

# **POVERTY MEASUREMENT IN TAJIKISTAN: METHODOLOGICAL ISSUES AND PRELIMINARY ESTIMATES**

Victor Sulla  
(The World Bank)

Giovanni Vecchi<sup>1</sup>  
(University of Rome “Tor Vergata”)

This draft: January, 2013

**Preliminary – Not to be quoted without the Authors’ permission**

## **Abstract**

The paper estimates inequality and poverty rates in Tajikistan using two quarters of the 2012 Household Budget Survey. In constructing a monetary welfare indicator we deal with two common, yet complicated data problems, namely a badly defined household identifier and the availability of partial information on households’ food consumption. We estimate a set of poverty lines based on the cost-of-basic-need method, and produce a timely and updated assessment of the poverty situation in Tajikistan. While results are preliminary, they are statistically robust. In 2012, more than one third of the Tajikistani population is classified as poor, while 14% face extreme poverty. When compared to previous estimates, our findings suggest that during the last three-four years poverty rates in Tajikistan might have declined at a rapid pace. Further analysis is required to identify the poverty time trend, however, as both data and methods used in previous studies are too heterogeneous to carry out poverty comparisons with any confidence.

**Keywords:** living standards; consumption aggregate; cost-of-basic-needs method; poverty; inequality; spatial deflators; Tajikistan.

**JEL classification:** I32, D31.

---

<sup>1</sup> We thank Nicola Amendola (University of Rome “Tor Vergata”), Nobuo Yoshida (The World Bank) and Umar Serajuddin (The World Bank) for useful comments. We also thank Maka Lomaia (The World Bank) for superb research assistance.

## 1 INTRODUCTION

In this note we illustrate the main steps involved in measuring poverty in Tajikistan on the basis of the 2012 Household Budget Survey (HBS). The survey – run by the National Statistical Committee (GosKomStat) on a yearly basis – visits each of 3,000 households four times in a year, with one third of the sample interviewed in January, April, July and October, one third in February, May, August and November, and one third in March, June, September and December. In this paper we use data from the 2<sup>nd</sup> and 3<sup>rd</sup> quarters.

The first task we accomplish is the construction of a monetary *welfare indicator* (section 2). Next, we estimate a *spatial survey-based price index* (SPI, section 2.5), required to adjust for the differences in the cost-of-living across the territory. In Section 3 we estimate the *inequality* of the distribution of living standards. In Section 4, we produce a set of absolute *poverty lines* using the so-called Cost-of-Basic-Needs (CBN) method [Ravallion 2012: 83]. Finally, we estimate a selection of *poverty indicators* both at the national and regional level (Section 5).

The empirical findings presented in sections 4 and 5 are preliminary. This is due to the fact that the analysis has focused on the development of a suitable methodology to carry out HBS-based poverty analysis in Tajikistan, while most of previous studies have used LSMS data [World Bank 2009]. Secondly, considerable effort was put on checking data quality issues. Thirdly, while we have used only two quarters of the 2012 HBS, no attempt was made to adjust for seasonal effects. Nor have we taken any action to solve potential small sample size problems (*e.g.*, urban GBAO only counts 60 observations per quarter).

A few results are worth mentioning here. Firstly, 2012 HBS data do not seem to suffer from any specific problems. After exploring a number of variables, we found no major anomalies to report (but see section 2.1 and Appendix E); we conclude that HBS data represent a viable alternative to LSMS data as far as poverty measurement in Tajikistan is concerned. Secondly, while the information available in the HBS questionnaire allows the analyst to construct a reliable welfare indicator, margins of improvement are large. The comprehensiveness of the consumption aggregate is severely limited by the absence of important components (housing, durable goods, and other non-food expenditures) and by the fact that food expenditures only refer to purchased quantities, and not to consumed quantities. Thirdly, the estimated *food poverty line* equals 97.9 Somoni/person/month (this is the minimum cost of purchasing 2,250 kcal/person/day and amounts to 48.7% of

average per capita expenditure). In 2012, 14% of the Tajikistani population could not afford the food poverty line, and is classified as “extremely poor”. Fourthly, following Ravallion and Bidani (1994), we estimated two additional poverty lines: (a) a *lower poverty line*, which equals 119.2 Somoni/person/month (59.3% of average per capita expenditure), and (b) an *upper poverty line*, which equals 134.2 Somoni/person/month (66.9% of average per capita expenditure). In 2012, 26.2% of the population fell below the lower poverty line, while 33.8% fell below the upper poverty line. The urban-rural divide is narrow, while regional variation in poverty rates is substantial (Table 1).

From a methodological standpoint, poverty lines estimates are sensitive to a number of discretionary choices made by the analyst, which affect both the overall “level” of poverty (how many poor are in the country) and its structure (the so called “poverty profile”). In the context of Tajikistan, we find that the single most important parameter affecting the poverty estimates is the *unit cost of calories*, which, in turn, crucially depends on the choice of the “reference group” (that is, the set of households over which the cost of one calorie is calculated). The “referencing problem” is a critical issue that deserves further investigation and discussion.

**Table 1 – Selected poverty and inequality estimates**

	Dushanbe	Sogd	Khatlon	RRP	GBAO	Urban	Rural	Tajikistan
Population (%)	10.8	25.9	31.3	23.6	8.4	33.3	66.7	100.0
Per capita expenditure (Somoni/person/month)	234	253	186	172	135	213	195	201
<b>INEQUALITY</b>								
Gini index (%)	24.9	36.2	29.7	29.6	22.2	31.6	32.2	32.1
<b>FOOD POVERTY LINE (97.9 Somoni/person/month)</b>								
Headcount ratio (%)	1.5	6.7	16.4	20.4	25.5	10.1	15.9	14.0
Distribution of the poor (%)	1.2	12.4	36.8	34.4	15.3	24.0	76.0	100.0
Poverty Gap (%)	0.2	1.1	3.2	3.6	5.5	1.7	3.1	2.6
Poverty Gap Squared (%)	0.0	0.3	0.9	1.0	1.8	0.5	0.9	0.8
<b>LOWER POVERTY LINE (119.2 Somoni/person/month)</b>								
Headcount ratio (%)	7.9	15.7	29.8	34.1	46.3	21.2	28.7	26.2
Distribution of the poor (%)	3.3	15.5	35.7	30.7	14.8	27.0	73.0	100.0
Poverty Gap (%)	1.0	2.9	6.8	7.9	10.9	4.2	6.5	5.8
Poverty Gap Squared (%)	0.2	0.8	2.2	2.5	3.9	1.3	2.2	1.9
<b>UPPER POVERTY LINE (134.2 Somoni/person/month)</b>								
Headcount ratio (%)	12.9	22.5	36.9	43.6	57.3	29.3	36.1	33.8
Distribution of the poor (%)	4.1	17.2	34.1	30.4	14.2	28.8	71.2	100.0
Poverty Gap (%)	2.1	4.7	9.8	11.4	15.5	6.6	9.4	8.5
Poverty Gap Squared (%)	0.5	1.5	3.5	4.0	5.9	2.1	3.4	3.0

Source: Authors' estimates on HBS 2012, 2<sup>nd</sup> and 3<sup>rd</sup> quarter data.

A final remark concerns the comparability of the estimates presented in this note with previous estimates. Briefly, our recommendation is to resist to compare LSMS-based estimated with HBS-based estimates, due to differences in survey design, the definition of the welfare aggregate, and estimation methods.

## 2 THE WELFARE INDICATOR

Schematically, a monetary welfare indicator based on household expenditures can be defined as follows:

$$\underbrace{\text{Welfare indicator}}_{\text{sect. 2.5}} = \frac{\overbrace{\text{nominal household expenditure}}^{\text{sect. 2.3}}}{\underbrace{\text{household equivalent size}}_{\text{sect. 2.1}} \times \underbrace{\text{temporal CPI}}_{\text{sect. 2.2}} \times \underbrace{\text{spatial CPI}}_{\text{sect. 2.4}}}$$

For each *individual*, the welfare indicator is defined as the ratio between the nominal household expenditure, and the product of three factors required to adjust for:

- (i) *differences in needs* caused by different demographic structures of the households and/or the presence of *economies of scale* that may be available to larger households (section 2.1);
- (ii) *inflation* (section 2.2);
- (iii) *cost-of-living differences* across geographical areas (section 2.4);

In what follows we describe each step in detail.

### 2.1 HOUSEHOLD EQUIVALENT SIZE

While the HBS collects expenditures at the *household* level (as most surveys around the world do), we are ultimately interested in *individual* welfare. As noted by Deaton and Zaidi (2002: 48): “If it were possible to gather data on consumption by individual family members, we could move directly from the data to individual welfare, but except for a few goods, such data are not available, even conceptually – think of public goods that are shared by all household members. As it is, the best that can be done is to adjust total household expenditure by some measure of the number of people in the household, and to assign the resulting welfare measure to each household member as an individual”. Thus,

household expenditures need to be adjusted for household *size* (large households can take advantage from sharing certain goods, that is, can exploit economies of scale) and/or household *composition* (two households of the same size may have different needs due to different demographic structures).

The simplest adjustment consists in dividing the total household expenditure by the number of household members, thereby obtaining per capita expenditure (PCE). This is clearly a partial solution, because economies of scale are taken care of, but no adjustment is made for differences in needs. Nevertheless, the use of PCE is both a common and best practice, in line with the recommendation by Deaton and Zaidi (2002: 49): “no calculation of welfare or poverty profile should ever be done without the calculation of per capita expenditure as at least one of the alternatives”. Consistently, in what follows *we will use the consumption aggregate expressed in per capita terms*<sup>2</sup>.

The HBS visits households four times during the year [Sulla and Muñoz, 2011]. As a consequence, household size tends not to remain constant across quarters. Some of the observed variation is genuine (household composition changes during the survey year for a number of perfectly legitimate reasons: births and deaths, marriages and separations, ins and outs of the family life, etc.), some of it is likely to be due to reporting errors. In the absence of additional information is impossible to tell the relative importance of genuine variation versus bad measurement. Table 2 shows the extent that household size varies between the second and third quarter in the 2012 HBS. At the national level, household size does not change for 85.5% of total households, while a unit *increase* is recorded in 5% and a unit *decrease* in 4% of households (see the last column of the table). Regional variation is more pronounced. The case of Dushanbe is, in fact, worth highlighting: in this region one third of households undergoes a change in size between the two quarters.

---

<sup>2</sup> While deserving a place in the agenda of future analysis, the use of *equivalence scales* and/or other adjustments for economies of scale does not seem a priority at this stage. As explained in Section 2.3.2, the consumption aggregate does not include housing expenditures (more precisely, it does not include the imputed/self-reported rental value of the house), which is the most important “public good” within the household, nor does it include durable goods.

**Table 2 – Differences in household size between the 2<sup>nd</sup>-3<sup>rd</sup> quarters (%)**

Q3 minus Q2	Dushanbe	Sogd	KHatlon	RRP	GBAO	Rural	Urban	Tajikistan
-11	0.3	0.0	0.0	0.0	0.0	0.0	0.1	0.0
-7	0.0	0.0	0.0	0.7	0.0	0.2	0.0	0.1
-6	0.0	0.0	0.0	0.5	0.0	0.1	0.1	0.1
-5	0.3	0.1	0.0	0.0	0.0	0.1	0.1	0.1
-4	1.5	0.0	0.0	0.5	0.8	0.3	0.5	0.4
-3	1.8	0.2	0.6	0.3	0.0	0.3	1.0	0.5
-2	4.5	0.0	0.6	1.2	1.7	0.8	1.7	1.1
-1	6.8	2.7	3.0	4.7	6.7	3.9	4.3	4.0
0	66.8	92.7	89.2	83.5	81.7	87.7	81.8	85.5
+1	7.3	4.2	4.0	6.0	5.8	4.4	6.1	5.0
+2	3.3	0.1	1.4	1.2	1.7	1.2	1.4	1.3
+3	3.5	0.0	0.3	0.5	1.7	0.4	1.4	0.8
+4	2.0	0.0	0.3	0.3	0.0	0.2	0.8	0.4
+5	0.8	0.0	0.3	0.0	0.0	0.2	0.3	0.2
+6	0.3	0.0	0.1	0.0	0.0	0.1	0.1	0.1
+7	0.5	0.0	0.0	0.0	0.0	0.0	0.2	0.1
+8	0.5	0.0	0.1	0.2	0.0	0.1	0.2	0.1
+9	0.3	0.0	0.0	0.3	0.0	0.1	0.1	0.1
+12	0.0	0.0	0.0	0.2	0.0	0.1	0.0	0.0
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: In column 1, zero means no change in household size between the 2<sup>nd</sup> and 3<sup>rd</sup> quarter, “+1” means that household size in q3 had one more member than it had in q2, while “-1” means that household size has decreased by one unit between quarters 2 and 3. Source: Auhtors’ calculation on HBS 2012, 2<sup>nd</sup> and 3<sup>rd</sup> quarters.

The fact that the household size varies by quarter poses a couple of problems when it comes to calculating the *yearly* expenditure in *per capita* terms. First, per capita expenditures must be calculated separately by quarter, then aggregated by adding up quarterly expenditures, and finally annualized by multiplying the results by two: by so doing, one obtains correct measures of yearly per capita expenditures for each household. For households with different sizes in the two quarters there is also a problem in assigning them a single value to be used as *the* household size. This is an important requirement for poverty analysis.

One solution consists in assigning each household the *modal household size*, *i.e.* the most frequent size observed over the quarters. In the present context, however, with only two quarters available, the use of the mode would amount to using either the minimum or the maximum household size. This is clearly an unsatisfactory solution. Another solution consists in using the *mean size*, that is the mean value (rounded to the closest integer) between the household sizes observed in each quarter. Both solutions are suboptimal. The use of either the minimum or the maximum size implies that in one of the two quarters we would systematically rely on a wrong household size. On the other hand, the use of the mean size implies that we would *always* use an incorrect household size. It may be argued that some inaccuracy can be accepted, provided that the choice of one method

over the other makes a little difference, empirically. Unfortunately, this does not seem to hold in the case of the Tajikistani data. Table 3 shows the frequency distribution of household size by region and urban areas after assigning households with different across-quarter values in household size (*a*) the lowest value, (*b*) the highest value and (*c*) the mean value. The table also compares HBS 2012 with the LSMS 2009 estimates<sup>3</sup>. Clearly, whether using the lowest or the highest or the mean value has a significant impact on the sample distribution of household size<sup>4</sup>.

**Table 3 – Household size by region and urban-rural area**

	HBS 2012 (quarters 2 and 3)			HBS 2012 (quarters 2 and 3)			HBS 2012 (quarters 2 and 3)			LSMS 2009 (4 quarters)		
	MIN			MAX			MEAN					
Region	Rural	Urban	Total	Rural	Urban	Total	Rural	Urban	Total	Rural	Urban	Total
Dushanbe		<b>4.8</b>	<b>4.8</b>		<b>5.6</b>	<b>5.6</b>		<b>5.3</b>	<b>5.3</b>		<b>5.3</b>	<b>5.3</b>
		<i>400</i>	<i>400</i>		<i>400</i>	<i>400</i>		<i>400</i>	<i>400</i>		<i>270</i>	<i>270</i>
Sogd	<b>5.9</b>	<b>5.5</b>	<b>5.7</b>	<b>6.0</b>	<b>5.6</b>	<b>5.8</b>	<b>6.0</b>	<b>5.6</b>	<b>5.8</b>	<b>6.4</b>	<b>5.1</b>	<b>6.0</b>
	<i>530</i>	<i>330</i>	<i>860</i>	<i>530</i>	<i>330</i>	<i>860</i>	<i>530</i>	<i>330</i>	<i>860</i>	<i>261</i>	<i>135</i>	<i>396</i>
Khatlon	<b>7.0</b>	<b>5.6</b>	<b>6.6</b>	<b>7.2</b>	<b>5.8</b>	<b>6.8</b>	<b>7.1</b>	<b>5.7</b>	<b>6.7</b>	<b>7.7</b>	<b>6.9</b>	<b>7.6</b>
	<i>660</i>	<i>240</i>	<i>900</i>	<i>660</i>	<i>240</i>	<i>900</i>	<i>660</i>	<i>240</i>	<i>900</i>	<i>315</i>	<i>63</i>	<i>378</i>
RRP	<b>7.6</b>	<b>6.3</b>	<b>7.4</b>	<b>8.0</b>	<b>6.5</b>	<b>7.7</b>	<b>7.9</b>	<b>6.5</b>	<b>7.6</b>	<b>8.3</b>	<b>6.5</b>	<b>8.0</b>
	<i>480</i>	<i>120</i>	<i>600</i>	<i>480</i>	<i>120</i>	<i>600</i>	<i>480</i>	<i>120</i>	<i>600</i>	<i>261</i>	<i>54</i>	<i>315</i>
GBAO	<b>6.8</b>	<b>5.9</b>	<b>6.6</b>	<b>7.2</b>	<b>6.0</b>	<b>6.9</b>	<b>7.1</b>	<b>6.0</b>	<b>6.8</b>	<b>6.5</b>	<b>4.1</b>	<b>6.2</b>
	<i>180</i>	<i>60</i>	<i>240</i>	<i>180</i>	<i>60</i>	<i>240</i>	<i>180</i>	<i>60</i>	<i>240</i>	<i>126</i>	<i>18</i>	<i>144</i>
Tajikistan	<b>6.8</b>	<b>5.4</b>	<b>6.3</b>	<b>7.0</b>	<b>5.7</b>	<b>6.5</b>	<b>7.0</b>	<b>5.6</b>	<b>6.5</b>	<b>7.4</b>	<b>5.5</b>	<b>6.7</b>
	<i>1,850</i>	<i>1,150</i>	<i>3,000</i>	<i>1,850</i>	<i>1,150</i>	<i>3,000</i>	<i>1,850</i>	<i>1,150</i>	<i>3,000</i>	<i>963</i>	<i>540</i>	<i>1,503</i>

*Note:* each cell contains the average household size (top number, in bold) and the number of observations (bottom number, in italics). “Min” means that households with different sizes in Q2 and Q3 have been assigned the minimum value; under the “Max” hypothesis, the maximum value has been used, while “Mean” denotes the case where we use the mean household size (rounded up to the closest integer).

In the rest of the analysis we explore a different approach. Instead of treating the quarterly observations as longitudinal data, we treat them as repeated cross-sections. The rationale for this choice stems from the “high” across-quarter variation in household size shown in Table 2, which suggests that something might have gone wrong with household identifiers. There is no need to speculate on the mechanisms that might be responsible for this phenomenon: sometimes interviewers replace households with “similar” ones and assign new households the same identifiers as the old households. Other problems might occur at the data management level. These circumstances cannot be excluded *a priori* and

<sup>3</sup> In this document we assume that the sample is self-weighted. Household-level estimates are therefore carried out without the need to specify a weight, while individual-level estimates use household size as a weighting variable.

<sup>4</sup> See Appendix A for an assessment of poverty rates sensitivity to different methods for calculating the household size (Table 3).

are worth consideration. If the accuracy of the household identifier is questioned, then, no matter the cause, one had better give up the usual strategy to *merge* households across quarters, and instead *append* them quarter by quarter. Observations can be merged only when there is confidence that the identifier is indeed an identifier. Given the evidence in Table 2 e Table 3 we decided to pool observations from different quarters and treat them as if they originated from independent quarterly samples. Table 4 shows the household size distribution separately by quarter and for the pooled sample.

**Table 4 – The distribution of household size by region and urban-rural area**

Region	2nd QUARTER			3rd QUARTER			2nd and 3rd QUARTERS		
	Rural	Urban	Total	Rural	Urban	Total	Rural	Urban	Total
Dushanbe		5.1 400	5.1 400		5.3 400	5.3 400		5.2 800	5.2 800
Sogd	5.9 530	5.5 330	5.8 860	5.9 530	5.5 330	5.8 860	5.9 1060	5.5 660	5.8 1720
Khatlon	7.0 660	5.6 240	6.7 900	7.1 660	5.7 240	6.7 900	7.1 1320	5.7 480	6.7 1800
RRP	7.8 480	6.4 120	7.5 600	7.8 480	6.5 120	7.5 600	7.8 960	6.4 240	7.5 1200
GBAO	7.0 180	5.9 60	6.7 240	7.0 180	6.0 60	6.7 240	7.0 360	5.9 120	6.7 480
Tajikistan	6.9 1,850	5.5 1,150	6.4 3,000	6.9 1,850	5.6 1,150	6.4 3,000	6.9 3,700	5.6 2,300	6.4 6,000

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

## 2.2 TEMPORAL CONSUMER PRICE INDEX

An adjustment for within-the-year inflation is required because the available data span six months of 2012. During this period prices have varied, and with them the households' purchasing power. Thus, prior to using monetary variables for the construction of the welfare aggregate, we deflate all monetary values using the official CPI (Table 5).

Given the sampling scheme described in Sulla and Muñoz (2011), the ideal solution to deflate expenditures would use *monthly* CPIs. The HBS datasets, however, do not contain the information on the month of the interview. For this reason we have deflated household expenditures using *quarterly* CPIs (calculated as a simple average of monthly total CPIs in Table 5).



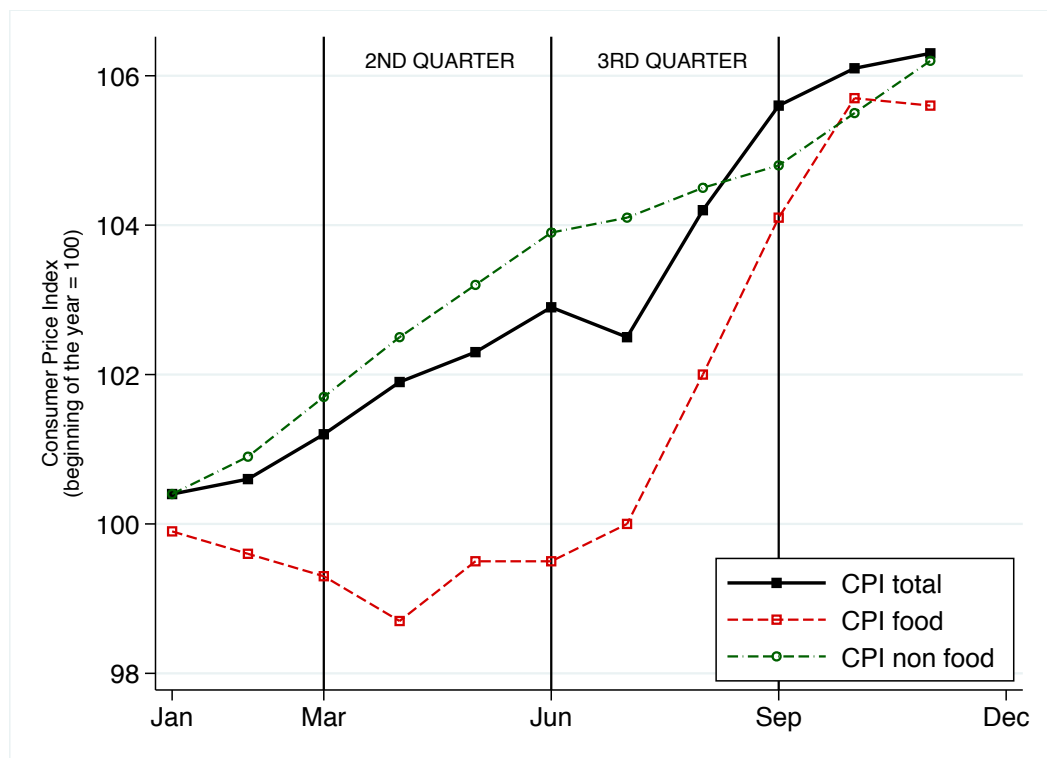
**Table 5 – Official Consumer Price Index (beginning of the year 2012 = 100)**

month	CPI total	CPI food	CPI non food
Jan	100.4	99.9	100.4
Feb	100.6	99.6	100.9
Mar	101.2	99.3	101.7
Apr	101.9	98.7	102.5
May	102.3	99.5	103.2
Jun	102.9	99.5	103.9
Jul	102.5	100.0	104.1
Aug	104.2	102.0	104.5
Sep	105.6	104.1	104.8
Oct	106.1	105.7	105.5
Nov	106.3	105.6	106.2

Source: Tajikistan Central Bank (2012).

Figure 1 shows that the dynamics of different price indices is significantly different between Q2 and Q3. We have used the total CPI, but one might consider using the food CPI, in which case the data in the 3<sup>rd</sup> quarter would be higher in real terms. As far as inequality and poverty analysis is concerned, however, the choice would by and large lead to the same estimates. This is guaranteed by the fact the most inequality indices are scale invariant, while most poverty indicators are homogeneous of degree zero in the poverty line and the consumption aggregate.

**Figure 1 – Official Consumer Price Indices (beginning of the year 2012 = 100)**



Source: Tajikistan Central Bank (2012).

## 2.3 THE NOMINAL CONSUMPTION AGGREGATE

The information collected in the HBS questionnaire makes it possible to construct (at least) two welfare indicators based on consumption expenditures:

- 1) one possibility is to use the expenditures for *purchased* goods and services;
- 2) a second possibility is to exploit the information on *actual consumption* of goods and services.

The advantage associated with option 2) is that actual consumption tends to fluctuate less than out-of-pocket expenditures (which are affected by seasonal and cyclical factors to a greater extent). If households have some capacity to smooth their consumption over time, option 2) qualifies as a more appropriate indicator of *permanent* (or average) living standards than option 1). Accordingly, *we opt for a welfare indicator based on actual consumption*.

The nominal consumption aggregate can be expressed as the sum of two sub-aggregates, namely food expenditures and non-food expenditures. Notation, convenient for later use, is as follows:

$$THE_h = \underbrace{FHE_h}_{\text{sect. 2.3.1}} + \underbrace{NFHE_h}_{\text{sect. 2.3.2}}$$

where “THE” is for “total household expenditure”, “FHE” is for “food household expenditure”, and “NFHE” is for “non-food household expenditure”. The suffix  $h$  denotes the  $h$ -th household, with  $h = 1, \dots, H$ .

### 2.3.1 Food consumption ( $FHE_h$ )

The HBS survey collects information on a) *purchased* quantities of foods (form 1, section 4), b) expenditures for purchased foods (form 1, section 4), and c) *consumed* quantities (form 2, section 1) during the last 30 days for 71 food items. However, while the questionnaire collects expenditures associated with purchased food quantities, it does not provide information on the value of consumed expenditures. Given our preference for measuring welfare based on actual consumption rather than on purchases this poses a problem: How to value the actual household consumption? The solution described in this section exploits the information available in the questionnaire to scale up or down purchases by means of a factor accounting for the discrepancy between purchased and actually consumed quantities.

We define an adjustment coefficient  $R_j^h$  for each household ( $h$ ) and for each food item ( $j$ ):

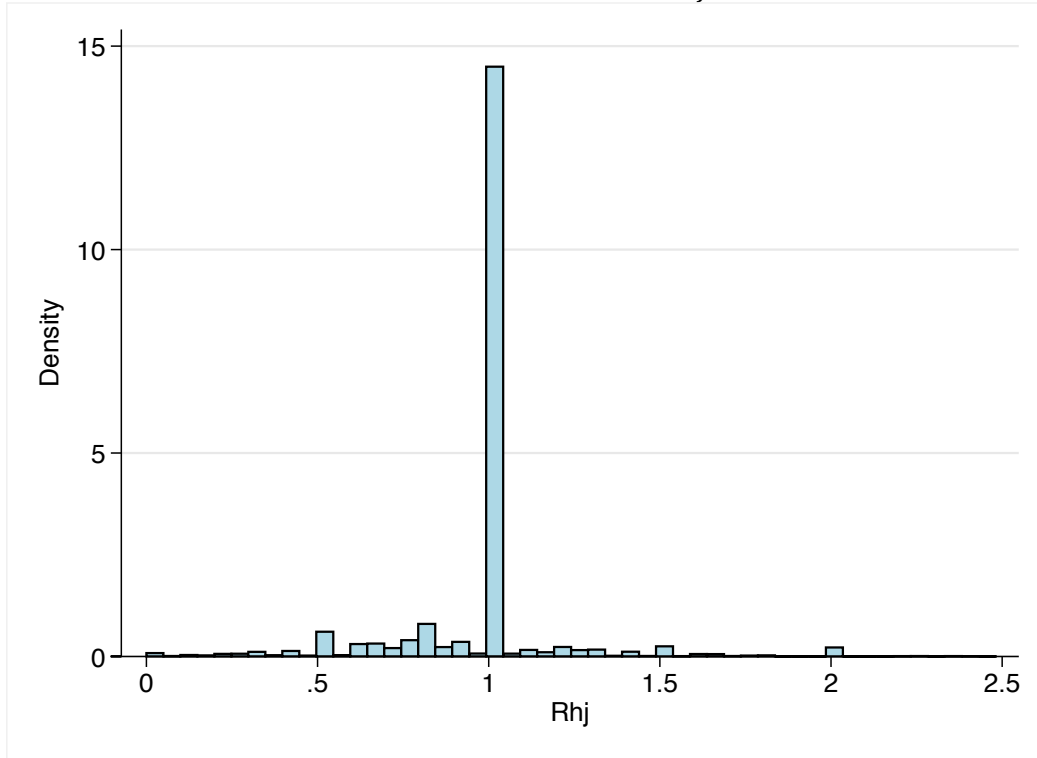
$$(1) \quad R_j^h = \frac{Q_j^h(cons)}{\widehat{Q}_j^h(pur)}$$

The coefficient  $R_j^h$  is the ratio between  $Q_j^h(cons)$ , the quantity of item  $j$  *actually* consumed by the  $h$ -th household (form 2, section 1, question 9) and  $\widehat{Q}_j^h(pur)$ , the *purchased* quantity (form 1, section 4, question 2) by the same household. In eq. 1 we use the hat to stress the fact that the data provide us with the *estimated* value of the purchased quantity  $Q_j^h(pur)$ . While  $\widehat{Q}_j^h(pur)$  is observed,  $Q_j^h(pur)$  is not, and the difference between the two can be attributed to a number of factors, including measurement error, misreporting and other<sup>5</sup>. The interpretation of  $R_j^h$  is straightforward. In equation 1, the condition  $R_j^h < 1$  implies that household  $h$  consumes less of product  $j$  than its monthly purchases. Vice versa, the condition  $R_j^h > 1$  indicates that the household's consumption during the reference period has exceeded its purchases (perhaps consuming self-production or running down the stock). Before illustrating how we use of this factor, we show its empirical distribution (Figure 2).

---

<sup>5</sup> This distinction might be of some relevance, in the context of Tajikistan, as the HBS collects information on these two variables in different sections of the questionnaire, with no cross-checks in the field, nor at the data entry stage.

**Figure 2 – The adjustment factor  $R_j^h$  (eq. 1)**



The histogram shows that the majority of households do not accumulate or deplete their food stocks ( $R_j^h = 1$ ), while a not negligible share of households consume less than their monthly purchasing. It is worth reminding that the data only refers to two quarters of 2012 and the distribution of  $R_j^h$  might be affected by seasonal factors. Note: values of  $R_j^h$  exceeding 2.5 (less than 1% of total observations) have been omitted from the graph. Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

The adjustment factor  $R_j^h$  is a useful device to scale upward or downward, as required, the expenditures on food purchases. To illustrate, let us write the total food expenditure for the  $h$ -th household as follows:

$$(2) \quad \widehat{FHE}_h = \sum_j \widehat{FHE}_j^h = \sum_j p_j^h Q_j^h(pur) R_j^h$$

where  $p_j^h$  is the price paid for item  $j$  by household  $h$ , and  $Q_j^h(pur)$  denotes the purchased quantity of item  $j$  by household  $h$ . In the absence of  $R_j^h$  the only option would be to use the purchases on food. With  $R_j^h$  we can exploit the information on the quantities actually consumed by the household. To illustrate, suppose that a household reports spending 1,000 Somoni for the purchase of 1.5 kilograms of beef. Also suppose that the same household reports consuming 3 kilograms of beef (remember that the questionnaire does

not ask the *value* of consumed quantities). Our procedure – as described in eq. 2 – calculates the ratio between consumed and purchased quantities ( $R_j^h = 3/1.5=2$ ) and use the result to rescale the observed expenditure for purchases ( $\widehat{FHE}_j^h = 1,000 \times 2 = 2,000$ ). Thus, in our example, the household is assigned an expenditure for beef equal to 2,000 instead of 1,000.

The above-described procedure has a number of advantages. Firstly, it produces a more accurate estimate of actual consumption than that used in previous reports, where only expenditures on purchased goods were used. Secondly, the estimated  $\widehat{FHE}_h$  does not depend on *unit values*. While playing a key role in estimating the value of in-kind consumption, unit values are prone to measurement errors – see Deaton (1997).

To further investigate the pros and cons associated with the proposed method let us replace eq. 1 in eq. 2:

$$(3) \quad \widehat{FHE}_h = \sum_j p_j^h Q_j^h(pur) \frac{Q_j^h(cons)}{\widehat{Q}_j^h(pur)} = \sum_j p_j^h Q_j^h(cons) \frac{Q_j^h(pur)}{\widehat{Q}_j^h(pur)}$$

If there were no difference between  $Q_j^h(pur)$  and its empirical/observed counterpart  $\widehat{Q}_j^h(pur)$ , then eq. 3 would deliver  $\widehat{FHE}_h = \sum_j p_j^h Q_j^h(cons)$ , the value of consumed quantities. What eq. 3 shows is that the adjustment factor  $R_j^h$  defined in eq. 1 does a good job in estimating actual household consumption provided the inaccuracy in reporting purchased quantities is small in size and/or errors are randomly distributed across products and households.

The practical implementation of eq. 3 requires an additional step. Note that  $\widehat{FHE}_h$  is obtained as the sum over all food items  $\widehat{FHE}_j^h$ . These terms can be estimated if and only if, in addition to expenditures for purchases, *both* purchased *and* actually consumed quantities are available. When this is not the case we have implemented the following rule:

$$(5) \quad \widehat{FHE}_h^j = \begin{cases} p_j^h Q_j^h(pur) R_j^h & \text{if } p_j^h Q_j^h(pur) \text{ and } R_j^h \text{ are not missing} \\ p_j^h Q_j^h(pur) & \text{if } p_j^h Q_j^h(pur) \text{ is not missing and } R_j^h \text{ is missing} \\ uv_j(h) Q_j^h(cons) & \text{if } p_j^h Q_j^h(pur) \text{ is missing} \end{cases}$$

where  $uv_j(h) = \text{median}(uv_j^h)$  is the median unit value for commodity  $j$ . The calculation of unit values in eq. 5 has followed a 3-level hierarchical procedure: the highest priority was given to median unit values calculated across the five regions (*oblasts*), separately by urban-rural areas. Next, unit values have been calculated at the regional level, and finally at the national level.

Table 6 shows average monthly expenditure (FHE in equation 2) by region and urban rural areas (left panel) and compares it with the expenditure for purchases (right panel). All estimates are in per capita terms. At the national level consumption expenditures are 22% higher than expenditures for purchases (126 versus 103 Somoni/person/month). The adjustment for sub-national aggregates is more substantial: the use of consumption slightly decreases expenditures in urban areas (minus 2%, on average, with respect to purchases), while it increases expenditures in rural areas (plus 40%, on average). Regional variation is also not uniform across regions: rural Khatlon +60%, compares to -5% in urban Sogd. Overall, the pattern in Table 6 is in line with expectations (actual consumption among rural households tend to be higher than purchases), a fact that supports the method described in this section.

**Table 6 – Food expenditure (Somoni/person/month)**

Region	CONSUMPTION (FHE <sub>h</sub> )			PURCHASES		
	Rural	Urban	Total	Rural	Urban	Total
Dushanbe		149	149		153	153
Sogd	123	117	121	98	123	107
Khatlon	139	107	132	87	108	91
RRP	115	139	119	85	134	94
GBAO	104	135	111	83	141	95
Tajikistan	125	129	126	89	132	103

Note: values are have been deflated with a temporal CPI but have *not* been deflated with the *spatial* CPI.

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

### 2.3.2 Non-Food Consumption

Household expenditures for non-food items can be defined as the sum of three main sub-aggregates:

$$NFHE_h = NFE_h + DURABLES_h + HOUSING_h$$

where  $NFE_h$  identifies all expenditures for non-durable non-food commodities and services,  $DURABLES_h$  is the consumption flow out of durable goods owned by the

household, and  $HOUSING_h$  is the monthly expenditure on housing. The 2012 HBS collects information on all these three sub-aggregates but, as we will see, only the first sub-aggregate ( $NFE_h$ ) can be reliably estimated.

## Non-food non-durable commodities and services

In the 2012 HBS questionnaire, sections 5, 6, 7 and 8 in Form 1 contain monthly expenditures on non-food commodities and services at the household level.

*Section 5 (“Clothing, hosiery, fabrics and shoes”).* All expenditures in this section have been included in the consumption aggregate.

*Sections 6 (“Furniture, cultural and home items”) and 7 (“Purchase of soap, synthetic detergents, ...”).* These sections contain expenditures on furniture, cultural and home items, soap, detergents, fuel, tobacco products, fodder for live-stocks and birdseeds for poultry, etc.. The total value of household consumption has been calculated by summing up purchases and values of received gift, and subtracting the value of commodities given away by the household as a gift (Deaton and Zaidi, 2002). We excluded all the expenditures for production (fodder for live-stocks, birdseeds for poultry, medicines for animals and the like).

*Section 8 (“Monetary expenses of payment for personal and production services”).* This module contains expenditures on actual rent, utilities, repairs, household services, cultural and educational services, transport services, health services and production services. All these expenditures have been included in the consumption aggregate with the exception of (1) production services and burying services (form 1, section 8, question 8010, variables 1-7, and question 804 variable 7), and (2) actual rent (Form 1, section 8, question 801, variable 1).

## Housing

The housing component of the consumption aggregate should measure the flow of services that households receive from occupying their dwelling and express it in monetary terms. The HBS questionnaire only collects this information for renters (question 801 of section 8, Form 1). Yet, HBS 2012 data show that in Tajikistan the largest majority of households own their houses. This implies that for most households in the survey it is necessary to estimate the *implicit rental value*, that is, the value that household owners should pay if they had to rent their own house.

Unfortunately, the HBS questionnaire does not ask households to estimate the rental values of their homes. On the other hand, small sample size problems prevent us from estimating the rental value by means of a *hedonic regression*. This method requires to use data on housing characteristics, estimate rental equivalences on the basis of a sample of renters, and finally estimate the rental value of homeowners housing units by means of the predicted value calculated using the estimated parameters of the regression (Citro and Michael, 1995: 246). Due to lack of suitable data, *we decided not to include housing expenditures in the welfare aggregate*. The same strategy was adopted in the LSMS poverty assessment report.

## Consumer Durable Goods

Form 6 of the HBS questionnaire collects information on durable goods. Households are asked to provide information on availability of durables and the age of the durables. There is no information, however, on the market value of these commodities nor are there estimates of their values. *The available information does not allow us to estimate the consumption flow from durables.*

Table 7 shows the distribution of non-food expenditures (NFHE) by region and rural/areas, and compares them with food and total expenditures. The Sogd region stands out for relatively high non-food expenditures.

**Table 7 – Household expenditures (Somoni/person/month)**

Region	FOOD			NONFOOD			TOTAL		
	Rural	Urban	Total	Rural	Urban	Total	Rural	Urban	Total
Dushanbe		149	149		102	102		250	250
Sogd	123	117	121	119	118	119	242	234	239
Khatlon	139	107	132	50	44	49	190	151	181
RRP	115	139	119	49	81	54	164	220	173
GBAO	104	135	111	45	64	49	149	198	160
Tajikistan	125	129	126	66	89	74	191	218	200

Note: values are have been deflated with a temporal CPI but have *not* been deflated with the *spatial* CPI. Totals might differ from the sum of food nonfood due to rounding. Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

## 2.4 SPATIAL CONSUMER PRICE INDEX

Welfare comparisons require that monetary welfare indicators be adjusted for differences in purchasing power due to differences in price levels across geographical areas. In this



section we describe the steps involved in estimating a Paasche index at the household level. While other price indices can be chosen, the advantage with the Paasche index is twofold. Firstly, its estimation is computationally simple and fast. Secondly, the use of a Paasche index is highly recommendable because “well-founded” in economic theory: *money metric utility* (Samuelson 1974) can be approximated by dividing total household expenditure by a Paasche price index (Deaton 1980; Deaton and Zaidi 2002).

The Paasche index for the  $h$ -th household is defined as follows:

$$(6) \quad P_h = \frac{\sum_j p_j^h Q_j^h}{\sum_j p_j^0 Q_j^h}$$

where  $p_j^0$  is the price of commodity  $j$  for the reference group 0. The index is the ratio between the cost of a bundle of goods and services purchased by the  $h$ -th household and the cost of the same bundle as paid by some reference household (the “average household”, indexed by 0). From the previous formula we obtain:

$$(7) \quad P_h = \left[ \sum_j \left( \frac{p_j^h}{p_j^0} \right)^{-1} w_j^h \right]^{-1}$$

where  $w_j^h$  is the budget share of household  $h$  for commodity  $j$ , and  $p_j^h/p_j^0$  is the relative price of the  $j$ -th item.

### **Price relativities $p_j^h/p_j^0$**

The HBS provides information on expenditures and purchased quantities for both food- and (a selection of) non-food items. Thus, *unit values* can be calculated as follows:

$$(8) \quad uv_j^h = \frac{x_j^h}{Q_j^h(pur)}$$

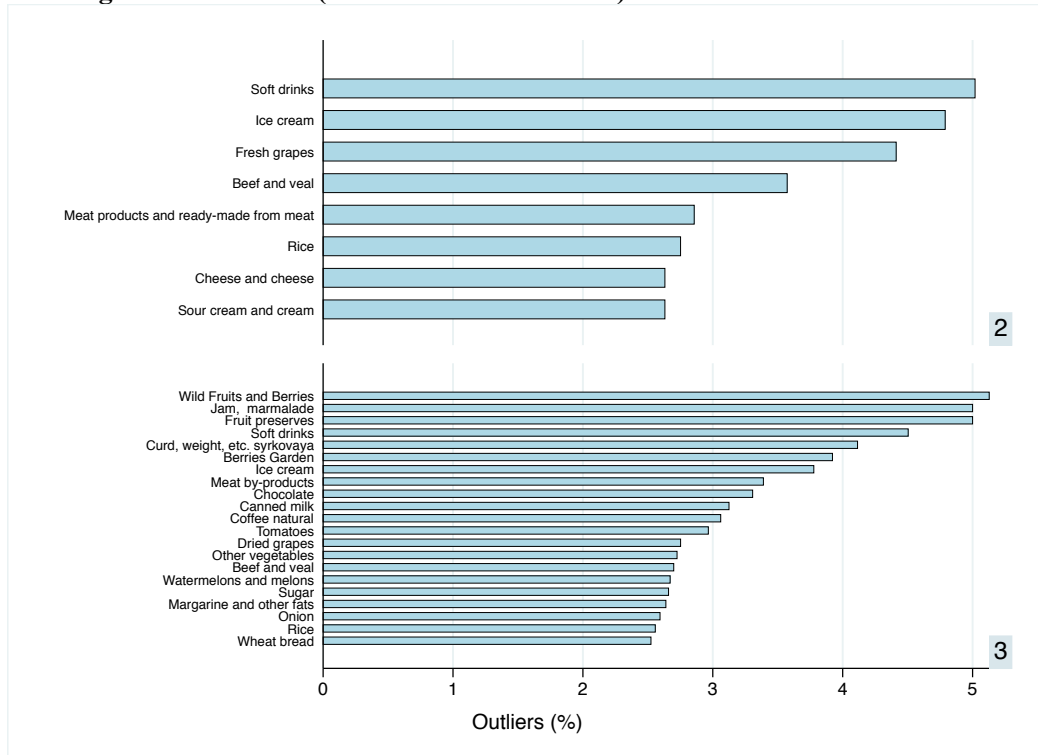
where  $x_j^h$  is the expenditure of household  $h$  on commodity  $j$ .

Before estimating unit values, we have detected outliers in the distribution of unit values at the product level using the following a rule, adapted from Deaton and Tarozzi (2008):

$$(9) \quad \left| \frac{\ln(uv_j^h) - \text{mean}[\ln(uv_j^h)]}{\text{std}[\ln(uv_j^h)]} \right| > 2.5$$

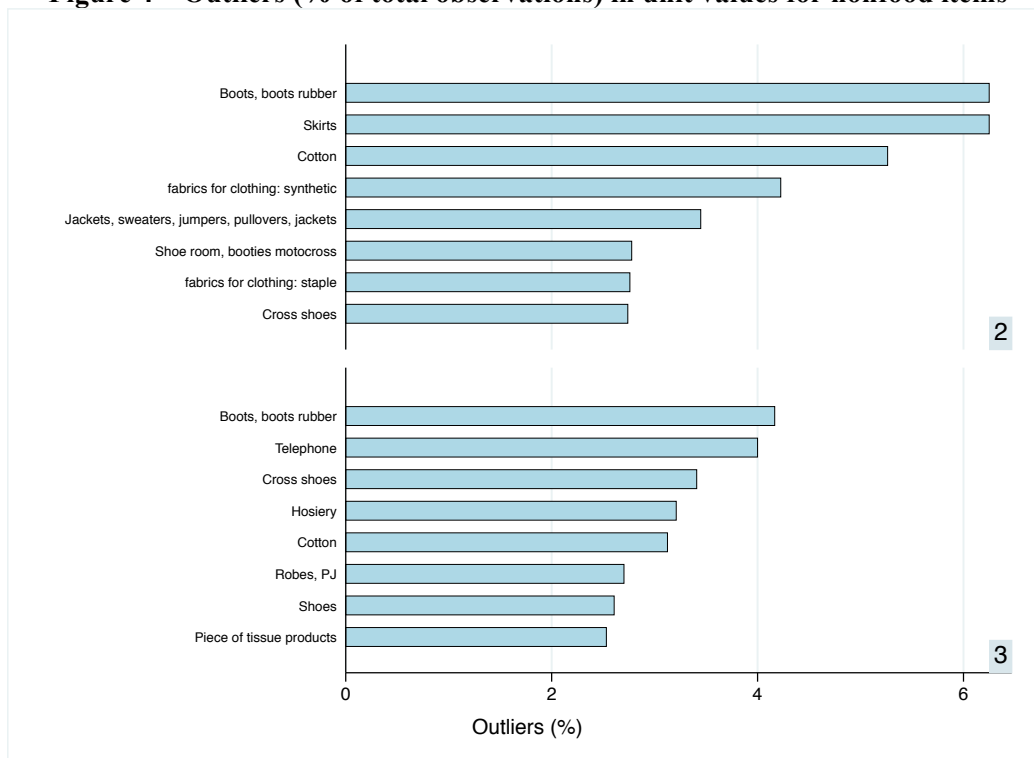
where  $\text{std}[\ln(uv_j^h)]$  is the standard deviation of  $\ln(uv_j^h)$ . Any observation that falls outside the interval defined in eq. 9 was set to missing and subsequently replaced with an imputed value (the procedure is described below). Figure 3 and Figure 4 show the distribution of outliers for food and nonfood items, respectively, by product.

**Figure 3 – Outliers (% of total observations) in unit values for food items**



Note: the top panel refers to quarter 2, the bottom panel to quarter 3. The figure shows items for which outliers are 2.5% of total observations or more. The full list is available in Appendix B.

**Figure 4 – Outliers (% of total observations) in unit values for nonfood items**

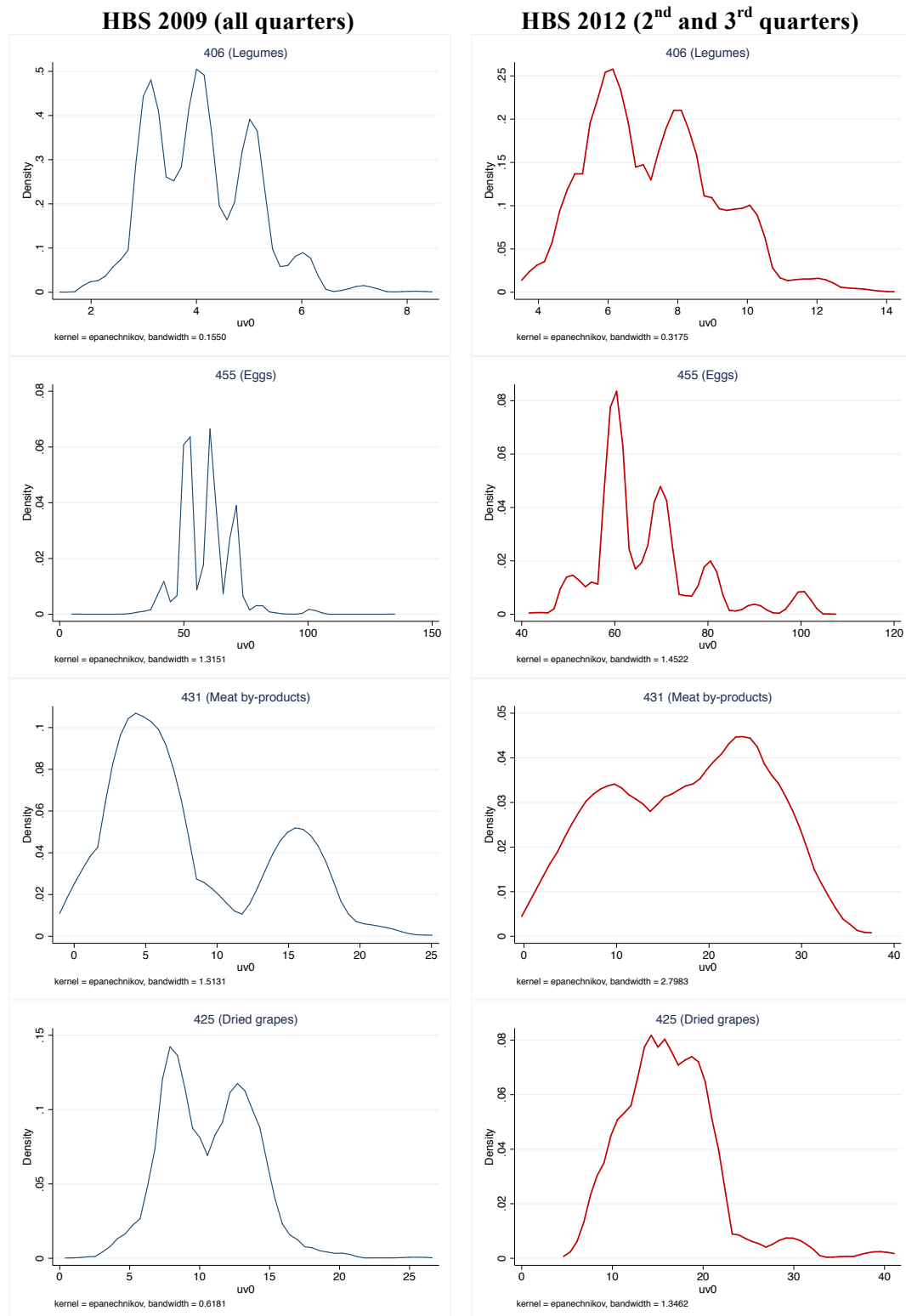


Note: the top panel refers to quarter 2, the bottom panel to quarter 3. The figure shows items for which outliers are 2.5% of total observations or more. The full list is available in Appendix B.

The distributions of unit values were also visually graphically for the presence of multi-modal distributions. Figure 5 shows a few examples of problematic distributions. While examples such as those shown in the figure exist, they are the exception. Most distributions are well-behaved. All things considered, we decided not to intervene on the data, not even in the presence of multi-modal distributions<sup>6</sup>.

<sup>6</sup> The same procedure was applied to non-food items: the distributions of unit values do not point to major inaccuracies in the data.

**Figure 5 – Empirical distribution functions of unit values for selected food items.**



Source: Authors' estimates.

Based on “cleaned” unit values, the ratio of price relativities  $p_j^h/p_j^0$  was estimated as follows:

$$\widehat{\left(\frac{p_j^h}{p_j^0}\right)} = \frac{uv_j^h}{uv_j^0}$$

where  $uv_j^0$  is the national average unit value of commodity  $j$ . Note that the  $j$ -th unit value  $uv_j^h$  can be missing even if the actual consumption of commodity  $j$  is strictly positive (self-production, running down the stocks, gifts received etc...). In all these cases, it is essential to impute the missing values. We followed the same hierarchical procedure described in section 2.3.1:

$$uv_j^h = \begin{cases} uv_j^h & \text{if } uv_j^h \text{ is not missing} \\ E[uv_j^h | region, area] & \text{if } uv_j^h \text{ is missing} \\ E[uv_j^h | region] & \text{if } E[uv_j^h | region, area] \text{ is missing} \\ E[uv_j^h | all h] & \text{if } E[uv_j^h | region] \text{ is missing} \end{cases}$$

where  $E[uv_j^h | H]$  denotes the average of  $uv_j^h$  over the reference group  $H$ .

### **Budget shares $w_j^h$**

The budget shares  ${}_F w_j^h$  needed to estimate a *food spatial price index* have been calculated as follows:

$${}_F w_j^h = \frac{\widehat{FHE}_h^j}{\sum_{j \in G_F} \widehat{FHE}_h^j}$$

where  $G_F$  denotes the set of food items covered by the index.

### **The Paasche food price index**

We have now all the elements to estimate a Paasche spatial price index for food items *at the household level*:

$$P_h^F = \left[ \sum_{j \in G_F} \left( \frac{uv_j^h}{uv_j^0} \right)^{-1} {}_F w_j^h \right]^{-1}$$

Household-level indices  $P_h^F$  can be aggregated by regional and area averaging over all the households that belong to a given region and area  $R$ :

$$P_R^F = E[P_h^F | h \in R]$$

Independently of the level of aggregation, all price indices have finally been normalized in order to give a unitary average value:

$$P_h^F(\text{normalized}) = P_h^F / E[P_h^F]$$

Table 8 shows the normalized food survey-based Paasche indices (SPI) by region and urban/rural areas.

**Table 8 – Spatial Food SPI (Tajikistan = 1)**

QUARTER 2						
Region	MEAN			MEDIAN		
	Rural	Urban	Total	Rural	Urban	Total
Dushanbe		1.0873	1.0873		1.0831	1.0831
Sogd	0.9276	0.9619	0.9408	0.9289	0.9545	0.9363
KHatlon	0.9521	1.0048	0.9661	0.9489	1.0015	0.9591
RRP	0.9894	1.0554	1.0026	0.9870	1.0536	0.9965
GBAO	1.1726	1.2314	1.1873	1.1438	1.2310	1.1738
Total	0.9762	1.0383	1.0000	0.9611	1.0293	0.9808
QUARTER 3						
Region	MEAN			MEDIAN		
	Rural	Urban	Total	Rural	Urban	Total
Dushanbe		1.0489	1.0489		1.0416	1.0416
Sogd	0.9498	0.9225	0.9393	0.9350	0.9070	0.9274
KHatlon	0.9737	0.9888	0.9777	0.9670	0.9887	0.9724
RRP	1.0207	1.0009	1.0168	1.0085	0.9782	1.0039
GBAO	1.1557	1.2431	1.1776	1.1522	1.2593	1.1775
Total	0.9968	1.0052	1.0000	0.9771	0.9970	0.9820
QUARTERS 2 AND 3						
Region	MEAN			MEDIAN		
	Rural	Urban	Total	Rural	Urban	Total
Dushanbe		1.0681	1.0681		1.0650	1.0650
Sogd	0.9387	0.9422	0.9400	0.9315	0.9340	0.9325
KHatlon	0.9629	0.9968	0.9719	0.9559	0.9953	0.9649
RRP	1.0051	1.0281	1.0097	0.9965	1.0267	1.0004
GBAO	1.1642	1.2373	1.1825	1.1460	1.2485	1.1745
Total	0.9865	1.0218	1.0000	0.9674	1.0137	0.9815

Source: Authors' estimates on Tajikistani HBS 2012.

### The Paasche total price index

The Paasche total spatial price index at the household level can be estimated following the same procedure as above. This time we include both food- and non-food items:

$$P_h^T = \left[ \sum_{j \in G_F \cup G_{NF}} \left( \frac{uv_j^h}{uv_j^0} \right)^{-1} {}^T W_j^h \right]^{-1}$$

where  $G_{NF}$  denotes the set of non-food items covered by the index. As before, all price indices have been normalized:

$$P_h^T(\text{normalized}) = P_h^T / E[P_h^F]$$

Table 9 shows the normalized total Paasche indices (SPI) by region and urban/rural areas.

**Table 9 – Spatial Total SPI (Tajikistan = 1)**

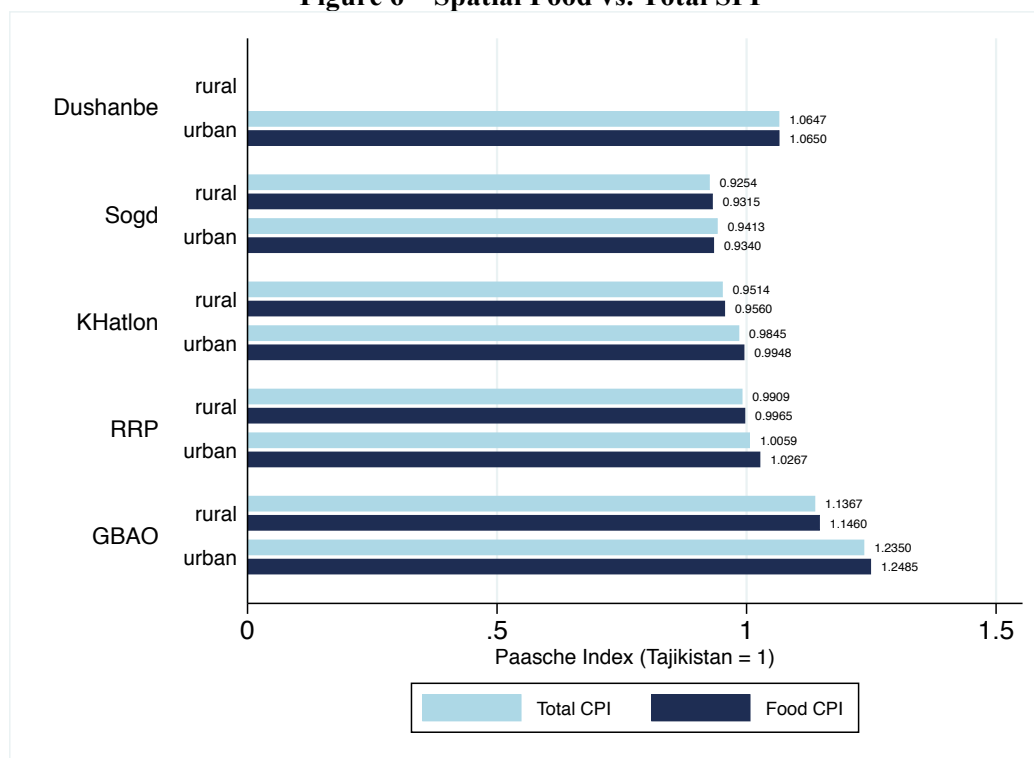
QUARTER 2						
Region	MEAN			MEDIAN		
	Rural	Urban	Total	Rural	Urban	Total
Dushanbe		1.0902	1.0902		1.0798	1.0798
Sogd	0.9282	0.9774	0.9471	0.9163	0.9551	0.9304
KHatlon	0.9483	0.9995	0.9619	0.9467	0.9920	0.9572
RRP	0.9881	1.0568	1.0018	0.9835	1.0393	0.9935
GBAO	1.1602	1.2292	1.1775	1.1411	1.2259	1.1670
Total	0.9735	1.0427	1.0000	0.9581	1.0274	0.9780
QUARTER 3						
Region	MEAN			MEDIAN		
	Rural	Urban	Total	Rural	Urban	Total
Dushanbe		1.0597	1.0597		1.0470	1.0470
Sogd	0.9536	0.9411	0.9488	0.9343	0.9215	0.9315
KHatlon	0.9712	0.9836	0.9745	0.9590	0.9807	0.9645
RRP	1.0083	0.9927	1.0052	0.9968	0.9693	0.9935
GBAO	1.1427	1.2380	1.1665	1.1311	1.2395	1.1663
Total	0.9925	1.0121	1.0000	0.9757	0.9960	0.9820
QUARTERS 2 AND 3						
Region	MEAN			MEDIAN		
	Rural	Urban	Total	Rural	Urban	Total
Dushanbe		1.0750	1.0750		1.0647	1.0647
Sogd	0.9409	0.9593	0.9480	0.9254	0.9413	0.9308
KHatlon	0.9597	0.9915	0.9682	0.9514	0.9845	0.9599
RRP	0.9982	1.0248	1.0035	0.9909	1.0059	0.9935
GBAO	1.1515	1.2336	1.1720	1.1367	1.2350	1.1666
Total	0.9830	1.0274	1.0000	0.9648	1.0115	0.9795

Source: Authors' estimates on Tajikistani HBS 2012.

How does the food SPI compare with the total SPI? Do they provide a consistent pattern of the cost-of-living in Tajikistan? The answer is positive (Figure 6). Due to the heterogeneity that one often finds in the unit values of non-food items *we have opted for using the survey based food Paasche index (SPI), calculated at the household level*. As results in Figure 6 suggest, in the context of Tajikistan the choice between the *food* SPI

versus *total* SPI is likely to have a negligible impact on the real consumption aggregate (Appendix C).

**Figure 6 – Spatial Food vs. Total SPI**



Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

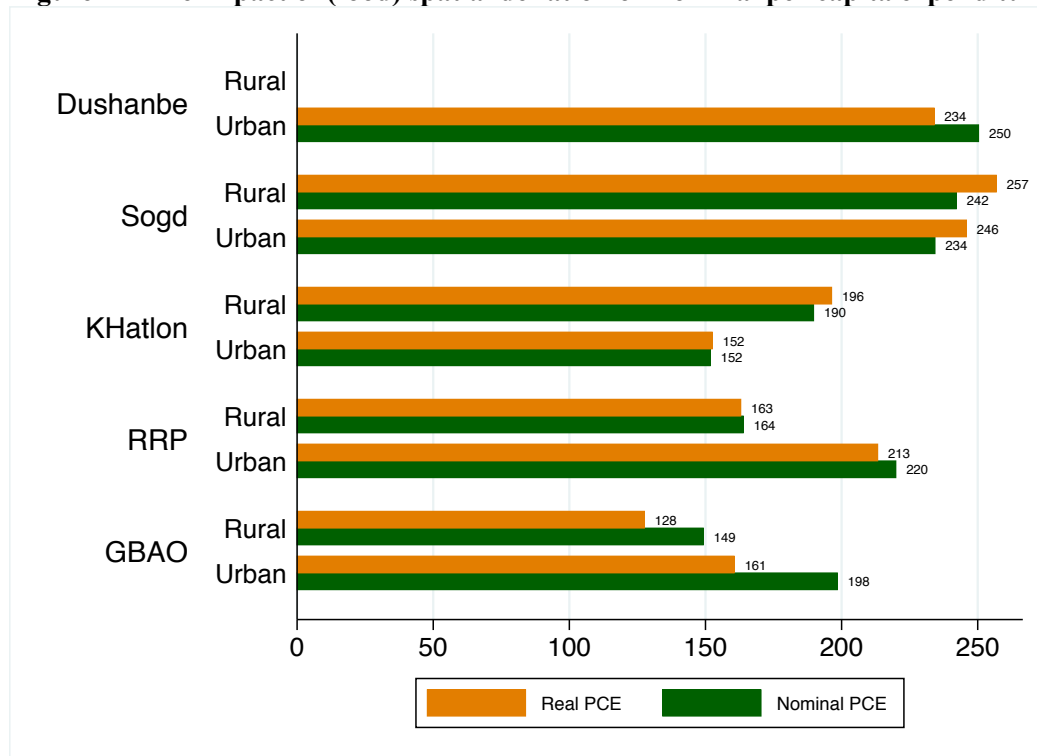
## 2.5 THE REAL CONSUMPTION AGGREGATE

The *nominal* consumption aggregate divided by the food spatial Paasche index (SPI) discussed in the previous section gives the *real* consumption aggregate, that is the welfare indicator defined in Section 1. Figure 7 shows the impact of spatial deflation on the nominal per capita expenditure.

Figure 8 shows the empirical probability density function of PCE separately by urban and rural areas, while Figure 9 shows the distribution of PCE for each region. At the national level, the empirical distribution functions of rural and urban households are remarkably similar, both in terms of levels (on average urban households' expenditures are ca. 9% higher than urban households' expenditures) and in terms of the shape of the distribution (more on this in Section 3).

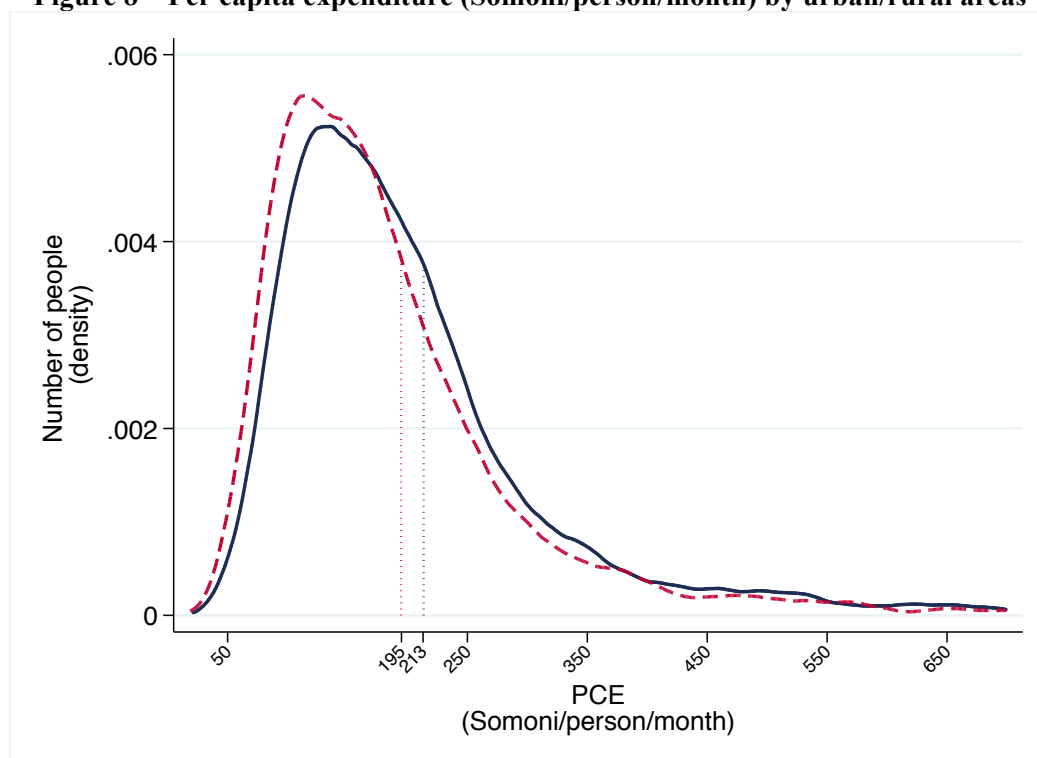


**Figure 7 – The impact of (food) spatial deflation on nominal per capita expenditure**



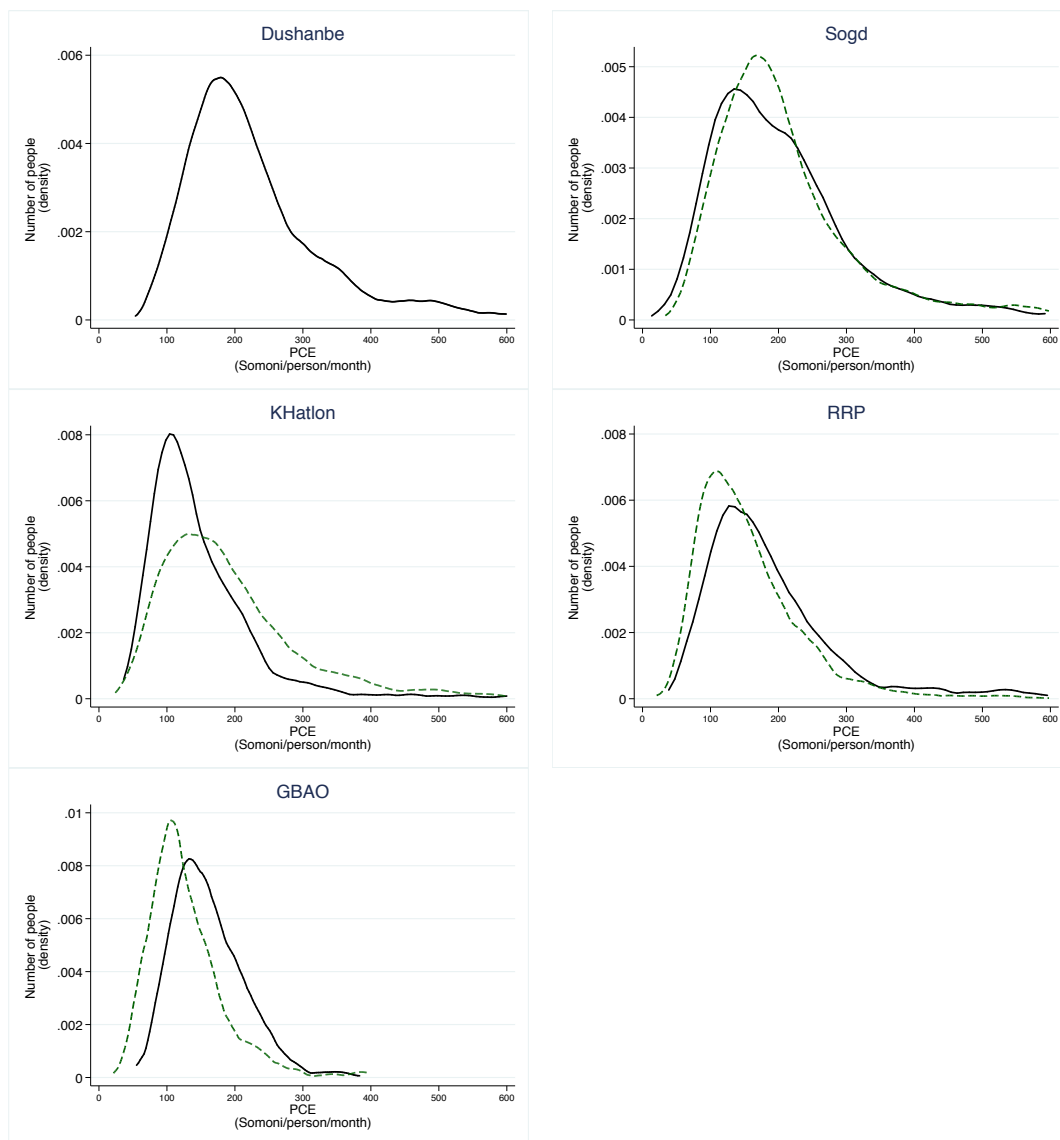
Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

**Figure 8 – Per capita expenditure (Somoni/person/month) by urban/rural areas**



Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

**Figure 9 – Empirical probability distribution functions of PCE by region.**



Note: solid line is for urban households, dashed line is for rural households. Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

Table 10 shows how the average consumption aggregate varies across national PCE deciles.

**Table 10 – Real per capita expenditure by national PCE deciles**

RURAL						
DECILE	Dushanbe	Sogd	KHatlon	RRP	GBAO	Total
1		79	73	75	70	73
2		100	99	99	100	99
3		117	117	117	117	117
4		137	136	136	136	136
5		155	155	155	154	155
6		174	175	176	176	175
7		197	199	199	200	199
8		229	230	230	228	230
9		284	283	276	271	281
10		644	456	516	369	545
Total		257	196	163	128	195
URBAN						
DECILE	Dushanbe	Sogd	KHatlon	RRP	GBAO	Total
1		77	74	75	78	75
2		100	100	100	97	100
3		116	118	117	118	117
4		137	136	135	137	136
5		156	154	156	155	155
6		175	176	175	172	175
7		198	198	198	197	199
8		229	231	226	225	229
9		284	280	284	275	281
10		452	646	505	604	544
Total		234	246	152	213	213
TOTAL						
DECILE	Dushanbe	Sogd	KHatlon	RRP	GBAO	Total
1		77	76	74	75	74
2		100	100	100	99	99
3		116	118	117	117	117
4		137	136	136	136	136
5		156	154	155	155	155
6		175	175	175	176	175
7		198	198	199	199	198
8		229	230	230	229	229
9		284	282	283	276	281
10		452	645	461	546	544
Total		234	253	186	172	201

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

### 3 INEQUALITY

Table 11 shows the estimated Gini indices by urban/rural areas and regions, together with standard errors and 95% bootstrapped confidence intervals. Inequality in Dushanbe and GBAO seems to be low compared to other regions, an issue that is further investigated in Appendix E.

**Table 11 – Gini index of real PCE by urban/rural areas and regions**

Group	Estimate	Std. Error	95% confidence interval	
			lower bound	upper bound
Rural	0.3225	0.0075	0.3078	0.3371
Urban	0.3165	0.0092	0.2984	0.3346
Dushanbe	0.2492	0.0071	0.2353	0.2631
Sogd	0.3624	0.0133	0.3364	0.3885
KHatlon	0.2969	0.0061	0.2848	0.3089
RRP	0.2965	0.0118	0.2734	0.3195
GBAO	0.2217	0.0078	0.2065	0.2369
Tajikistan	0.3213	0.0058	0.3099	0.3327

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

The difference between urban and rural Gini indices, however, is *not* statistically significant (Table 12).

**Table 12 – Testing the difference between Gini indices in urban/rural areas**

index	estimate	Std. Err.	t	P>t	[95% Confidence Interval]	
urban	0.3165	0.0092	34.2973	0.0000	0.2984	0.3346
rural	0.3225	0.0075	43.1568	0.0000	0.3078	0.3371
difference	0.0060	0.0119	0.5046	0.6138	-0.0173	0.0293

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

The inequality ranking of regions does not depend on the choice of the inequality measure (Table 13).

**Table 13 – Inequality indices by region**

region	Theil	MLD	GEI(1)	GEI(2)	Atkinson (0.5)	Atkinson (1)	Atkinson (2)
Dushanbe	0.10147	0.09919	0.10720	0.13075	0.05022	0.09443	0.16871
Sogd	0.21891	0.21954	0.29053	0.61198	0.11730	0.19711	0.30451
KHatlon	0.15402	0.14251	0.15267	0.19790	0.07098	0.13282	0.23550
RRP	0.14704	0.14594	0.17696	0.29545	0.07642	0.13579	0.22726
GBAO	0.08851	0.08049	0.07989	0.08620	0.03928	0.07734	0.15039
Tajikistan	0.17683	0.17158	0.2106	0.39531	0.08952	0.15767	0.26127

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters). Note: MLD = mean logarithmic deviation, GEI(a) = Generalized Entropy Index with parameter a.

## 4 THE POVERTY LINE

This section illustrates the procedure followed to set an absolute poverty line for Tajikistan based on the second quarter data of the 2012 HBS. Ravallion (1994) provides a clear exposition of the theoretical rationale of the so-called “cost-of-basic-needs method” (CBN).

According to the CBN method, the total poverty line ( $Z$ ) is defined as the sum of two components, namely a *food poverty line* ( $Z_F$ ) and an *allowance for non-food consumption* ( $Z_{NF}$ ):

$$Z = \underbrace{Z_F}_{\text{sect. 4.1}} + \underbrace{Z_{NF}}_{\text{sect. 4.2}}$$

where  $Z_F$  is the “food poverty line” (section 4.1), and  $Z_{NF}$  is a non-food allowance (section 4.2).

### 4.1 THE FOOD POVERTY LINE ( $Z_F$ )

The determination of a food poverty line is done in three steps:

- 1) We first estimate the *average kilocalorie requirement* (AKR), defined as the average number of kilocalories (per person per day) needed to meet daily energy requirements in Tajikistan.
- 2) Next we estimate the *minimum cost of one kilocalorie*. The unit calorie cost can be approximated by the average cost of one kilocalorie for a “reference group”, as close to the poor as possible.
- 3) Once steps 1) and 2) are completed, a *food poverty line*  $Z_F$  is obtained as the product of the AKR (step 1) times the minimum kilocalorie unit cost (step 2):

$$Z_F = \underbrace{\text{AKR}}_{\text{sect. 4.1.1}} \times \underbrace{(\text{cost of one kilocalorie})}_{\text{sect. 4.1.2}}$$

Section 4.1.1 illustrates the procedure used to set the AKR, while section 4.1.2 explains how the unit cost of kilocalories was calculated.

#### 4.1.1 The Average Kilocalorie Requirement (AKR)

While there are many methods available to *estimate* the AKR [Kakwani 2003], in this note we interpret AKR as a *normative value*, that is as the minimum amount of kilocalories that – on average – Tajikistani people ought consume daily in order to live a healthy and active social life. *We set AKR equal to 2,250 kilocalories per person per day*. This is the same threshold used in the 2009 poverty assessment report<sup>7</sup>.

#### 4.1.2 The Minimum Cost Of One Kilocalorie

The unit cost of kilocalories was estimated by dividing the total expenditure on food items evaluated at the market prices by the total caloric intake for each household in the sample. The calorie intake was calculated by converting the purchased quantities into kilocalories using the conversion factors reported in the Appendix D. Table 14 shows the distribution of calories by national PCE deciles.

**Table 14 – Distribution of kilocalories (person/day) by PCE deciles**

PCE decile	PURCHASED			CONSUMED		
	Rural	Urban	Tajikistan	Rural	Urban	Tajikistan
1	1296	1379	1315	1345	1352	1346
2	1570	1563	1568	1669	1558	1638
3	1685	1652	1674	1810	1622	1748
4	1789	1873	1818	2001	1800	1931
5	1807	1966	1856	2097	1895	2034
6	1974	2069	2005	2204	1963	2124
7	1933	2098	1993	2248	2032	2169
8	2068	2334	2173	2414	2128	2302
9	2234	2498	2332	2614	2251	2480
10	2461	3035	2681	2900	2666	2810
Total	1861	2102	1941	2102	1971	2058

Note: The left panel shows the distribution of calories from *purchased* food, while the right panel provides an estimate of calories from *consumed* food. Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

The cost of one calorie can be calculated by dividing the total expenditure on purchased food of a household by the total number of purchased kilocalories. Table 15 shows that the cost of one kilocalorie monotonically increases as one moves from a lower to a higher decile.

<sup>7</sup> Appendix F shows the sensitivity of poverty rates to the choice of the AKR.

**Table 15 – Average cost of one calorie by PCE deciles**

PCE decile	MEDIAN			MEAN		
	Rural	Urban	Tajikistan	Rural	Urban	Tajikistan
1	0.00134	0.00148	0.00137	0.00141	0.00151	0.00144
2	0.00137	0.00158	0.00143	0.00147	0.00164	0.00151
3	0.00146	0.00165	0.00150	0.00158	0.00178	0.00164
4	0.00146	0.00174	0.00157	0.00161	0.00186	0.00170
5	0.00157	0.00183	0.00165	0.00173	0.00195	0.00179
6	0.00154	0.00193	0.00170	0.00174	0.00202	0.00184
7	0.00162	0.00203	0.00180	0.00185	0.00214	0.00196
8	0.00167	0.00215	0.00190	0.00185	0.00229	0.00202
9	0.00162	0.00232	0.00188	0.00181	0.00239	0.00202
10	0.00182	0.00256	0.00208	0.00199	0.00266	0.00224
Total	0.00152	0.00192	0.00163	0.00169	0.00207	0.00182

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

In order to estimate the *minimum* cost of the average caloric requirement we need to choose a *reference group* of households. Table 15 shows that, at the national level, the cost of one calorie varies in the range of 0.00137-0.00208 Somoni/kilocalorie (left panel of the table), depending on the reference decile. How to choose the reference group? Clearly this is a crucial decision. “Self-fulfilling prophecies”, as Pradhan, Suryahadi, Sumarto and Pritchett (2000) have named it, is the risk associated with the discretionary choice of a reference group made by the analyst: “Two researchers working on the same country with exactly the same data using exactly the same method but simply having different prior beliefs on headcount poverty will produce different poverty estimates. The one who believes poverty is high will choose a wealthier reference population. This richer reference group will consume a more luxurious food basket. Hence the calories per rupiah will be lower so the cost of obtaining a fixed amount of calories will be higher. (...) This researcher will most likely get a higher estimated headcount poverty compared to the researcher who started off with a low prior.” (p. 7) The implication of the above argument is that “the “standard” poverty methodology is incomplete and not well specified. Without a procedure for fixing the reference group, the “standard” method applied to the same country with the same data can produce different outcome” (p. 8).

One solution consists in implementing the iterative method designed by Pradhan et al. (p. 39). This minimizes the chances that poverty rates are estimated in a somewhat circular way. However, the implementation of the 11-step procedure poses a high computational burden on the analysis. Not only this, the method is not exempt from arbitrary decisions that the analyst must take to implement the procedure. Aware of the pros and cons associated to different methods, we have opted for the simplest, most transparent and easy-to-replicate solution: we calculated the calorie unit cost as the *median unit value of*

calorie over the bottom three national PCE deciles. By so doing, we find that *the cost of 1 kcal is equal to 0.0014303 Somoni*. This is the value used to price the AKR<sup>8</sup>.

By multiplying the AKR value (2,250 kcal/person/day) by the minimum average cost of 1 kcal (0.0014303 Somoni/kcal) and transforming this last value into monthly terms, we estimate that *the food poverty line is equal to 97.9 Somoni/person/month*. If we use  $Z_F = 97.9$  as a poverty line, we obtain a set of estimates of the so-called *extreme poverty* in Tajikistan 2012 (Table 16).

**Table 16 – Extreme poverty (food poverty line = 97.9 Somoni/person/month)**

Region	Poverty Headcount Rate	95% confidence interval		Distribution of the poor	Distribution of the population	Poverty Gap	95% confidence interval		Squared Poverty Gap	Sen-Shorrocks-Thon Index	Watts Index
<b>Tajikistan</b> <i>s.e.</i>	14.0 (0.54)	12.9	15.0	100.0	100.0	2.6 (0.14)	2.3	2.9	0.8 (0.06)	5.0 (0.26)	3.1 (0.18)
<b>Urban</b> <i>s.e.</i>	10.1 (0.77)	8.6	11.6	24.0	33.3	1.7 (0.18)	1.4	2.1	0.5 (0.07)	3.3 (0.33)	2.0 (0.22)
<b>Rural</b> <i>s.e.</i>	15.9 (0.71)	14.5	17.3	76.0	66.7	3.1 (0.19)	2.7	3.4	0.9 (0.08)	5.8 (0.35)	3.7 (0.24)
<b>Dushanbe</b> <i>s.e.</i>	1.5 (0.57)	0.4	2.6	1.2	10.8	0.2 (0.07)	0.0	0.3	0.0 (0.01)	0.3 (0.13)	0.2 (0.08)
<b>Sogd</b> <i>s.e.</i>	6.7 (0.75)	5.2	8.1	12.4	25.9	1.1 (0.17)	0.8	1.4	0.3 (0.07)	2.2 (0.32)	1.3 (0.21)
<b>KHatlion</b> <i>s.e.</i>	16.4 (0.98)	14.5	18.3	36.8	31.3	3.2 (0.25)	2.7	3.7	0.9 (0.09)	6.1 (0.45)	3.9 (0.31)
<b>RRP</b> <i>s.e.</i>	20.4 (1.38)	17.7	23.1	34.4	23.6	3.6 (0.37)	2.9	4.3	1.0 (0.15)	6.7 (0.67)	4.2 (0.48)
<b>GBAO</b> <i>s.e.</i>	25.5 (2.27)	21.0	29.9	15.3	8.4	5.5 (0.68)	4.1	6.8	1.8 (0.30)	10.1 (1.18)	6.8 (0.89)

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters). Note: standard errors are calculated using the software DASP described in Araar and Duclos (2007).

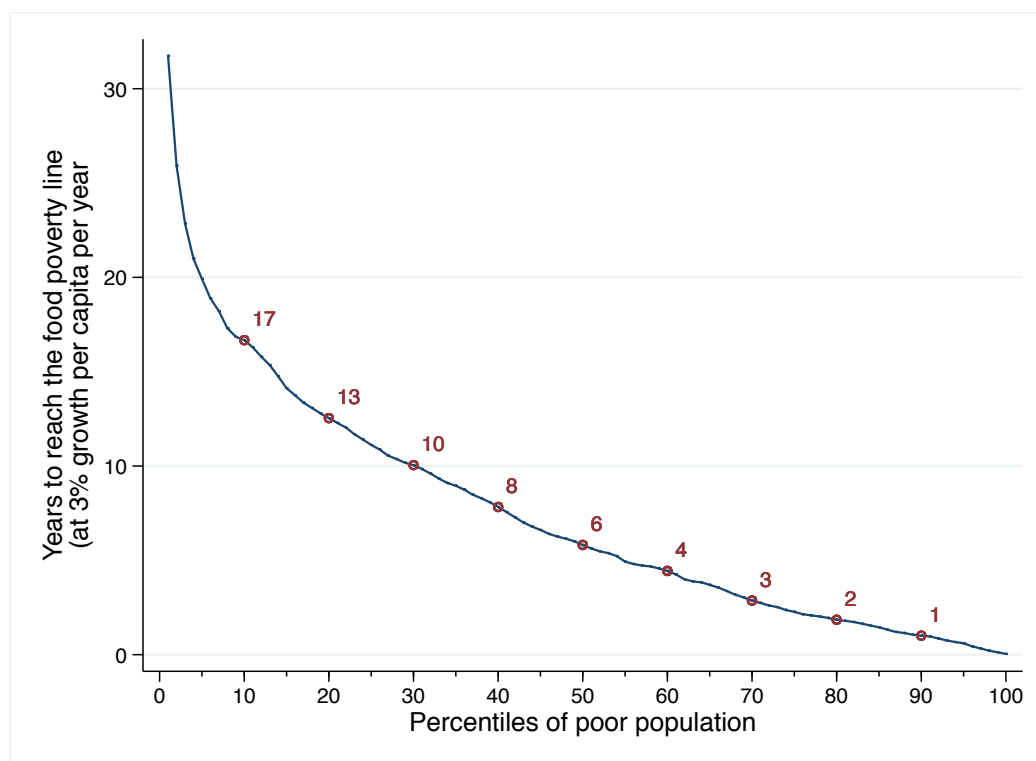
The Watts index (last column in Table 16), has a number of appealing ordinal properties – as discussed in Zheng (1997). Morduch (1994, 1998) has shown that a simple linear transformation of the Watts index gives another useful indicator – named the “*average exit time*” – which can be interpreted as “the average number of years that it would take the population to exit poverty if it were possible to ensure that all incomes grow at rate  $g$ ” [Morduch 1998: 386]. In the Tajikistani context, according to the estimates in Figure 10, the poorest 10% of the poor would take 17 years to reach the poverty line under the assumption that per capita consumption increases at the rate of 3% per year. In Figure 10

<sup>8</sup> Appendix F shows the sensitivity of poverty rates to the choice of the reference group.



the number of years required to escape poverty monotonically decreases with the PCE deciles of the poor.

**Figure 10 – Average exit time for the Tajikistani poor.**



Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

## 4.2 THE NON-FOOD COMPONENT OF THE POVERTY LINE

After estimating the food poverty line, the CBN approach yields the *total* poverty line by adding an allowance for non-food basic needs to the *food* poverty line. We followed Ravallion (1994) and Ravallion and Bidani (1994), and calculated a lower- and an upper-bound for the total poverty lines. We will denote these two measures by  $Z_L$  and  $Z_U$ , respectively.

### Lower bound ( $Z_L$ )

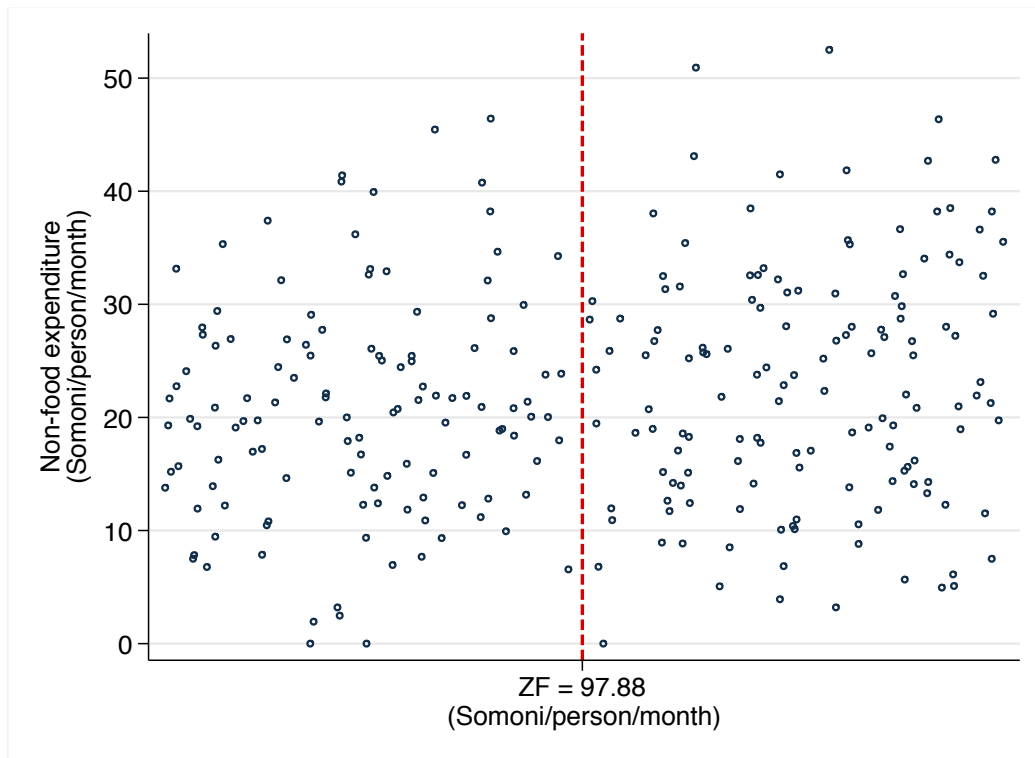
The calculation can be implemented following as a two-step procedure. In step 1 the households whose *total* expenditures are approximately equal to the *food* poverty line are identified. In step 2 the non-food allowance is estimated as an average over the expenditures on non-food items across the subset of households identified in step 1. The corresponding poverty line,  $Z_L$ , is therefore given by:

$$Z_L = Z_F + E(X_{\text{non food}} | X \cong Z_F)$$

where  $X$  denotes the total household expenditure,  $Z_F$  is the food poverty line, and the symbol  $\cong$  is short for “approximately equal to”.

Using the 2012 HBS data, we can visualize (Figure 11) the set of households for which the expenditure on food is approximately equal to the food poverty line ( $Z_F = 97.9$  Somoni/person/month) estimated in section 2.1.3.

**Figure 11 – Non-food expenditure of households with PCE approximately equal to the food poverty line.**

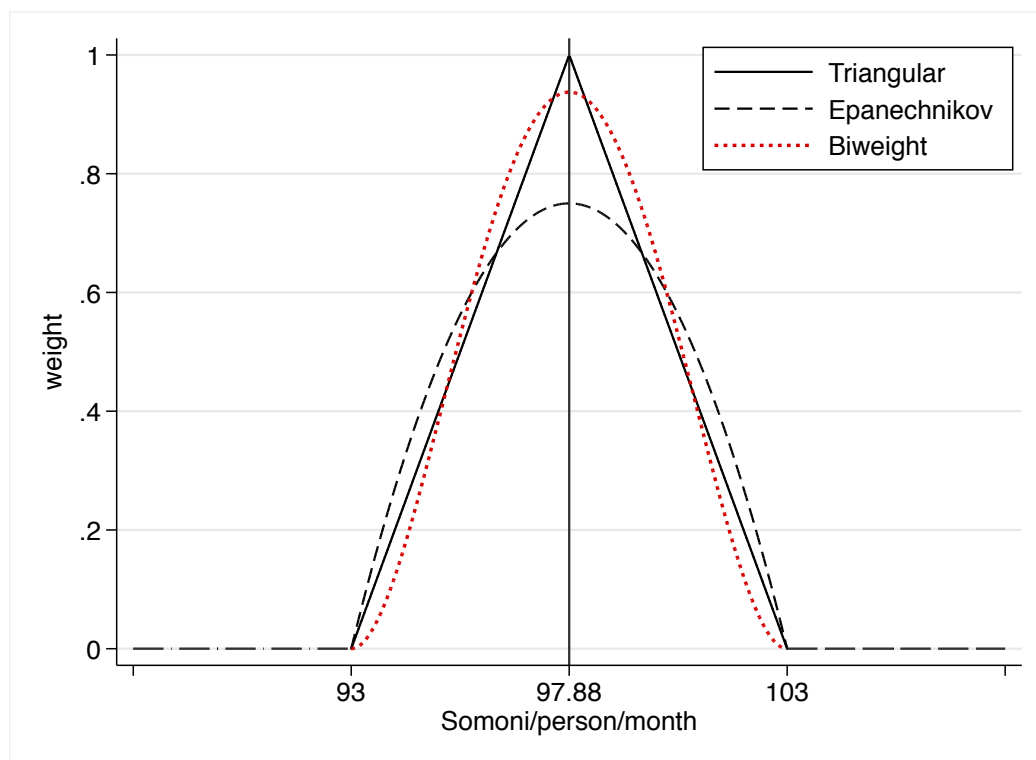


Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters). Note: the interval centered around the food poverty line is  $\pm 5$  Somoni/person/month.

The simplest method for estimating the expected value in the above equation consists in defining a symmetric interval around the food poverty line and calculating the average non-food expenditure over the households whose food expenditures belong to this interval. One might consider using the *median*, and not the mean, as the former is less sensitive to extreme values. Using this method the non-food allowance equals 21.9 Somoni/person/month. The so-called Lower Poverty Line ( $Z_L$ ) is then equal to 119.8 Somoni/person/month (Table 17, row 3).

Following Ravallion and Chen (1996), we have also calculated the average non-food expenditure within each of a range of intervals (in increments of 1 Somoni up to 5 Somoni) and taken the average of those. The results are not significantly different from those obtained with just one interval. While the method introduced by Chen and Ravallion (1996) has the advantage of increasing the robustness of results to the choice of the width of the interval centered at the food poverty line, averaging over different values of non-food expenditures introduces an *implicit* weighing scheme. Specifically, households that are closer to  $Z_F$  are assigned higher weights than those that are further away. This feature may be seen as a desirable property of the procedure, but then again a shortcoming is that we do not control for *how* the weights decrease when moving away from the food poverty line. A solution consists in using an *explicit* weighing scheme, for instance by means of the *kernel functions* (Figure 12).

**Figure 12 – Selected kernel functions.**



Note: the kernel functions are centered at the food poverty line, and assign decreasing weights to households the more total expenditures are far away from the food poverty line. Households with total expenditures outside the range [93, 103] Somoni/person/month are assigned a zero weight, that is are excluded from the calculation of the non-food allowance.

Using the Biweight kernel we estimate the non-food allowance equal to 21.3 Somoni/person/month. The lower poverty line  $Z_L$  is equal to 119.2 Somoni/person/month. The choice of the kernel function does not affect significantly the estimate of the poverty line (Table 17, rows 4-6).

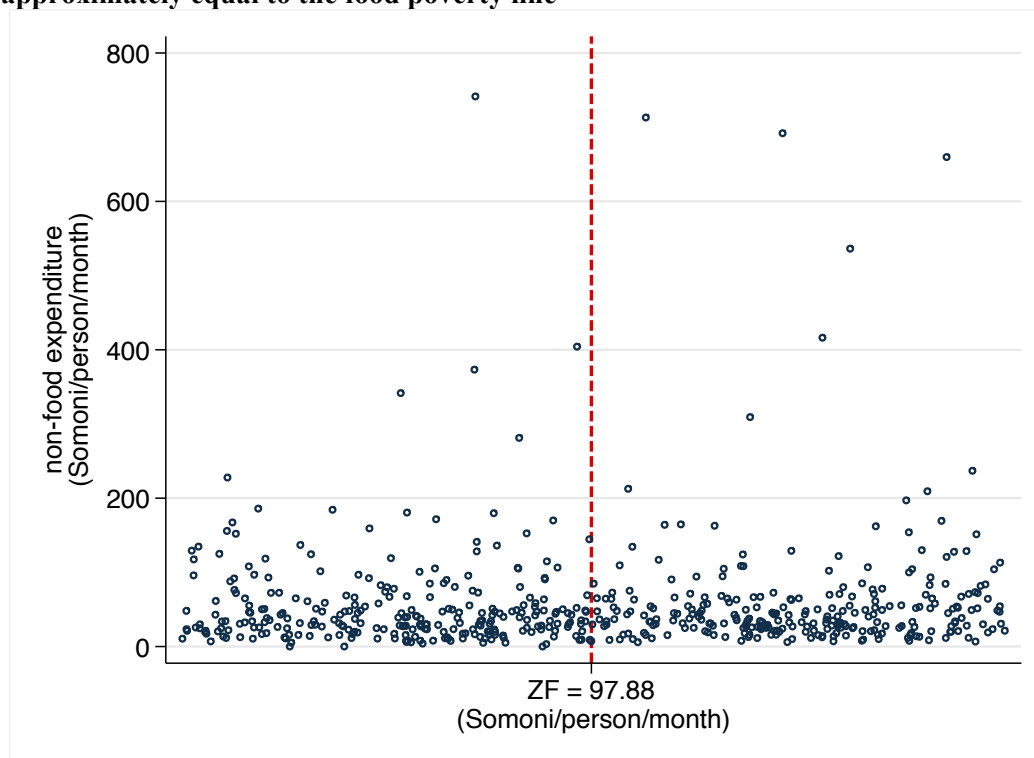
### Upper bound ( $Z_U$ )

Ravallion (1994) suggests a second method that makes a more generous allowance for non-food spending. This alternative procedure focuses on households whose actual *food* spending is equal to the food poverty line, and defines the total poverty as follows:

$$Z_U = Z_F + E(X_{\text{non food}} | X_{\text{food}} \cong Z_F)$$

where  $X_{\text{food}}$  and  $X_{\text{non food}}$  denote, respectively, the food and non-food household expenditures,  $Z_F$  is the food poverty line, and, as before, the symbol  $\cong$  denotes “approximately equal to”.

**Figure 13 – Non-food expenditure of households with food expenditure approximately equal to the food poverty line**



Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters). One outlying observation was dropped before producing the graph.

Using the Biweight kernel we estimate the non-food allowance equal to 36.4 Somoni/person/month (Table 17, row 10). *The upper poverty line  $Z_U$  is equal to 134.2 Somoni/person/month.* In Section 4 (Table 16) we used the food poverty line in row 2; in Section 5 we will use the lower- and upper- poverty lines in rows 6 and 10.

**Table 17 – Sensitivity of poverty rates to poverty lines**

	method	deflated with spatial CPI	food poverty line	non food allowance	total poverty line	H0	PG	PG2
1	FOOD poverty line	no	97.5	.	.	14.60	2.80	0.80
2	FOOD poverty line	yes	97.9	.	.	14.00	2.60	0.80
3	LOWER bound - Absolute Intervals	yes	97.9	21.9	119.8	26.60	5.90	1.90
4	LOWER bound - Epanechnikov	yes	97.9	21.3	119.2	25.50	5.60	1.80
5	LOWER bound - Triangular	yes	97.9	21.3	119.2	25.50	5.60	1.80
6	LOWER bound - Biweight	yes	97.9	21.3	119.2	25.60	5.60	1.80
7	UPPER bound - Absolute Intervals	yes	97.9	35.1	133.0	33.00	8.10	2.80
8	UPPER bound - Epanechnikov	yes	97.9	36.8	134.7	33.80	8.40	2.90
9	UPPER bound - Triangular	yes	97.9	36.5	134.4	33.70	8.30	2.90
10	UPPER bound - Biweight	yes	97.9	36.4	134.2	33.70	8.30	2.90

Note: poverty lines are expressed in Somoni/person/month. Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

## 5 POVERTY ESTIMATES

Table 18 and Table 19 summarize poverty estimates corresponding to the lower and upper poverty lines described in Section 4.2.

**Table 18 – Poverty estimates: lower poverty line (119.2 Somoni/person/month)**

Region	Poverty Headcount Rate	95% confidence interval		Distribution of the poor	Distribution of the population	Poverty Gap	95% confidence interval		Squared Poverty Gap	Sen-Shorrocks-Thon Index	Watts Index
<b>Tajikistan</b> <i>s.e.</i>	26.2 (0.65)	24.9	27.5	100.0	100.0	5.8 (0.20)	5.4	6.2	1.9 (0.09)	10.6 (0.34)	7.1 (0.26)
<b>Urban</b> <i>s.e.</i>	21.2 (1.00)	19.3	23.2	27.0	33.3	4.2 (0.27)	3.7	4.7	1.3 (0.11)	7.9 (0.48)	5.1 (0.35)
<b>Rural</b> <i>s.e.</i>	28.7 (0.84)	27.0	30.3	73.0	66.7	6.5 (0.26)	6.0	7.1	2.2 (0.12)	11.9 (0.45)	8.1 (0.36)
<b>Dushanbe</b> <i>s.e.</i>	7.9 (1.20)	5.6	10.3	3.3	10.8	1.0 (0.19)	0.7	1.4	0.2 (0.05)	2.0 (0.37)	1.2 (0.22)
<b>Sogd</b> <i>s.e.</i>	15.7 (1.03)	13.7	17.7	15.5	25.9	2.9 (0.26)	2.4	3.4	0.8 (0.11)	5.6 (0.48)	3.5 (0.34)
<b>KHatlon</b> <i>s.e.</i>	29.8 (1.18)	27.5	32.1	35.7	31.3	6.8 (0.36)	6.1	7.5	2.2 (0.16)	12.3 (0.60)	8.4 (0.47)
<b>RRP</b> <i>s.e.</i>	34.1 (1.55)	31.1	37.2	30.7	23.6	7.9 (0.51)	6.9	8.9	2.5 (0.24)	14.0 (0.83)	9.6 (0.69)
<b>GBO</b> <i>s.e.</i>	46.3 (2.43)	41.5	51.0	14.8	8.4	10.9 (0.87)	9.2	12.6	3.9 (0.44)	18.8 (1.35)	13.8 (1.21)

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters). Note: standard errors are calculated using the software DASP described in Araar and Duclos (2007).

**Table 19 – Poverty estimates: upper poverty line (134.2 Somoni/person/month)**

Region	Poverty Headcount Rate	95% confidence interval		Distribution of the poor	Distribution of the population	Poverty Gap	95% confidence interval		Squared Poverty Gap	Sen-Shorrocks-Thon Index	Watts Index
<b>Tajikistan</b> <i>s.e.</i>	33.8 (0.69)	32.5	35.2	100.0	100.0	8.5 (0.24)	8.0	8.9	3.0 (0.11)	15.1 (0.39)	10.6 (0.32)
<b>Urban</b> <i>s.e.</i>	29.3 (1.09)	27.1	31.4	28.8	33.3	6.6 (0.33)	5.9	7.2	2.1 (0.15)	11.9 (0.56)	8.1 (0.44)
<b>Rural</b> <i>s.e.</i>	36.1 (0.87)	34.4	37.8	71.2	66.7	9.4 (0.31)	8.8	10.1	3.4 (0.15)	16.7 (0.50)	11.9 (0.43)
<b>Dushanbe</b> <i>s.e.</i>	12.9 (1.44)	10.1	15.7	4.1	10.8	2.1 (0.30)	1.5	2.7	0.5 (0.09)	4.0 (0.55)	2.4 (0.35)
<b>Sogd</b> <i>s.e.</i>	22.5 (1.15)	20.2	24.7	17.2	25.9	4.7 (0.33)	4.1	5.4	1.5 (0.14)	8.8 (0.58)	5.8 (0.44)
<b>KHatlon</b> <i>s.e.</i>	36.9 (1.23)	34.4	39.3	34.1	31.3	9.8 (0.42)	8.9	10.6	3.5 (0.20)	17.2 (0.67)	12.3 (0.58)
<b>RRP</b> <i>s.e.</i>	43.6 (1.58)	40.5	46.7	30.4	23.6	11.4 (0.59)	10.2	12.5	4.0 (0.30)	19.5 (0.90)	14.3 (0.82)
<b>GBO</b> <i>s.e.</i>	57.3 (2.33)	52.8	61.9	14.2	8.4	15.5 (0.96)	13.7	17.4	5.9 (0.53)	25.3 (1.37)	20.0 (1.40)

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters). Note: standard errors are calculated using the software DASP described in Araar and Duclos (2007).

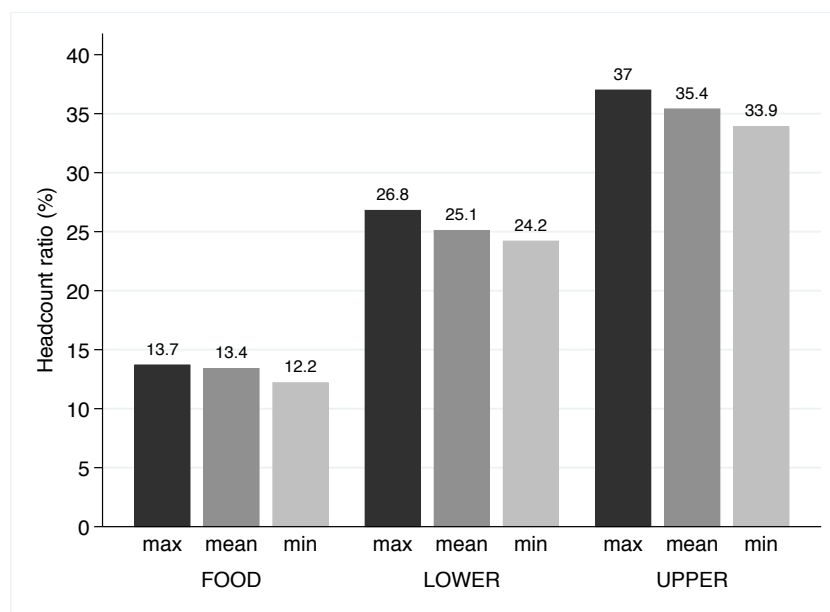
Appendix G shows quarterly poverty rates and compares Tajikistan across-quarter poverty rates changes with other countries.

## APPENDIX A

### SENSITIVITY OF POVERTY RATES TO ERRORS IN THE ESTIMATED HOUSEHOLD SIZE

The figure shows the estimated poverty rates obtained by implementing different solutions to the problem with the household size variable described in section 2.1. Dark bars (label “max” in the figure) refer to the case where each household is assigned the *maximum* size among the values reported across quarters. Light bars (label “min” in the figure) refer to the estimated headcount poverty rates after assigning households the minimum value between the sizes reported across quarters. The bars in the middle of each group (label “mean”) corresponds to poverty rates when households are assigned the across-quarter average household size – see Table 3.

**Figure 14 – Headcount poverty rates under different hypotheses for estimating the household size**



Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

## APPENDIX B OUTLIERS IN UNIT VALUES

**Table 20 – Outliers (%) for unit values of food items**

PRODUCT	Quarter		Total
	2	3	
Beef and veal	3.6	2.7	3.1
Beets, carrots, radishes Other Edible ro	0.7	0.9	0.8
Berries Garden	2.2	3.9	3.1
Butter	0.4	0.9	0.6
Cabbage	1.8	0.4	1.1
Candy, halva, etc.	2.4	2.4	2.4
Canned milk	1.4	3.1	2.3
Canned vegetables	0.6	1.2	0.9
Cereals, except rice	1.4	0.8	1.1
Cheese and cheese	2.6	0.0	1.3
Chocolate	2.2	3.3	2.8
Citrus	2.4	0.0	1.2
Coffee natural	1.9	3.1	2.5
Cookies, cakes, etc.	1.7	1.0	1.3
Cucumbers	0.4	2.4	1.4
Curd, weight, etc. syrkovaya	0.6	4.1	2.3
Dried Fruits and Berries	0.5	0.3	0.4
Dried grapes	0.8	2.8	1.8
Eggs	1.1	1.2	1.2
Fish, fresh	0.0	2.2	1.1
Fresh grapes	4.4	1.1	2.8
Fruit Garden	1.8	1.5	1.7
Fruit preserves	0.0	5.0	2.5
Honey bee	1.1	0.6	0.9
Ice cream	4.8	3.8	4.3
Jam, marmalade	0.0	5.0	2.5
Legumes	0.5	1.5	1.0
Margarine and other fats	1.4	2.6	2.0
Meat by-products	0.0	3.4	1.7
Meat products and ready-made from meat	2.9	2.2	2.5
Meats and myasokopchenosti	0.6	0.4	0.5
Mutton and goat meat	0.0	0.9	0.5
Onion	0.8	2.6	1.7
Other vegetables	2.2	2.7	2.4
Pasta	2.1	1.7	1.9
Potatoes	2.3	2.3	2.3
Poultry	1.4	1.9	1.7
Rice	2.8	2.6	2.7
Salt	1.0	0.7	0.8
Soft drinks	5.0	4.5	4.8
Sour cream and cream	2.6	0.9	1.8
Sugar	2.5	2.7	2.6
Tea	1.8	1.4	1.6
Tomatoes	0.0	3.0	1.5
Vegetable oil	1.3	1.6	1.4
Vodka	0.0	1.6	0.8
Watermelons and melons	2.1	2.7	2.4
Wheat Flour	1.0	0.8	0.9
Wheat bread	1.9	2.5	2.2
Whole milk	1.5	1.4	1.4
Wild Fruits and Berries	0.0	5.1	2.6
Zucchini, squash, pumpkins, etc.	0.0	1.7	0.9
Total	1.2	1.8	1.5

Note: total number of observations is 408,000.



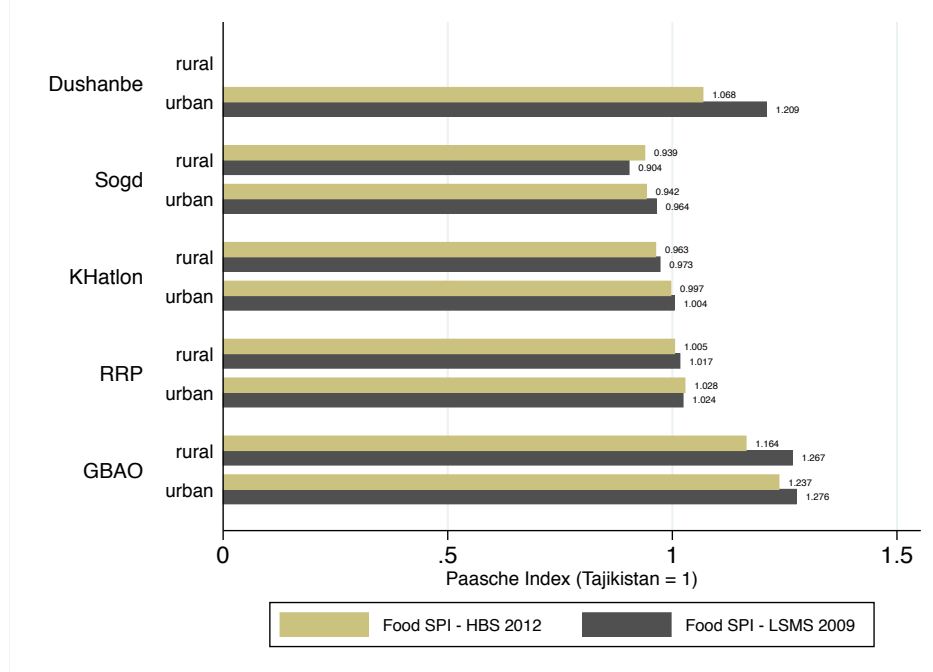
**Table 21 – Outliers (%) for unit values of nonfood items**

PRODUCT	Quarter		Total
	2	3	
Boots, boots rubber	6.2	4.2	5.2
Combinations, shirts, jerseys, t-shirts	0.5	0.0	0.2
Costumes	0.0	0.5	0.2
Cotton	5.3	3.1	4.2
Cross shoes	2.7	3.4	3.1
Dresses, dresses, suits	0.7	1.8	1.2
Gasoline	0.3	1.6	1.0
Hosiery	2.2	3.2	2.7
Jackets, sweaters, jumpers, pullovers, j	3.4	2.1	2.8
Laundry soap	0.7	1.4	1.0
Other Footwear	1.2	0.6	0.9
Other types of clothing	1.2	0.7	0.9
Other underwear	2.4	0.0	1.2
Pants	1.9	1.1	1.5
Pants, pants, trousers	0.0	0.8	0.4
Piece of tissue products	1.2	2.5	1.9
Robes, PJ	2.1	2.7	2.4
Sandals	1.7	1.2	1.4
Shawls, scarves	1.3	1.4	1.4
Shirts for men, women's blouses	2.0	1.7	1.9
Shoe room, booties motocross	2.8	1.9	2.3
Shoes	2.1	2.6	2.4
Skirts	6.2	0.0	3.1
Soap	1.8	1.5	1.7
Telephone	0.0	4.0	2.0
Tracksuits	0.0	1.4	0.7
fabrics for clothing: cotton	1.8	1.2	1.5
fabrics for clothing: silk	1.8	0.0	0.9
fabrics for clothing: staple	2.8	0.7	1.7
fabrics for clothing: synthetic	4.2	0.8	2.5
tapes for the VCR and tape	1.0	1.9	1.4
Total	1.0	0.8	0.9

Note: total number of observations is 627,601.

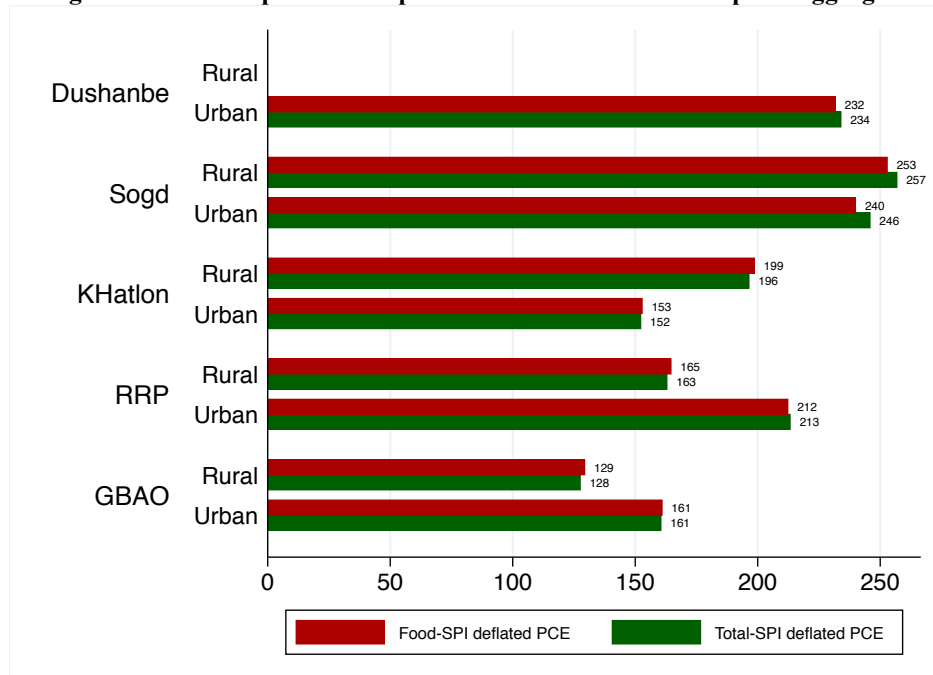
## APPENDIX C FURTHER EVIDENCE ON SPATIAL DEFLATORS

**Figure 15 – 2012 HBS-based Food Paasche Index vs. 2009 LSMS-based food Paasche Index**



Source: Authors' estimates.

**Figure 16 – The impact of the spatial deflator on the consumption aggregate**



Note: PCE denotes “per capita expenditure”, expressed in Somoni/person/month. Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> and 3<sup>rd</sup> quarters).

## APPENDIX D

### FOOD CALORIE CONVERSION FACTORS

**Table 22 – Food calorie conversion factors**

<b>Product</b>	<b>kilocalories (per Kg or unit)</b>
Wheat bread	2890
Rye bread and other	2200
Wheat Flour	3400
Rye flour	3110
Flour, other	3400
Legumes	2900
Rice	3500
Cereals, except rice	3400
Pasta	3500
Potatoes	850
Cabbage	240
Cucumbers	140
Tomatoes	190
Beets, carrots, radishes, other edible roots	350
Onion	260
Other vegetables	250
Canned vegetables	400
Zucchini, squash, pumpkins, etc.	180
Watermelons and melons	260
Fruit Garden	450
Berries Garden	386
Wild Fruits and Berries	400
Dried Fruits and Berries	6000
Fresh grapes	610
Dried grapes	2830
Citrus	360
Fruit preserves	600
Beef and veal	1200
Mutton and goat meat	1530
Meat by-products	8910
Poultry	1710
The meat of other animals	1500
Fat	8910
The meat of wild animals and birds	1200
Meats and myasokopchenosti	450
Meat products and ready-made from meat	2575
Meat and canned myasorastitelnaye	2575
Whole milk	650
Skim milk	360
Sour cream and cream	3370
Butter	8500
Curd, weight, etc. syrkovaya	3600
Cheese and cheese	3600
Canned milk	640
Ice cream	1900
Vegetable oil	8840
Margarine and other fats	7660
Sugar	3920
Candy, halva, etc.	4380
Chocolate	5100
Honey bee	3040

<b>Product</b>	<b>kilocalories (per Kg or unit)</b>
Jam, marmalade	2220
Cookies, cakes, etc.	3530
Eggs	53
Fish, fresh	990
Fish, salted, dried herring including	1280
Canned fish	1250
Fresh mushrooms	280
Dried mushrooms	1200
Tea	60
Coffee natural	300
Soft drinks	380
Salt	0
Vodka	2400
Liqueur	2400
Fruit and berry wine factory	5500
Fruit and berry wine homemade	3300

Source: FAO (2004) and the Italian National Institute of Nutrition ([www.inran.it](http://www.inran.it)).

## APPENDIX E

### HBS 2012 DATA QUALITY ASSESSMENT

In this Appendix we investigate data quality issues. A previous investigation concluded that neither 2009 nor 2010 HBS datasets provide a reliable basis for carrying out welfare analysis in Tajikistan. We analyze the data from the 2<sup>nd</sup> quarter of the 2012 HBS with the aim of testing whether the data suffer from the same problem.

Table 12 in Section 3 highlighted the fact that inequality in both Dushanbe and GBAO seems somehow low with respect to the values observed in other regions. To further investigate this issue our strategy is to use the raw data, and focus on elementary variables (as opposed to constructed variables such as the consumption aggregate). We begin our investigation focusing on *form 1, section 4, question 3* (expenditure on purchased items).

The simplest approach consists in examining the data for a single food item (e.g., wheat flour) in one region (e.g., Dushanbe) using the dataset at the product-level.

Out of the 70 food items for which expenditures are available, we pick *wheat flour* (coicop code 403), the single most important item in the food bundle of Tajikistan's households. The HBS 2012 2<sup>nd</sup> quarter data contains 1,001 records for purchases of wheat flour in urban areas (Table B1).

**Table 23 – Wheat flour (code 403), sample size by region (urban areas only).**

Region	N. obs.
Dushanbe	366
Sogd	243
KHatlon	232
RRP	112
GBAO	48
Total	1,001

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> quarter).

The variability of mean expenditures on wheat flour by region is shown in Table 24.

**Table 24 – Inequality in expenditure on wheat flour.**

Region	GE(-1)	GE(0)	GE(1)	GE(2)	Gini
Dushanbe	0.27274	0.12969	0.0966	0.09134	0.21324
Sogd	0.38589	0.24571	0.20312	0.21288	0.3332
KHatlon	0.21139	0.11395	0.09489	0.09543	0.22394
RRP	0.15264	0.14174	0.14769	0.17231	0.29312
GBAO	0.06156	0.06269	0.06552	0.07031	0.19096

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> quarter).

Inequality in both Dushanbe and GBAO is slightly lower than in other regions, but the pattern does not seem robust to the choice of the inequality index. A similar story holds true for purchased **quantities** (form 1, section 4, question 2). Table B3 illustrates.

**Table 25 – Inequality in purchased quantities of wheat flour.**

Region	GE(-1)	GE(0)	GE(1)	GE(2)	Gini
Dushanbe	0.29645	0.13537	0.09841	0.09198	0.20085
Sogd	0.43716	0.2669	0.21339	0.2149	0.33115
Khatlon	0.26557	0.12334	0.09851	0.09795	0.21757
RRP	0.15621	0.14334	0.14836	0.17237	0.28984
GBAO	0.05504	0.0565	0.05928	0.06361	0.17482

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> quarter).

Overall, and in contrast to what was found for the 2009 HBS, there is no evidence in support of the hypothesis of an abnormally low variance in distribution of expenditures and quantities of wheat flour in Dushanbe and GBAO.

The evidence in Table 24 and Table 25 shows that the lowest variability is observed in urban GBAO; for this reason we now focus on GBAO and consider variable 3 in form1, section 4 (expenditure on purchased items), code 403 (wheat flour).

**Table 26 – Frequency distribution of expenditures on wheat flour in urban GBAO**

Expenditure (somon/month)	frequency	%
11309	14	29.2
12784	4	8.3
13276	10	20.8
13768	5	10.4
22618	3	6.3
24585	2	4.2
25568	1	2.1
26552	7	14.6
27535	2	4.2
Total	48	100.0

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> quarter).

Table 26 shows the frequency distribution of expenditures for wheat flour in urban GBAO. Column 1 shows the distribution of *distinct values* of expenditures on wheat flour observed in the sample. For instance, the expenditure level 11,309 (Somon/person/month) is observed 14 times; the value “12,784” is observed 4 times, and so forth.

One way to summarize the information in Table 26 is proposed in Table 27. Table 27 turns out to be a useful tool to identify anomalies in the distribution of expenditures of a selection of items (the 10 most important food items (in terms of national budget shares) across regions.<sup>9</sup> For each item in each table we have examined the number of distinct values and have calculated how many times the same values have been replicated. This has allowed us to calculate the “average duplication rate”, separately for each region and each item. To illustrate, consider Table 27, row 1. The ADR

<sup>9</sup> In order to compare ADRs across regions we should (i) standardize the ratios (standardization is required to make ADR independent of the number of regional observations) and (ii) adjust for heaping. Both adjustments go beyond the scope of this note.

for wheat flour in Dushanbe equals 4; this implies that in Dushanbe – on average – a single value for expenditure on wheat flour is repeated 4 times. The ADR in GBAO is 5. The same index ranges between 3 and 4 in the remaining regions.<sup>10</sup> The last row contains a weighted average of the ADR indices by region (the weighting system is based on the national budget shares in column 3). We interpret the weighted average as an overall index of data replication: the higher its value, the more likely is the chance that data repetition has taken place in a given region.

**Table 27 – Average Duplication Rates by region and urban-rural areas**

urban			ADR (average duplication rate)				
product	coicop code	budget share	Dushambe	Sogd	Khatlon	RRP	GBAO
wheat flour	403	16.43878	4	4	4	3	5
beaf	428	11.02809	5	4	6	3	6
vegetable oil	447	8.819638	4	3	4	2	5
potatoes	410	5.711864	5	4	5	2	2
rice	407	5.407801	6	4	5	3	5
sugar	449	4.659647	5	4	4	2	7
wheat bread	401	3.092508	2	3	2	1	0
onions	415	3.086171	6	6	4	2	4
milk	439	3.066482	5	5	4	2	6
orchard fruits	420	2.843611	3	4	2	3	2
average bundle		64.154592	4.5	4.0	4.3	2.5	4.2

rural			ADR (average duplication rate)				
product	coicop code	budget share	Dushambe	Sogd	Khatlon	RRP	GBAO
wheat flour	403	23.50634		6	9	5	8
vegetable oil	447	11.25893		5	7	5	5
beaf	428	10.67387		7	9	6	6
rice	407	7.088975		5	9	6	7
sugar	449	5.639004		5	10	5	10
potatoes	410	5.192524		7	8	4	3
onions	415	3.45975		9	11	6	7
candy, halva, etc	450	3.392347		5	9	6	4
pasta	409	2.763679		11	15	8	6
poultry	432	2.540397		5	9	5	7
average bundle		75.515816		6.1	9.0	5.4	6.3

Source: Authors' estimates on Tajikistani HBS 2012 (2<sup>nd</sup> quarter).

Rural Khatlon shows an above-average ADR, but we found no evidence of data duplication. The analysis, not reported here, is available from the authors.

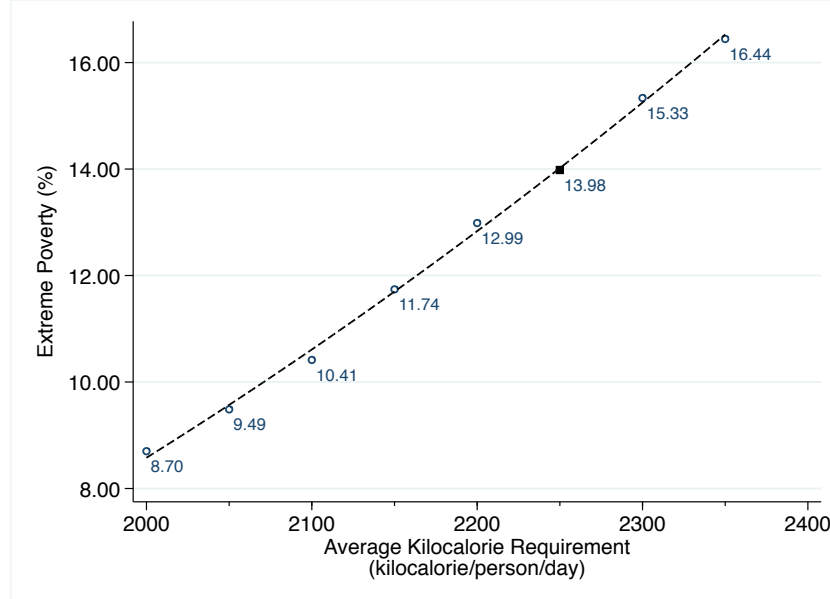
<sup>10</sup> Caution is required in interpreting GBAO's figures, owing to potential small sample size problems.

## APPENDIX F SENSITIVITY ANALYSIS OF POVERTY ESTIMATES

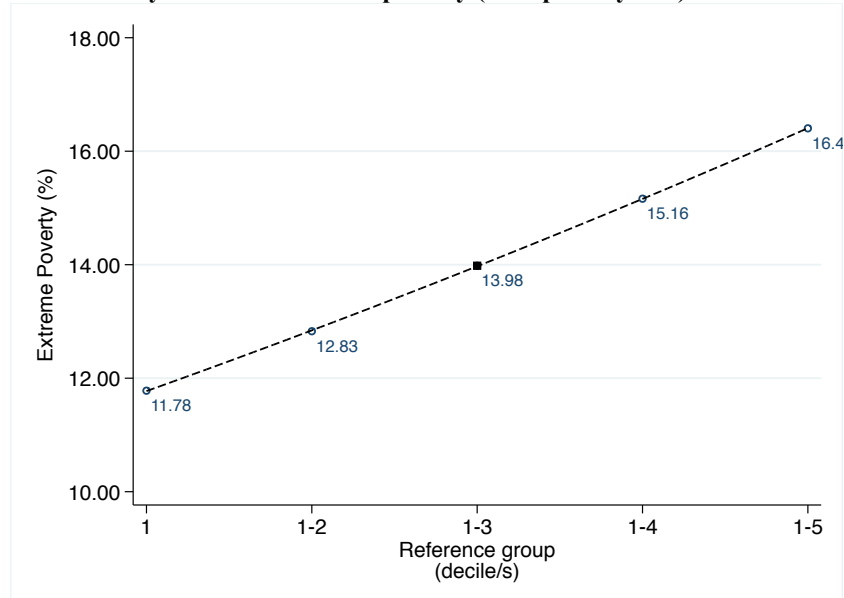
**Figure 17** shows the sensitivity of the incidence of extreme poverty (that is, the national headcount poverty ratio corresponding to the *food* poverty line) to changes in the average calorie requirements (AKR).

**Figure 18** shows the sensitivity of the incidence of extreme poverty to changes in the PCE deciles chosen as a reference group, that is the deciles used to calculate the calorie unit cost.

**Figure 17 – Sensitivity of the incidence of poverty (food poverty line) to the AKR**

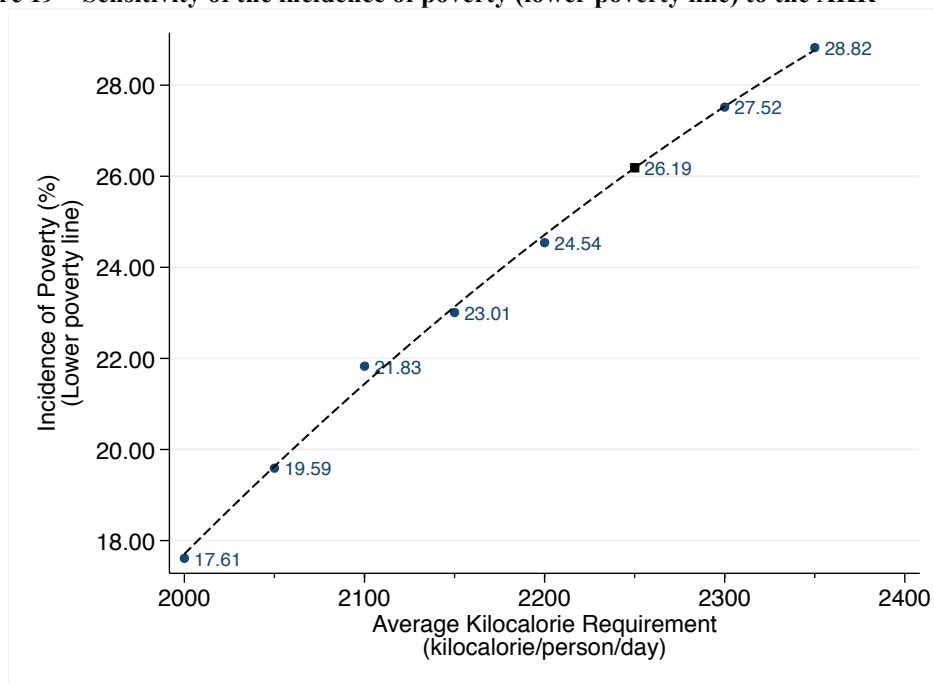


**Figure 18 – Sensitivity of the incidence of poverty (food poverty line) to the reference group**

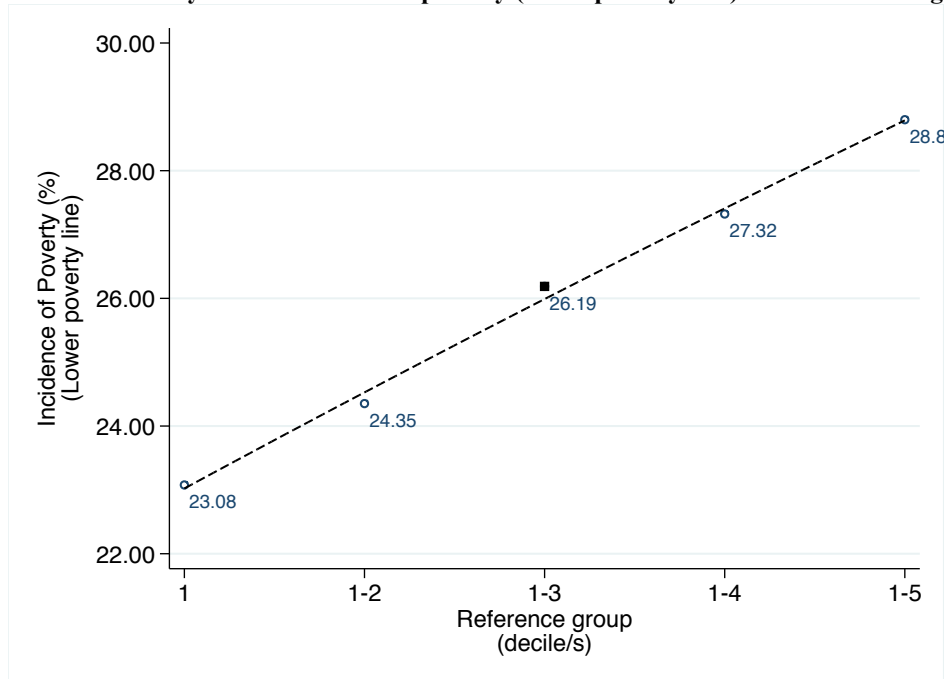




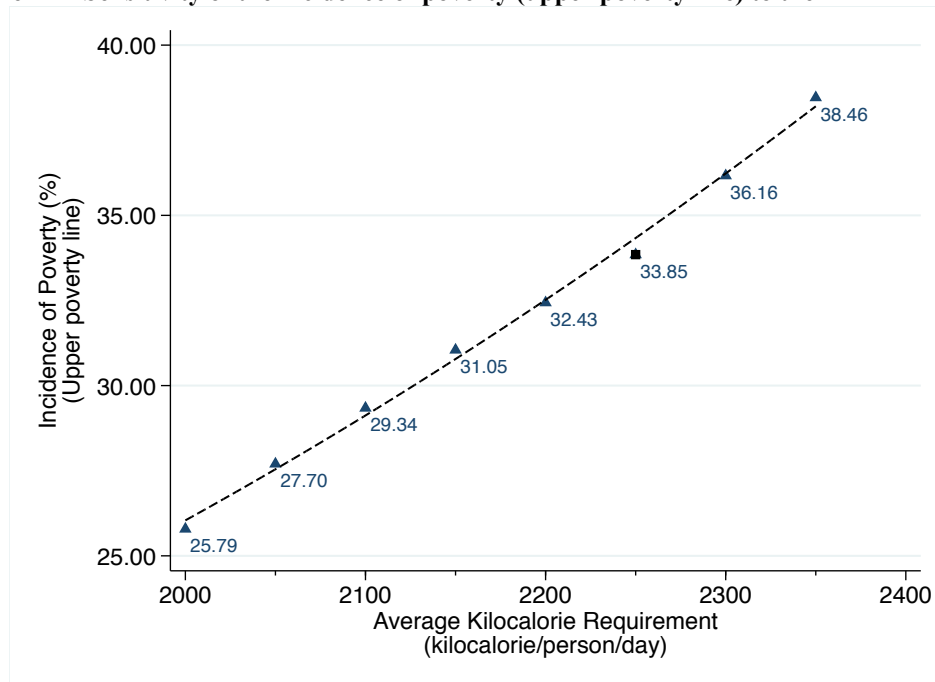
**Figure 19 – Sensitivity of the incidence of poverty (lower poverty line) to the AKR**



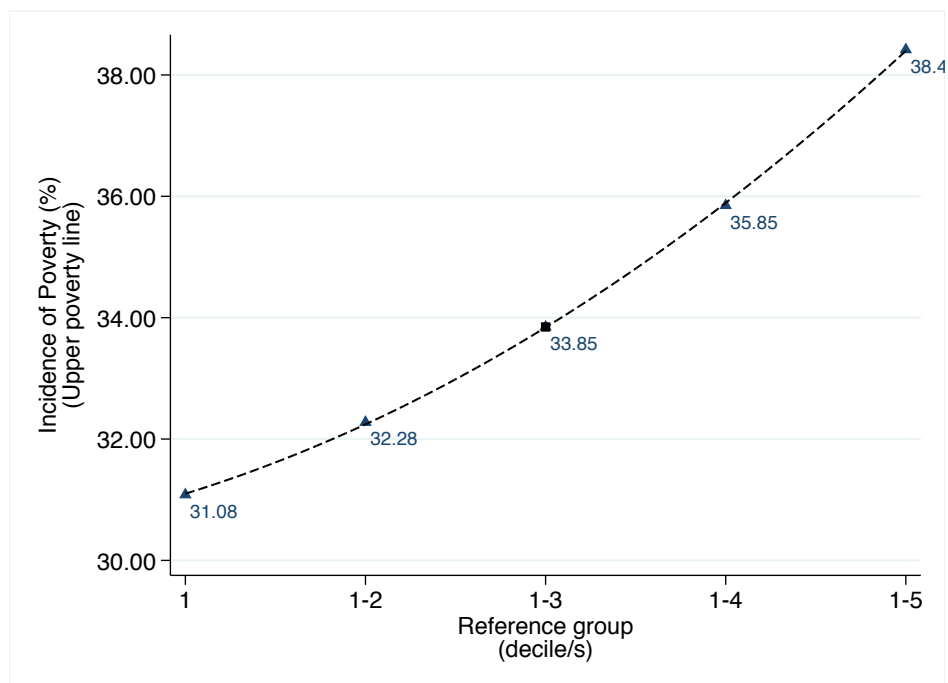
**Figure 20 – Sensitivity of the incidence of poverty (lower poverty line) to the reference group**



**Figure 21 – Sensitivity of the incidence of poverty (upper poverty line) to the AKR**



**Figure 22 – Sensitivity of the incidence of poverty (upper poverty line) to the reference group**



## APPENDIX G

### QUARTERLY POVERTY RATES

**Figure 23 – Headcount poverty ratio by quarters in selected ECA countries**

		QUARTERS							
	year	1	2	3	4	Total	Q2/Q1	Q3/Q2	Q4/Q3
<b>KAZ</b>	2009	0.572	0.634	0.591	0.499	0.574	1.11	0.93	0.84
<b>RUS</b>	2009	0.150	0.146	0.132	0.103	0.133	0.97	0.90	0.78
<b>GEO</b>	2010	0.547	0.558	0.543	0.511	0.540	1.02	0.97	0.94
<b>KRG</b>	2011	0.395	0.428	0.362	0.256	0.361	1.08	0.85	0.71
<b>TAJ</b>	2012		0.403	0.274				0.68	
<b>URBAN</b>		QUARTERS							
		1	2	3	4	Total	Q2/Q1	Q3/Q2	Q4/Q3
<b>GEO</b>	2009	0.411	0.457	0.510	0.441	0.455	1.11	1.12	0.86
<b>KAZ</b>	2009	0.359	0.384	0.379	0.284	0.351	1.07	0.99	0.75
<b>KRG</b>	2010	0.229	0.292	0.265	0.172	0.240	1.28	0.91	0.65
<b>RUS</b>	2011	0.082	0.077	0.072	0.051	0.071	0.94	0.94	0.71
<b>TAJ</b>	2012		0.337	0.249				0.74	
<b>RURAL</b>		QUARTERS							
		1	2	3	4	Total	Q2/Q1	Q3/Q2	Q4/Q3
<b>GEO</b>	2009	0.676	0.653	0.576	0.578	0.621	0.97	0.88	1.00
<b>KAZ</b>	2009	0.602	0.701	0.626	0.533	0.616	1.16	0.89	0.85
<b>KRG</b>	2010	0.514	0.527	0.432	0.317	0.448	1.03	0.82	0.73
<b>RUS</b>	2011	0.229	0.229	0.201	0.165	0.206	1.00	0.88	0.82
<b>TAJ</b>	2012		0.436	0.287				0.66	

Note and sources: “KAZ” (Kazakistan): 2009 HBS; “RUS” (Russia): 2009 HBS; “GEO” (Georgia): 2010 HBS; “KRG” (Kyrgyzstan): 2011 HBS.

## References

- Araar A. and J.-Y. Duclos (2007), “*DASP: Distributive Analysis Stata Package*”, PEP, World Bank, UNDP and Université Laval.
- Citro, C.F. and R.T. Michael, (eds.) (1995), *Measuring Poverty: A New Approach*, Washington, National Academy Press.
- Deaton, A. (1980), “The Measurement of Welfare. Theory and Practical Guidelines”, *Living Standards Measurement Study Working Paper n. 7*. The World Bank, Washington, DC.
- Deaton A. (1997), *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*, World Bank, Baltimore, Johns Hopkins University Press.
- Deaton A., and A. Tarozzi (2008)
- Deaton A. and S. Zaidi (2002), “Guidelines for Constructing Consumption Aggregates for Welfare Analysis.” *Living Standards Measurement Study Working Paper n. 135*. The World Bank, Washington, DC.
- FAO (2004), *Human Energy Requirements; Report of a Joint FAO/WHO/UNU Expert Consultation*. FAO Food and Nutrition Technical Report Series No.1. Food and Agriculture Organization: Rome.
- Kakwani, N. (2003), “Issues in Setting Absolute Poverty Lines”, *Poverty and Social Development Papers*, no. 3, Asian Development Bank.
- Morduch, J. (1994), “A Positive Measure of Poverty”, *Development Discussion Paper No. 478*, Harvard Institute for International Development.
- Morduch, J. (1998) “Poverty, economic growth, and average exit time”, *Economics Letters*, 58: 385-90.
- Pradhan, M., A. Suryahadi, S. Sumarto, and L. Pritchett (2000), “Measurements of Poverty in Indonesia. 1996, 1999, and Beyond”, *Policy Research Working Paper no. 2438*, The World Bank.
- Ravallion, M. (2012), “Poverty Lines Across the World”, in P.N. Jefferson (ed.), *The Oxford Handbook of the Economics of Poverty*, Oxford University Press, ch. 3, pp. 75-104.
- Ravallion, M. (1994), *Poverty Comparisons*, Harwood Academic Publishers.

- Ravallion, M. and B. Bidani, B. (1994), "How robust is a poverty profile?", *World Bank Economic Review*, 8, 1: 75-102.
- Ravallion, M. and S. Chen (1996) "Data in transition: Assessing Rural Living Standards in Southern China", *China Economic Review*, 7, 1: 23-56.
- Samuelson, P. (1974) "Complementarity. An Essay on the 40<sup>th</sup> Anniversary of the Hicks-Allen Revolution in Demand Theory", *Journal of Economic Literature*, 12, 4: 1255-89.
- Sulla, V. and J. Munoz (2011), "Tajikistan Household Budget Survey. Recommendations of a World Bank Mission", mimeo.
- World Bank (2009), *Republic of Tajikistan. Poverty Assessment*. Washington: The World Bank.
- Zheng, B. (1993), "An Axiomatic Characterization of the Watts Poverty Index", in *Economics Letters*, 42: 81-6.
- Zheng, B. (1997), "Aggregate Poverty Measures", in *Journal of Economic Surveys*, 11, 2: 123-162.