household members: Number of respondents in the household

Occupation of the individual:

The main occupation of the individual has 8 modalities:

- Agricultural worker
- Self-employed non agricultural
- Government worker
- Private sector wage earner
- Education
- Too young or too old to have an occupation
- Non-activity due to illness
- Others

This variable is available for all the individuals aged 5 or more in each household. For consistency, the last four modalities are grouped for adults in most specifications. For children, all modalities for economic activity are grouped (Agricultural worker, Self-employed non agricultural, Government worker and Private sector wage earner), and the last 2 modalities are grouped (Illness and others).

Reported disease:

The question asks whether the household member was ill or injured during the last 2 weeks. It takes value 1 if a disease is reported, 0 otherwise.

Sex: takes value 1 for females and 0 for males.

School enrollment rate of children aged 7-14: Share of the children aged 7-14 enrolled at school during the current school year. This variable is not available for the households with no child children aged 7-14. In regression Tables 4, 5 10 and 11, it is arbitrarily set to 0 for these

households, and the regressions include the dummy “There is a child of school age (7-14) in the household”.

Share of children (0-17) in the household: Ratio of the number of respondents reportedly aged 0-17 on the total number of respondents in the household.

The household head reported a recent disease:

Reported disease variable for the household head

There is a child of school age (7-14) in the household:

This dummy takes value 1 if one of the household members reported an age between 7 and 14 y.o. and 0 otherwise.