Food and Agriculture Organization of the United Nations

FAO Statistics Division

Working Paper Series

ESS / 14-08

ESTIMATING FOOD CONSUMPTION PATTERNS BY RECONCILING FOOD BALANCE SHEETS AND HOUSEHOLD BUDGET SURVEYS

December 2014

ESTIMATING FOOD CONSUMPTION PATTERNS BY RECONCILING FOOD BALANCE SHEETS AND HOUSEHOLD BUDGET SURVEYS

Klaus Grünberger

Food and Agriculture Organization of the United Nations
Rome, 2014

Estimating Food Consumption Patterns by Reconciling Food Balance Sheets and Household Budget Surveys

Klaus Grünberger

Consultant, Statistics Division, FAO

Abstract

Food Balance Sheets (FBS) are one of the most important sources of data on food availability for human consumption. This paper presents a method to improve the information on food consumption patterns of FBS by using national household budget surveys (HBS).

In this paper, food commodities are categorized into 16 major food groups. For each food group, the contribution to the overall caloric intake is represented in shares. Item group shares of 64 surveys from 51 low and middle income countries are compared with shares from country-specific FBS. Given the countries represented in the data, the analysis evaluates food consumption of over 3 billion persons worldwide.

A model based on a cross-entropy measure of information has been developed in order to reconcile aggregate food consumption patterns suggested by FBS and HBS. The latter model accounts for the fact that data from both data sources are prone to measurement errors.

Overall, the results of the reconciliation suggest that average consumption of cereals, eggs, fish products, pulses and vegetables are likely to be underestimated in FBS, while fruits, meat, milk and sugar products are likely to be overestimated in FBS. Even though the suggested changes in average food consumption are moderate, the results imply considerable relative changes in the aggregate consumption of single food groups. Furthermore, the results imply that the aggregate consumption of fats is 2% higher than currently assumed.

The updated consumption patterns provide valuable information from an agro-industrial perspective. Differences in updated consumption pattern with respect to the original FBS might suggest a re-evaluation of FBS elements of the value chain, starting from production and ending at food losses.

Key words: Food Balance Sheets, Household Budget Surveys, Generalized Cross-Entropy Estimation, Food Consumption

JEL codes: I12, L66, O13, Q11, Q18

1 Introduction

Food balance sheets (FBS) provide a comprehensive picture of national food supply and are of fundamental importance to measure global food security (FAO et al., 2013). Furthermore, FBS are widely used in research of agriculture, nutrition and public health. In FBS, food availability for human consumption is calculated by taking into account production, imports, exports, stock variation and utilization elements such as feed, seed, losses and industrial uses. In most countries it is a big challenge to collect accurate data about all the elements and the quality of the data cannot always satisfy high quality standards. As a result, FBS figures of food consumption are prone to measurement errors.

This paper presents a method that seeks to improve the quality of FBS by making use of 64 national household budget surveys (HBS) from 51 low and middle income countries. Altogether, these 51 countries have over 3 billion inhabitants. Hence, the analysed data represent the average food consumption of more than 40% of the world population.

While FBS provide data from a macro perspective, HBS are looking at food availability at the micro level. Each HBS collects information on household food acquisition or consumption[1] of a presumably representative sample of the country's population. However, like FBS, also HBS have their specific problems in providing a comprehensive picture of a country's food consumption. For this reason information on food is consolidated by combining the strength of both data sources.

Food data from HBS and FBS are aggregated into 16 major commodity groups (cereals, fruits, etc.) and compared in terms of calories. The reconciliation of HBS and FBS data will be performed on the basis of calorie shares of food item groups, i.e. the contribution of each food item group to the total calorie consumption. Hereby an estimation procedure based on a cross-entropy measure will be employed, allowing for measurement errors in HBS shares.

The procedure produces updated FBS with adjusted item group calorie shares. Overall, the results suggest that the consumption of cereals, eggs, fish products, pulses and vegetables might be higher than supposed by the original FBS, while the consumption of fruits, meat, milk, starchy roots and sugar products might be lower. The reconciliation results can be taken as benchmark for reviewing consumption patterns in FBS figures.

In Section 2 describes the data and Section 3 presents the reconciliation model. Results are presented in Section 4 and Section 5 concludes.

[1] In this paper 'consumption' is not defined as food eaten (commonly assessed by specific nutrition surveys), but as food available for actual human consumption. Moreover, the definitions of food consumption slightly differ between FBS and HBS. Section 2 will discuss the data in detail.

2 The data

Both, FBS and HBS, have their own strengths and weaknesses. Neither of the two data sources can be regarded as a ’gold standard’ for making inference about the food consumption in a country. When comparing FBS and HBS it is important to identify their main shortcomings and to highlight the difference between the two data sources.

2.1 FBS data

FBS are aggregate data which include information about food commodities for more than 180 countries and about 100 food items. The main components of FBS are provided by national statistical offices. In case of missing data, imputation techniques are used to fill the data gaps. FBS reflect the countries’ food supply during a specified reference period by subtracting utilization from supply. Domestic food supply is given by the sum of production (already harvested crops), stock variation and the foreign trade balance. Domestic utilization consists of the following elements: feed, seed, post-harvest losses and industrial uses. FBS’s food supply reflects therefore food available for human consumption, without considering food wasted at the household level. Furthermore, there is some evidence that in FBS, food losses during distribution at the retail level might not be sufficiently captured (Naska et al., 2009; Grünberger, 2013 and Sibrián et al., 2006). As a result, FBS’s food supply might overestimate the amount of food available for human consumption.

Each component of both sides, elements of supply as well as from utilization, are prone to measurement error. Since the estimate of food consumption is in most cases derived as a residual of these (often highly uncertain) elements, its reliability is often called into question (Jacobs et al., 2002).[2]

2.2 HBS data

In principle, household data should solve most of the problems related to the measurement of countries’ food availability. A perfectly representative household survey which assesses food consumption over a whole calendar year might be regarded as a ’gold standard’ for the measurement of food availability. However, in practise, surveys have their own shortcomings too.[3]

It begins with the fact that most surveys are designed as household expenditure surveys and are not primary planned for measuring food expenditure alone, but measure all types of consumption. Since they do not primarily focus on nutrition, the picture of the country’s food consumption is often incomplete. In fact, it is difficult to obtain complete information of what an entire household has consumed. The reference person who responses to the survey questionnaire may either have incomplete information about the consumption of

[2] FAO is currently working on a new framework to compile FBS (see Mahjoubi et al., 2012).
[3] A comprehensive overview of household surveys and their use for estimating food consumption is provided in Smith et al. (2014).

other household members, or simply forgets to mention apparently insignificant items (like small snacks).

Some surveys assess food consumption, while others measure food acquisition. Consumption comprehends all purchases, auto-produced food and stock withdrawals minus food put into stocks. Expenditure/acquisition surveys take into account auto-produced food but do account for food put into, or taken out of stocks.[4] However, after taking the average of all observations, both, consumption and expenditure surveys, should get the same result. For that reason, in the following discussion, no difference will be made between food consumption and food purchase.

Even if most HBS assess food consumed away from home, it can be expected that not all food flowing into the non-household sector (restaurants, canteens, schools, etc.) is captured. Furthermore, food waste in the non-household sector is a factor that cannot be assessed in HBS. As a result, food consumption in countries with a high fraction of food away from home might be underestimated in HBS.

Finally, representation of the country's population and the coverage of seasonal variation might be incomplete in the HBS. A lack of representativeness of timing is therefore an additional source of measurement error of HBS.

On the other hand, a clear advantage of household surveys is that own consumption is explicitly captured in most HBS. Under the definition of own consumption are falling domestic produces and food gathered wild. In many countries own consumption represents a relevant fraction of countries total food consumption. By definition of the FBS, own consumption should be already included in countries' food production. However, collecting data on own consumption is difficult and FBS might not always sufficiently capture these factors in its accounts. In this respect, HBS may provide valuable information to identify possible measurement errors in FBS production figures.

This study uses 64 HBS from 51 low and middle income countries. The sample of surveys covers all major regions of the developing world: Caribbean (1), Central America (4), Central Asia (1), Eastern Africa (8), Eastern Europe (3), Melanesia (1), Middle Africa (2), Northern Africa (2), Northern Europe (1), South-Eastern Asia (6), South America (7), Southern Asia (5), Southern Europe (1), Western Africa (7) and Western Asia (3).[5] A table listing all surveys can be found in Appendix A.

[4] In both cases household waste is not detracted.

[5] The number of countries represented in the sample is reported in parenthesis. Sub-regions are categorized according to the United Nations geoscheme (based on the M49 coding classification). The latter categorization assigns, in contrast to other geoschemes, Lithuania to Northern Europe and Albania to Southern Europe.

2.3 Data construction

Household surveys used for this analysis are converted to the ADePT[6] format and contain detailed information on household's food consumption. Nutritional properties of the food items are retrieved by food composition tables. Food composition tables assign calories and macronutrients to each food item and differ from country to country. In case of generic food purchases, like undefined meals away from home, calories and macronutrients are imputed (Moltedo et al., 2014).

It is not possible to compare commodity-specific data from HBS and FBS, because an exact matching of specific food items between HBS and FBS is not possible. Item descriptions in HBS are often very generic and cannot meet the precision of FBS. However, food items of household surveys are categorized in 19 major food groups. Since the classification of food groups of HBS is similar to those of FBS, food can be compared on this level of aggregation. Table 1 shows how food groups of HBS and FBS are harmonized to a common categorization. Finally we end up with 16 common food groups on whose basis the comparison will be performed.

Table 1: Formation of item groups

(1) **FBS Item Groups**	(2) **HBS/ADePT Item Groups**		(3) **Common Groups**
Cereals	Cereals		Cereals
Starchy Roots	Roots & Tubers		Starchy Roots
Sugar Crops	Sugar & Syrups	→	Sugar & Products
Sugar & Sweeteners	Soft drinks		
Pulses	Pulses		Pulses
Tree Nuts	Tree Nuts	→	Oil Crops & Tree Nuts
Oil Crops	Oil Crops		
Vegetable Oils	Vegetable Oils		Vegetable Oils
Vegetables	Vegetables		Vegetables
Fruits - Excl. Wine	Fruits		Fruits
Alcoholic Beverages	Alcoholic Beverages		Alcoholic Beverages
Meat	Meat & Offals	→	Meat & Offals
Offals			
Eggs	Eggs		Eggs
Milk - Excl. Butter	Milk & Cheese		Milk & Cheese
Animal Fats	Animal Fats		Animal Fats
Fish, Seafood	Fish & Seafood	→	Fish & Seafood
Aquatic Products			
Stimulants	Stimulants		Stimulants
Spices	Spices		Spices
Miscellaneous	Food Preparations	(Distributed to other items groups)	

Note: The lists of single items included in food groups can be consulted in FAO (2001) and Moltedo et. al (2014). Food from category 'Miscellaneous' or 'Food Preparations' respectively has been distributed proportionally to all other food groups. Oil crops and tree nuts are reduced to a single food group, because either oil crops, or tree nuts was often missing in the HBS.

[6] ADePT is a 'Software Platform for Automated Economic Analysis' (for further information see http://econ.worldbank.org). It has a module on food security which was developed by the FAO Statistics Division and the World Bank.

National per capita food consumption of HBS is calculated by dividing food consumption of each household by the number of household members and then taking the mean (applying sample weights). Per capita food consumption of FBS is available on the FAOSTAT website.[7] If a survey was conducted over several years, a weighted average of the FBS of the corresponding years has been calculated. Hereby the FBS of a certain year has been weighted according to the size of population which it is represented in the corresponding year of the survey.

2.4 Comparison of HBS and FBS

There are various ways how to compare two data sources like HBS and FBS. First, food can be compared either by single food items, or by major food groups. As previously argued, the comparison of food groups is preferred because of the problems regarding the identification of specific food items in the surveys.

Second, in this study the amount of food will be measured in calories, instead of weight. The reason for this choice is that food groups might include food commodities which are heterogeneous in terms of moisture content and calorie density. If a measure based on weight would be considered, commodities with a low calorie density would influence strongly the outcome of the whole food group (e.g., water melons for fruits). A measurement error of this low calorie-dense commodity would therefore heavily distort the measurement of the whole food group. In this respect, calories provide a more robust metric than weight (assuming that conversion into calories is done properly).

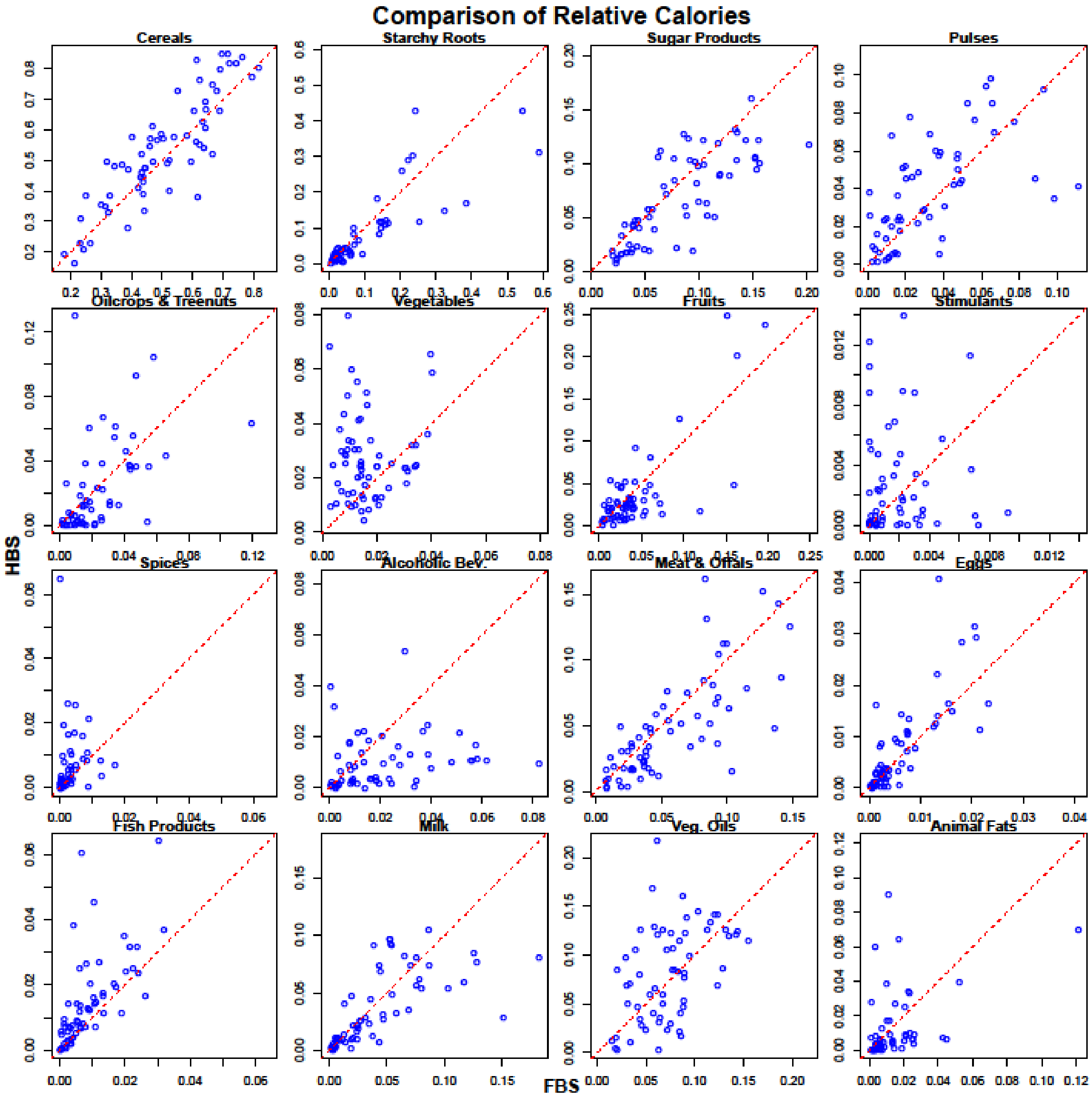
Furthermore, there is an additional, more practical, reason why a calorie metric should be applied. In cases where only monetary values for a food item are provided, calories can be imputed by applying a single price per calorie. This is a big advantage in cases where the food item is a prepared meal that contains various ingredients.

A drawback of dealing with calories is the fact that surveys have their own food composition tables. The latter tables impose a calorie calculation which might be different from the calculation underlying the calorie data in FBS. This might add some additional noise to the comparison.

Third, food can be compared either in absolute terms, i.e. the number of calories per food group, or in relative terms, i.e. the food group's contribution to the total amount of calories. As aforementioned, the levels of food are not exactly comparable between HBS and FBS. As a consequence only food in relative terms will be considered in the following analysis.

[7] http://faostat.fao.org/site/368/default.aspx#ancor. Downloaded on 16 July 2014. For the Democratic Republic of Congo and Papua New Guinea, unofficial data from the SUA Working System are used.

Figure 1: Comparison of calorie shares by item group



Note: Scatterplots show item group shares of all (64) available HBS/FBS pairs. The y-axis refers to HBS shares and the x-axis to FBS shares. The red dashed 45 degree line defines the points at which shares of the two sources are equal.

2.4.1 Comparison of relative calories of HBS and FBS

The further analysis is restricted to calories in relative terms. With relative terms is intended the food group's contribution to the total amount of calories (they will be called 'shares'). The formula below shows how the share of an item group *i* is calculated, where N is the set of all 16 item groups:

$$SHARE_i = \frac{kcal_i}{\sum_j^N kcal_j} , \qquad \forall i \ \in N.$$

An analysis based on shares is no longer sensitive to systematic differences in the level of calories. It just requires that food groups are equally well represented in the HBS. In case that an item group is poorly represented in the survey this would certainly distort the analysis. An evident example of a poorly represented food group in the HBS, are alcoholic beverages. In almost all countries respondents are expected to (either consciously or unconsciously) underreport alcohol consumption. Looking at the scatterplot of alcoholic beverages in Figure 1, one can note that in most cases the share of calories is dramatically lower according to HBS than to FBS. Because of the evident lack of reliability, alcoholic beverages are omitted from the analysis.

Apart from alcoholic beverages, there are other item groups whose consumption is generally not well captured in the HBS too. Figure 1 shows that the distribution of spices, additives and stimulants are very disperse, with a large fraction of very low HBS shares. This cast doubts on the reliability of the later data. Fortunately this two food groups contribute a minor amount of calories and can be omitted from the analysis without any concern.

Additionally oil crops, vegetable oils and animal fats seem to be poorly surveyed in many cases. The latter item groups are characterized by a large fraction of cases with very low HBS share and a high variability. They are therefore as well excluded from the analysis, even if vegetable oils are a big contributor to the overall dietary energy.[8]

The HBS shares of meat and sugar are in most cases lower than the FBS shares. The HBS shares of sugar may be too low because sugar snacks might be included in the category of miscellaneous food, or forgotten to be mentioned in the questionnaire. As argued in FAO (1983) sugar products 'are often purchased by members of the family without coming to the notice of the person how keeps the records'. Additionally, soft drinks might be often part of prepared meals consumed away from home. Since the calories from prepared food are distributed among all food items, the share of sugar might be additionally underestimated. The latter argument holds also for meat, which might be often part of prepared meals, especially when it is consumed outside home. Even though, sugar products and meat are not excluded from the analysis. Instead, the model presented in Section 3 will control for the bias of the latter item groups.

Ultimately, HBS data from oil crops, tree nuts, spices, alcoholic beverages, fish products, vegetable oils and animal fats are not considered in the further analysis. Furthermore, if for a given survey the average consumption of an item group is less than 5 kcal, the item group is excluded from the comparison. Finally, only cereals, starchy roots, sugar products, pulses, vegetables, fruits, meat & offal, fish products and milk are considered in the reconciliation which will be presented in the following chapter.

Given that some food groups are excluded from the analysis, food group shares have to be calculated on the basis of included items only. These 'partial' food group shares are

[8] Food group shares have been analyzed for correlations between each other. For a certain survey, a low HBS share of a food group (low with respect to the FBS share) might be systematically linked to a higher HBS share of another food group. E.g. shares of oil crops and vegetable oils (i.e. the processed product of many oil crops) might be correlated. However, no significant correlation could be found in this regard. Results are available on request.

calculated as indicated in the formulas below, where H_i and $\tilde{F}_i$ represent item group *i*'s (partial) shares of HBS and FBS respectively. *HBSkcal* and *FBSkcal* are the calories of the food group and K is set of included food groups.

$$H_i = \frac{HBSkcal_i}{\sum_j^K HBSkcal_j} \text{ and } \tilde{F}_i = \frac{FBSkcal_i}{\sum_j^K FBSkcal_j}, \qquad \forall i \in K.$$

Table 2 shows averages of caloric intake per person/day, the item group shares and the partial shares of HBS and FBS data. The means are weighted by population of each country. In case that multiple surveys are available for a country, only the data of the most recent survey is considered. Overall, the data represent the food consumption of more than 3 billion people, which is more than 40% of the world population.[9] According to HBS, the average calorie consumption is 2071, while in FBS the food consumption is 2459. Hence, HBS calories are over 16% lower than suggested by FBS. This high difference might by possibly due to an overestimation calories in FBS. But considering that HBS data refer to purchased food and not to direct intake, also the mean of 2071 kcal might be too low. This confirms the reservation expressed above, regarding the potential of HBS to inform calorie levels of FBS.

[9] Considering the population numbers of 2011.

Table 2: Sample means of HBS and FBS data

Item group	Dietary Energy in kcal		Consumption pattern: Total Shares in %		Consumption pattern: Partial Shares in %	
	NHS	FBS	NHS	FBS	NHS	FBS
Cereals	1216	1325	58.7	53.9	66.8	62.8
Eggs	11	12	0.5	0.5	0.6	0.6
Fish Products	21	19	1	0.8	1.2	0.9
Fruits	46	80	2.2	3.2	2.5	3.8
Meat & Offals	67	98	3.2	4	3.7	4.6
Milk	106	99	5.1	4	5.8	4.7
Pulses	79	85	3.8	3.5	4.3	4
Starchy Roots	96	138	4.6	5.6	5.2	6.5
Sugar Products	134	217	6.5	8.8	7.4	10.3
Vegetables	47	37	2.3	1.5	2.6	1.8
Alcoholic Bev.	11	27	0.5	1.1	.	.
Animal Fats	11	47	0.5	1.9	.	.
Oilcrops & Treenuts	33	51	1.6	2.1	.	.
Spices	15	13	0.7	0.5	.	.
Stimulants	4	3	0.2	0.1	.	.
Veg. Oils	175	207	8.5	8.4	.	.
Grand Total	2071	2459	100	100	100	100

Note: For countries with multiple HBS, only the most recent survey and the corresponding FBS is considered in the calculation. Country-specific data are weighted by the population. Overall, the countries in the sample represent 3,113 million people (UN estimates of 2011).

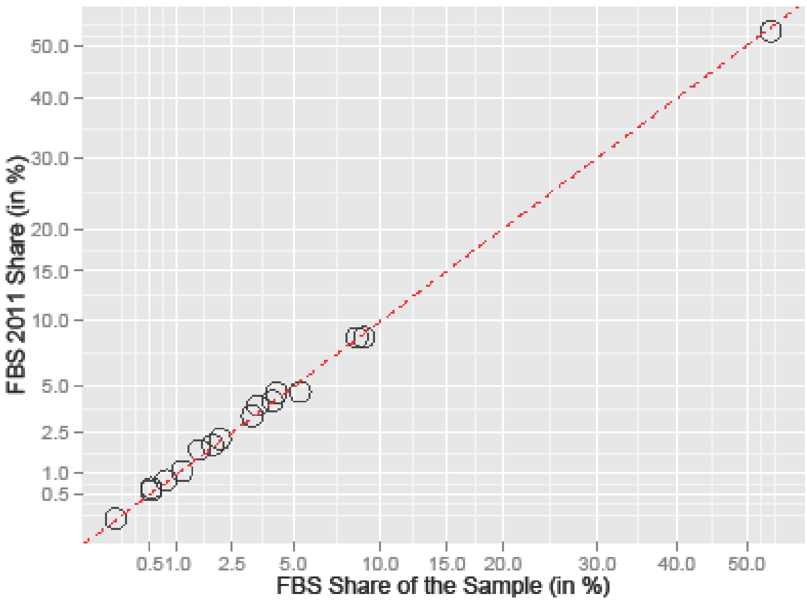
Cereals have by far the biggest item group share, with 58.7% in HBS and 53.9% in FBS (expressed in partial shares, 66.8% and 62.8%). The second biggest amount of calories comes from sugar products according to FBS or, following HBS, from vegetable oils. Thereby, the significantly lower calories from sugar in HBS might be due to the earlier discussed bias in capturing sugar-containing food.

2.4.2 Informative value of the data

As abovementioned, the sample of household surveys represents a large proportion of the world population. However, there is a potential drawback. Some of the surveys have been carried out comparably long time ago. Food consumption patterns of the countries might have changed significantly since the time when the survey was carried out. As a consequence, the data we use might not be representative anymore.

To check for the representativeness of the data, FBS shares used in the analysis are compared with the FBS shares of the year 2011 (the most recent data available at the present). Figure 2 plots the FBS item group shares used in our analysis against the FBS shares of 2011. One can see little variation, which can be regarded as a proof that current food consumption pattern is well reflected by the data we use.

Figure 2: Correlation between mean FBS shares of the sample in use and the FBS of the year 2011



Note: The mean shares are weighted for the country population. For illustrative purposes the axes are rescaled to the square root of the share (in %). The only significant outlier is the share of starchy roots, which is 4.4% in 2011 FBS and 5.3% in the FBS in use.

3 A Model for the reconciliation of HBS and FBS

HBS and FBS shares are reconciled using a Generalized Cross-Entropy (GCE) approach. The foundations of the GCE estimation lie in information theory (Shannon, 1948; Jaynes, 1957), whereas the methodology has been introduced to the area of applied econometrics by Golan et al. (1998).[10] This method allows to estimate models even if they are underdetermined and makes it possible to include prior knowledge about parameters which are going to be estimated.

$$\min_{F,w} \sum_i^K F_i \ln\left(\frac{F_i}{\tilde{F}_i}\right) + \sum_{i,l}^{K\times L} w_{i,l} \ln\left(\frac{w_i}{\tilde{w}_i}\right) \quad (1)$$

$$s.t.\ F_i = H_i + \varepsilon_i \quad \forall i \in K \quad (2)$$

$$\text{with } \varepsilon_i = \sum_l^L w_{i,l}\, \bar{v}_{i,l}$$

$$\sum_i^K F_i = 1 \quad (3)$$

$$\sum_i^L w_{i,l} = 1 \quad \forall i \in K \quad (4)$$

$$\min\{\tilde{F}_i, H_i\} \leq F_i \leq \{\tilde{F}_i, H_i\} \quad \forall i \in K \quad (5)$$

[10] The GCE method is especially attractive to estimate social accounting matrices (SAM) by incorporating and reconciling different data sources (Robinson et al. 2001; Fofana et al., 2002).

The GCE model developed for this reconciliation takes into account that neither FBS, nor HBS can serve as a general benchmark. The model is inspired by Robillard et al. (2003), even though the context of its application is quite different to the latter.

In particular, our model consists of an objective function (1) which is minimized subject to constraints (2)-(5). The reconciliation is performed individually for each of the 64 HBS/FBS comparisons. $\tilde{F}_i$ is the original (partial) FBS share of the item group i. H_i is the (partial) HBS share and F_i the new (partial) FBS share which has to be found by the procedure. It is important to note, that the shares are calculated on the basis of a subset K of item groups, i.e. cereals, starchy roots, sugar products, pulses, vegetables, fruits, meat & offal, fish products and milk.[11]

A special feature of this model is that constraint (2) is stochastic, where H_i is assumed to have a measurement error ε_i. This error has a discrete support set (with L=5 elements) defined by $\bar{v}_i$, where

$$\Pr(\varepsilon_i = \bar{v}_{i,l}) = w_{i,l} \;\; \forall \, l \in L.$$

The distributional form of the error term ε_i is defined by w_i. The error term is therefore a weighted sum of the discrete error support set $\bar{v}_i$ and the error weights w_i (see constraint (2)).

While the elements of $\bar{v}$ remain fixed, the error weights *w* are updated simultaneously with the FBS shares. The model requires the input of prior errors weights $\widetilde{w}$. If the H_i is assumed to be unbiased, the prior error, i.e. $\tilde{\varepsilon}_i = \sum_l^L \widetilde{w}_{i,l} \, \bar{v}_{i,l}$, is set to 0. If $\widetilde{w}$ are chosen such that $\tilde{\varepsilon}_i > 0$, the HBS is implicitly assumed to underestimate the respective item group. Whereas $\tilde{\varepsilon}_i < 0$ implies that the HBS_i is positively biased.

Constraints (3) and (4) are only assuring that updated shares and probabilities sum up to one. Constraint (5) makes sure that F_i lies between the original FBS and the HBS shares. The minimization is performed by a non-linear optimization algorithm (for more details, see Appendix C).

[11] Furthermore, if $HBSkcal_i$ is smaller than 5 kcal, item group *i* is not included in *K*.

The discrepancy metric in the objective function (1) is called Cross-Entropy measure or Kullback-Leibler divergence (Kullback et al., 1951). This objective function penalizes distributions F that are far from the original FBS distribution $\tilde{F}$. Or, to say it differently, the objective function 'attracts' updated FBS distribution towards the original FBS consumption pattern. On the other hand, constraint (2) restricts the updated FBS to lie within an error range around HBS shares. In the case that all prior errors have mean 0, H 'attracts' F as well. For a deviation of share F_i away from H_i, the 5 error weights in w_i have to move away from the priors error weights $\tilde{w}_i$. Since the term at the right-hand side of the objective function penalizes any deviation of w_i from $\tilde{w}_i$, F_i is not unlimitedly allowed to move away from H_i. Hereby, $\tilde{w}_i$ determines the strength of the 'attraction' of H_i. As it will be shown subsequently, prior HBS error weights $\tilde{w}_i$ will be constructed, which have a variance that is inversely related with the reliability the HBS share. Hence, a less (more) reliable HBS share will have a bigger (smaller) prior variance and as a consequence F_i can easier (harder) approach the original FBS share.

Since the prior errors influence the search for updated FBS shares, it is crucial to establish error weights $\tilde{w}$ which correspond to the uncertainty of the item group shares of HBS. It can be seen from Figure 1, the way how HBS and FBS shares are correlated varies from item group to item group. The information from this observed cross-survey deviations between $\tilde{F}_i$ and H_i is used to build the error support set $\bar{v}_i$ and the priors $\tilde{w}_i$. The less correlated the item group shares of $\tilde{F}_i$ and H_i, the higher is the variance of the prior error term.

The components of the error term, $\bar{v}$ and $\tilde{w}$, are constructed for each item group and differ from survey to survey. For that reason it will be temporarily introduced a notation to identify the elements of each of the 64 HBS-FBS comparisons with an index.

The error term of the model can be imagined as a histogram with 5 bins, where $\bar{v}$ are the locations of the 5 bins on the x-axis and $\tilde{w}$ the probabilities associated to the bins. Figure 3 shows an example how the error parameters of an item group are constructed using relative cross-survey deviations between $\tilde{F}_i$ and H_i.

In the first step, for each item group i and survey s, a sample $D_{i,s}$ of relative cross-survey deviations is constructed, with

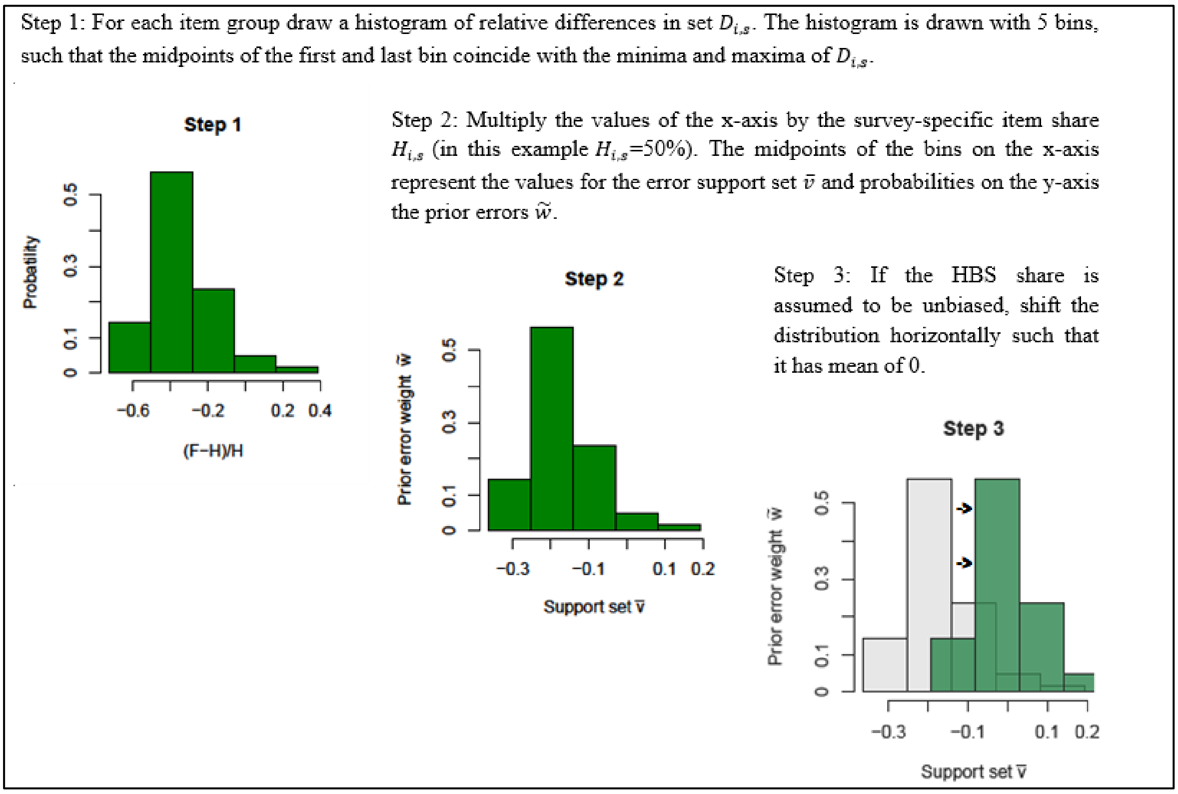
$$D_{i,s} = \{(\tilde{F}_{i,1} - H_{i,1})/H_{i,1}, \dots, (\tilde{F}_{i,S} - H_{i,S})/H_{i,S}\} \quad \forall\, i \in K,\ \ s \in S,$$

where S is the set of HBS-FBS comparisons. A histogram with 5 bins is drawn on the basis of this sample. Each histogram bin is then used to build an element of the discrete error term.

At step 2, the error supports $\bar{v}_{i,l,s}$ are obtained by multiplying $H_{i,s}$ (the HBS share of a specific survey s) with the midpoints of the 5 bins at the x-axis. In that way the distribution is resized to $H_{i,s}$.The probabilities on the y-axis represent the values for $\tilde{w}_{i,l,s}$.

If the HBS share is assumed to be unbiased, a third step is required. In this case, the distribution is shifted horizontally, such that the mean of the distribution equals 0. By this shifting, only the support set $\bar{v}$ is changed, while the weights $\widetilde{w}$ remain unchanged. In practice it will be assumed that HBS shares of all item groups, except sugar and meat, are unbiased.

Figure 3: Steps to obtain the error support set and prior errors weights for a specific item group and survey



The prior errors for the biased items sugar and meat are modelled in a way that its distribution depends on the ratio of food away from home reported in the survey. Cases with a high ratio of food consumed away from home are found to have larger relative differences for meat and sugar. This difference is assumed to exist due to a measurement error in HBS. To correct this underestimation of sugar and meat in HBS, the prior error is given a positive mean, which will correct the negative bias of HBS (see Appendix B for further details).

In order to further augment the error distributions with additional information, they will be refined by an additional modification. For some countries, surveys of different years are available. The countries' real consumption patterns are usually changing slowly over time. If consumption patterns of two HBS of the same country are very similar, it can be regarded as an evidence that the surveys are measuring well food consumption. If instead the consumption patterns are very different for the same country, the survey results might be very noisy. As a consequence, prior error weights will be chosen such that the algorithm will more (less) likely accept values close to HBS shares, if HBS consumption patterns are similar (different) within the surveys of the same country. All details about the construction of the error term can be found in the Appendix B.

The updated FBS share $F_{i,s}$ is a partial share and for the final representation it needs to be rescaled. The rescaling is simply done by multiplying the partial share $F_{i,s}$ with the share of all included item groups over the total FBS energy consumption:

$$UPDATE_{i,s} = F_{i,s} \frac{\sum_j^K FBSkcal_{j,s}}{\sum_j^N FBSkcal_{j,s}} \ , \quad \forall i \in K, s \in S.$$

In this way the sum of all partial shares (which is 100%) is reduced by the size of the excluded FBS shares (which do not change in the procedure). E.g., if the cumulated FBS share of all excluded items groups is 20%, all updated partial FBS shares are rescaled by the factor 0.8.

4 Results

Figure 4 and 5 show the results of the reconciliation procedure. Table 3 summarizes all results. The averages are weighted by country population. In case that multiple surveys are available for a country, only the result of the most recent survey is considered. The reconciliation procedure is performed for all HBS-FBS pairs individually. The results presented here are therefore the weighted mean of 51 independent optimization procedures.[12] FBS shares of item groups, which have been excluded from the reconciliation, are not included in the figures (their shares remain unchanged). The filled blue bars indicate the mean change of the updated FBS with respect to the original FBS. The red frames indicate the difference between the original FBS and the HBS.

Figure 4a shows that the shares of cereals, eggs, fish products and vegetables increase after the reconciliation. Fruits, meat, milk, pulses, starchy roots and sugar products show a negative change. Except of pulses, all means of updated shares are in between the means of original FBS shares and the ones of HBS. Generally, the updated shares are closer to the original FBS than to the HBS. With respect to the original FBS, the updated item group distribution change only moderately. Cereals are the item group that change the most after the reconciliation from 53.9% to 54.9%.

The change of each share is proportional to a change in calories. Changes in shares just have to be multiplied by the average FBS calories (2459 kcal) to quantify the average change in calories. Hence, the energy from cereals increases by about 24 kcal per day (see Table 3), while the calories form starchy roots decreased by about 13 kcal. Other items like eggs, meat and pulses change very little. By construction of the model, the overall calories of the updated FBS equal to the original FBS.

Even tough the average change of item groups is moderate, updated consumption patterns of single countries vary considerably from those of original FBS. Table 3 reports the

[12] The other 13 HBS-FBS pairs have been used solely for the construction if the prior error terms.

median absolute deviations of updated shares from original FBS. In case of cereals the median absolute deviation of shares is 5%, with a maximum absolute deviation of 23%.

As well mean relative changes in consumption of item groups is in some cases considerably large. As shown in Figure 4b, the share of fruits falls on average from 3.2% to 2.9%, which is a relative change of about -12%. For fish products the relative change is +15%, although the mean calorie share of fish changes only from 0.8% to 0.9%

Consecutively, the change in the consumption pattern can be quantified in terms of quantities. This is done by multiplying the relative change of each food item share by the sum of FBS food quantities consumed in all countries under investigation (i.e. of 3,133 million people). It has to be mentioned, that this calculation implicitly assumes that the distribution of specific commodities within the item groups remains fixed.[13] This said, the results expressed in quantities have to be interpreted with caution. Figure 5 shows the changes in quantities (in 1000 Metric tons per year) implied by the reconciled calorie shares. As previously shown, the change of calories is by far the highest for cereals. However, in terms of quantities, cereals are supposed to change by just 10 million tons, which is much less than other item groups. The biggest change occurs for fruits with a decrease of 23 million tons per year. The consumption of vegetables is supposed to increase by 15 million tons, while the consumption of starchy roots decreases by the same amount.

Finally, also the change of fat and protein is evaluated and reported in Table 4. FBS measure only the consumption of these two macronutrients in grams per person. The variation of fat and proteins is obtained analogously to the calculation of quantity changes in Figure 5.[14] Quantities of macronutrients consumed by each item group are simply multiplied by the relative change of item group shares. According to this calculation the total amount of protein consumption remains almost unchanged at about 55.5 grams/person/day. The average consumption of fats increased from 61.4 to 62.7 grams, which is a growth of a little more than 2%. The increase in fats is mainly due to an increase in consumption of fish, cereals and milk.

[13] As reported in Figure 4b, the updated item group share of vegetable was 7.4% higher than the original FBS share. If this increase in calories of vegetables was, e.g, just because of a higher consumption of tomatoes, the quantity increase of vegetables should be higher than the current calculation suggests. This is because tomatoes have a high moisture content compared to other vegetables and provide little calories per kg (Wu Leung, 1968). Therefore, a relatively higher quantity of tomatoes would be needed to compensate for the increase of calories. This has to be kept in mind when interpreting Figure 5.

[14] For the calculation of changes in macronutrients, the same reservation are advisable as for the calculation of changes in food quantities.

Figure 4a: Mean differences between shares of the updated FBS/HBS and original FBS

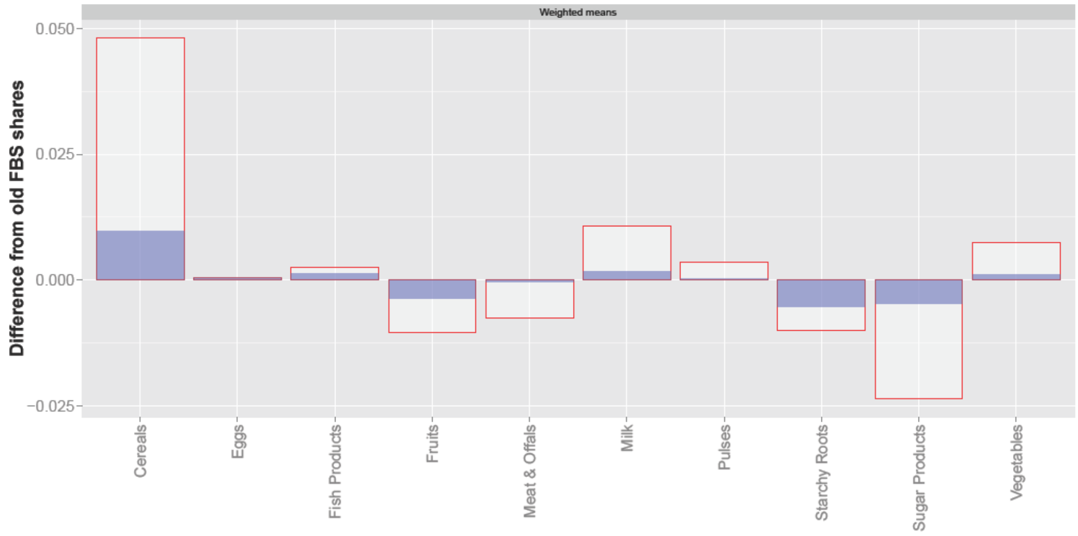
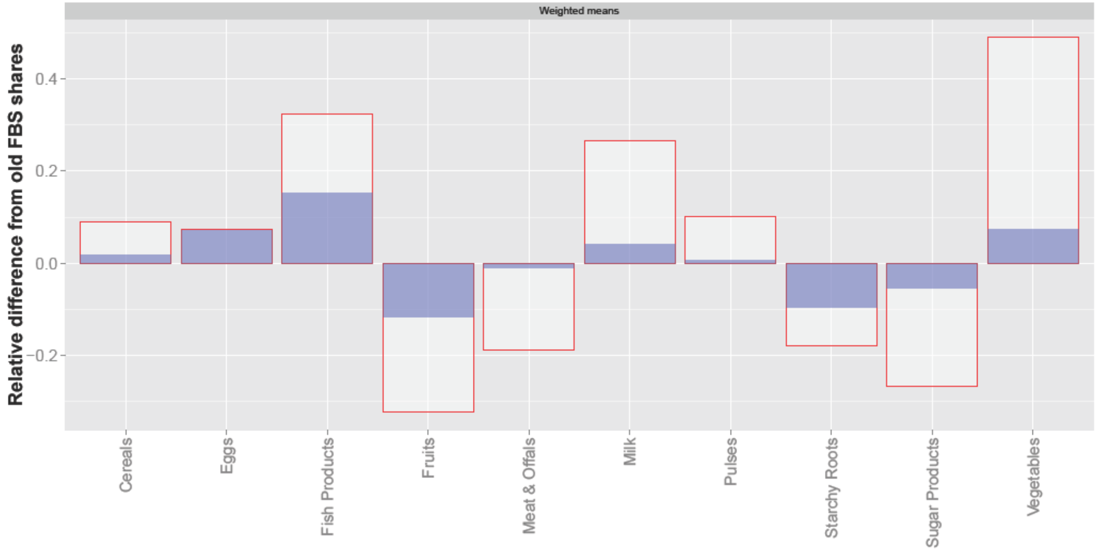
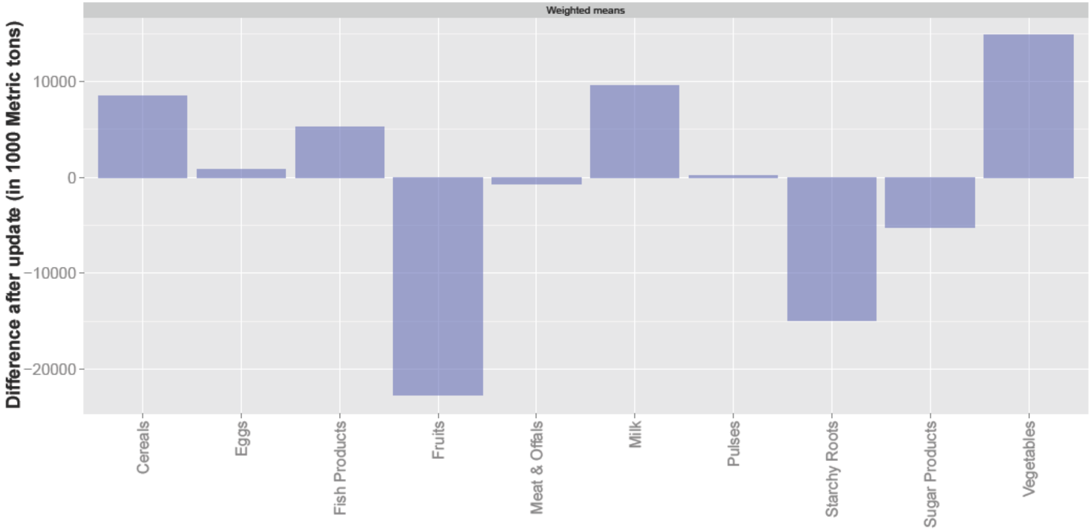


Figure 4b: Relative mean differences between calories of the updated FBS/HBS and original FBS



Note: Mean differences/mean relative differences are weighted for country populations. The red frames denote the mean difference between the HBS and FBS shares. The blue areas denote the mean difference between the updated FBS and the original FBS shares.

Figure 5: Quantity difference between updated FBS and original FBS



Note: Averages are weighted for country populations. The blue areas denote the mean difference between the quantities according to updated FBS and the original FBS.

5 Conclusion

When evaluating the results of this study it should be taken into consideration, that this reconciliation is based on a model and its underlying assumptions. The decision not to include certain item groups in the reconciliation, the bias assumption for sugar and meat and the construction of the pior error distribution in general, are the most important assumptions. Changing the rule of item group exclusion or setting the prior error distribution, might lead to different results. However, all choices are outlined and discussed in the paper. Finally, it is up to reader to interpret the results.

On average, the results of the reconciliations suggest no dramatic correction of the overall food consumption pattern. However, due to the big size of the population represented by the study, even small changes have a large impact in absolute terms. Cumulatively 82.5 million tons of food would be either over or underestimated.

Contrary to small changes in the average consumption pattern, for single countries updated and original FBS shares sometimes differ considerably. When FBS are compiled the updated shares can be used as a benchmark to inform FBS. Original and updated FBS can be compared case by case for all 51 countries. Large discrepancies in the food consumption of an item group may bear suggestions for improvement of the current FBS. Utilization elements of FBS like industrial use, feed, seed, or post-harvest losses might be corrected in this regard.

Table 3: Mean results of the reconciliation procedures

Item group	Updated	Updated FBS share in %	Change of update vs. original FBS					
			Difference in shares (in %)			Average relative diff. in %	Change in kcal	Change in Metric tons
			Mean	Median absolute deviation	Max. absolute devitation			
Cereals	YES	54.9	1	5	23	1.8	23.9	8499
Eggs	YES	0.5	0	0	1	7	0.9	856
Fish Products	YES	0.9	0.1	0.1	4.3	15.1	2.9	5263
Fruits	YES	2.9	-0.4	0.8	4.4	-11.6	-9.3	-22734
Meat & Offals	YES	3.9	0	1	8	-0.9	-0.9	-636
Milk	YES	4.2	0.2	0.3	3.2	4.2	4.1	9506
Pulses	YES	3.5	0	0.4	6.1	0.6	0.5	186
Starchy Roots	YES	5.1	-0.5	0.8	12.1	-9.6	-13.3	-14905
Sugar Products	YES	8.4	-0.5	1	4.4	-5.3	-11.6	-5166
Vegetables	YES	1.6	0.1	0.4	5.5	7.4	2.7	14799
Alcoholic Bev.	NO	1.1	0	0	0	0	0	0
Animal Fats	NO	1.9	0	0	0	0	0	0
Oilcrops & Treenuts	NO	2.1	0	0	0	0	0	0
Spices	NO	0.5	0	0	0	0	0	0
Stimulants	NO	0.1	0	0	0	0	0	0
Veg. Oils	NO	8.4	0	0	0	0	0	0
Grand Total	NO	100	0	0	0	0	0	82550*

Note: Averages are weighted for country populations. For countries with multiple HBS, only the most recent survey and the corresponding FBS is considered.
*The total change in Metric tons is calculated by taking the sum of absolute values.

Table 4: Effects of the reconciliation on consumption of fat and protein

Item group	Updated	Original FBS Macronutrients (in g)		Updated Macronutrients (in g)		Diff. between updated and original FBS (in	
		Fat	Protein	Fat	Protein	Fat	Protein
Cereals	YES	31.5	6.2	32.0	6.2	0.5	0.1
Eggs	YES	1.0	0.9	1.1	1.0	0.1	0.1
Fish Products	YES	3.0	0.7	3.5	0.8	0.5	0.1
Fruits	YES	1.0	0.5	0.8	0.4	-0.1	-0.1
Meat & Offals	YES	7.3	7.5	7.4	7.2	0.1	-0.2
Milk	YES	6.1	4.9	6.4	5.1	0.3	0.2
Pulses	YES	5.4	0.6	5.4	0.6	0.1	0.0
Starchy Roots	YES	1.6	0.2	1.4	0.2	-0.1	0.0
Sugar Products	YES	0.1	0.1	0.1	0.1	0.0	0.0
Vegetables	YES	1.9	0.3	2.0	0.3	0.1	0.0
Alcoholic Bev.	NO	0.2	0.0	0.2	0.0	0.0	0.0
Animal Fats	NO	0.0	5.3	0.0	5.3	0.0	0.0
Oilcrops & Treenuts	NO	1.6	4.1	1.6	4.1	0.0	0.0
Spices	NO	0.5	0.4	0.5	0.4	0.0	0.0
Stimulants	NO	0.3	0.1	0.3	0.1	0.0	0.0
Veg. Oils	NO	0.0	23.7	0.0	23.7	0.0	0.0
Grand Total	NO	61.4	55.4	62.7	55.5	1.3	0.2

Note: Averages are weighted for country populations. For countries with multiple HBS, only the most recent survey and the corresponding FBS is considered.

References

FAO Statistics Division. 1983. A comparative study of food consumption data from food balance sheets and household surveys. *FAO Economic and Social Development Paper* 34. FAO Statistics Division.

FAO. 2001. *Food balance sheets. A handbook.* Rome, FAO(also available at http://www.fao.org/docrep/003/x9892e/x9892e00.htm)

FAO, IFAD & WFP. 2013. *The State of Food Insecurity in the World. The multiple dimensions of food security.* Rome, FAO. (also available at http://www.fao.org/docrep/018/i3434e/i3434e.pdf)

FAOSTAT Website, retrieved on 19.07.2014 http://faostat.fao.org/site/368/default.aspx#ancor

Fofana, I, A. Lemelin, & J. Cockburn. 2002. "Balancing a Social Accounting Matrix." *Centre de Recherche en économie et Finances Appliquées (CREFA)* Université Laval.

Golan, A., G. Judge, & D. Miller. 1998. "*Maximum Entropy Econometrics. Robust Estimation with Limited Data.*" John Wiley & Sons, New York.

Grünberger, K. 2013. "Imputation of Loss Ratios for SUA/FBS Commodities. A Technical Report." *Paper produced as part of the FAO CapEx project.* Mimeo.

Gustavsson, J., Cederberg, Ch., Sonesson, U., van Otterdijk, R. & Meybeck, A. 2011. *Global Food Losses and Food Waste: Extent, Causes and Prevention.* FAO, Rome. (also available at http://www.fao.org/docrep/014/mb060e/mb060e.pdf)

Jacobs, K. & D.A. Sumner. 2002. The Food Balance Sheets of the Food and Agriculture Organization: A Review of Potential Ways to Broaden the Appropriate Uses of the Data. *FAO Review*. (also available at http://182.160.112.236/agridrupal/sites/default/files/FBS_Review_of_Potential_Ways_to_Broaden_the_Uses_of_Data.pdf)

Jaynes, E.G. 1957. Information Theory and Statistical Mechanics. *Physics Review 106*, 620-630.

Johnson, S.G. 2013. *The NLopt nonlinear-optimization package.* http://ab-initio.mit.edu/nlopt

Kraft, D. 1994. Algorithm 733: TOMPFortran modules for optimal control calculations. *ACM Transactions on Mathematical Software 20 (3)* 262-281.

Kullback, S. & R. A. Leibler. 1951. On Information and Sufficiency. *The Annals of Mathematical Statistics 4*, 99-111.

Naska, A., Berg, M.A., Cuadrado, C., Freisling, H., Gedrich, K., Gregorič , M., ... & A. Trichopoulou. 2009. Food balance sheet and household budget survey dietary data and mortality patterns in Europe. *British journal of nutrition*, 102(01), 166-171.

Mahjoubi, R. & A. Prakash. 2012. New framework to compile FAO's Food Balance Sheets. *Paper presented at the Twenty-fourth Session of the Asia and Pacific Commission on Agricultural Statistics.* Da Lat, Viet Nam, October 2012. (Available at http://www.fao.org/fileadmin/templates/ess/ess_test_folder/Workshops_Events/APCAS_24/Paper_after/APCAS-12-17-_Food_Balance_Sheets.pdf)

Moltedo, A., Troubat, N., Lokshin, M. & Z. Sajaia. 2014. *Analyzing Food Security Using Household Survey Data: Streamlined Analysis with ADePT Software.* Washington, DC: World Bank.

R Development Core Team. 2013. *R: A language and environment for statistical computing. R Foundation for Statistical Computing.* http://www.R-project.org

Robillard, A.S. & S. Robinson. 2003. Reconciling Household Surveys and National Accounts Data Using a Cross Entropy Estimation Method. *Review of Income and Wealth* 49 (3), 395-406.

Robinson, S., A. Cattaneo & M. El-Said. 2001. Updating and Estimating a Social Accounting Matrix Using Cross Entropy Methods. *Economic Systems Research 13*, 47-67.

Runarsson, T.P. & X. Yao. 2005. Search biases in constrained evolutionary optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 35 (2)* 233-243.

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27, 379-423.

Sibrián, R., J. Komoroska & J. Mernies. 2006. Estimating household and institutional food wastage and losses: Measuring food deprivation and food excess in the total population. *Statistics Division Working Paper Series no:* ESS/ESSA/001e. Rome, Italy: FAO. (also available at http://www.fao.org/fileadmin/templates/ess/documents/food_security_statistics/working_paper_series/WP001e.pdf)

Smith, L. C., Dupriez, O. & N. Troubat. 2014. Assessment of the Reliability and Relevance of the Food Data Collected in National Household Consumption and Expenditure Surveys. *International Household Survey Network, Food and Agriculture Organization of the United Nations, World Bank.* Mimeo

Wu Leung, W. T., Busson, F., & C. Jardin. 1968. Food composition table for use in Africa. *FAO-Nutrition Information Documents Series, (3).* (also available athttp://www.fao.org/docrep/003/X6877E/X6877E00.htm)

Ypma, J. 2013. *Introduction to NLoptr: An R interface to NLopt*. (Available at http://cran.r-project.org/web/packages/nloptr/vignettes/nloptr.pdf)

Appendix A: List of surveys

Country	Survey Name	Period
Albania	**Living Standards Measurement Survey**	**2005**
Azerbaijan	**Household Budget Survey**	**2006**
Bangladesh	**Household Income and Expenditure Survey**	**2000**
Bangladesh	**Household Income and Expenditure Survey**	**2005**
Bolivia	**Encuesta Continua de Hogares**	**2003-2004**
Brazil	**Pesquisa de Orçamentos Familiares**	**2008-2009**
Bulgaria	**Integrated Household Survey**	**2001**
Burkina Faso	**Enquête Burkinabé sur les Conditiones de Vie des Ménages**	**2003**
Cambodia	**Socio-Economic Survey**	**2003-2004**
Cambodia	**Household Socio-Economic Survey**	**2009**
Cape Verde	**Inquérito às Despesas e Receitas Familiares**	**2001**
Colombia	**Encuesta Nacional de Ingresos y Gastos**	**2006-2007**

Dem. Rep. of Congo	**Enquête Congolaise auprès des Ménages pour l'Evaluation de la Pauvreté**	**2004-2005**
Côte d'Ivoire	**Enquête sur le Niveau de Vie des Ménages**	**2002**
Ecuador	**Encuesta Condiciones de Vida & Quinta Ronda**	**2005-2006**
Egypt	**Integrated Household Survey**	**1999**
Ethiopia	**Household Income Consumption and Expenditure Survey**	**1999-2000**
Gabon	**Enquête Gabonaise pour l'Evaluation et le Suivi de la Pauvreté**	**2005**
Ghana	**Living Standards Survey**	**1998-1999**
Guatemala	**Encuesta Nacional de Condiciones de Vida**	**2006**
Haiti	**Enquête Budget Consommation des Ménages**	**1999-2000**
Hungary	**Hungarian Household Budget Survey**	**2002**
India	**National Sample Survey**	**1993-1994**
India	**National Sample Survey**	**2004-2005**
Indonesia	**National Social Economic Survey**	**2008**
Indonesia	**National Social Economic Survey**	**2011**
Iraq	**Household Socio-Economic Survey**	**2007**
Kenya	**Integrated Household Budget Survey**	**2005-2006**
Laos	**Lao Expenditure and Consumption Survey**	**2007-2008**
Lithuania	**Household Budget Survey**	**2002**
Malawi	**Second Integrated Household Survey**	**2004**
Mali	**Enquête Malienne Sur l'Evaluation de la Pauvreté**	**2000-2001**
Mexico	**Encuesta Nacional de Ingresos y Gastos de los Hogares**	**2004**
Mexico	**Encuesta Nacional de Ingresos y Gastos de los Hogares**	**2006**
Mexico	**Encuesta Nacional de Ingresos y Gastos de los Hogares**	**2008**
Moldova	**Household Budget Survey**	**2003**
Moldova	**Household Budget Survey**	**2006**
Mozambique	**Inquérito aos Agregados Familiares sobre Orçamento Familiar**	**2002-2003**
Nepal	**Living Standards Survey**	**1995-1996**
Nicaragua	**Encuesta Nacional de Hogares sobre Medición de Nivel de Vida**	**2005**
Niger	**Enquête Nationale sur le Budget et la Consommation des Ménages**	**2007**
Pakistan	**Social and Living Standards Measurement Survey**	**2005-2006**
Occ. Terr. of Palestine	**Palestinian Consumption and Expenditures survey**	**2005**
Panama	**Encuesta de Niveles de Vida**	**2008**
Papua New Guinea	**Household Survey**	**1996**
Paraguay	**Encuesta Integrada de Hogares**	**1997-1998**
Peru	**Encuesta Nacional de Hogares**	**2003**
Philippines	**Integrated Farm Household Survey**	**2003**
Sri Lanka	**Integrated Survey**	**1999-2000**
Sudan (former)	**National Baseline Household Survey**	**2009**
Tajikistan	**Household Budget Survey**	**2005**
Tajikistan	**Living Standards Survey**	**2007**
Tanzania	**Household Budget Survey**	**2000-2001**
Tanzania	**Household Budget Survey**	**2007**
Timor-Leste	**Living Standards Measurement Survey**	**2001**
Togo	**Questionnaire des Indicateurs de Base du Bien-être**	**2006**
Uganda	**National Household Survey**	**2002-2003**
Uganda	**National Household Survey**	**2005-2006**
Uganda	**National Panel Survey**	**2010-2011**
Venezuela	**Encuesta de Hogares por Muestreo**	**2004-2005**
Vietnam	**Living Standards Survey**	**1992-1993**
Vietnam	**Living Standards Survey**	**1997-1998**
Vietnam	**Living Standards Survey**	**2006**
Zambia	**Living Conditions Monitoring Survey**	**2002-2003**

Appendix B: Construction of error parameters

In Section 3 the construction of the error term has been already outlined. However, some technical details need to be added in order to complete the description.

B.1 Special Treatment of Countries with Multiple Household Surveys

For some countries, more than one survey is available. Assume that for a specific country food consumption has been measured by a survey S_1 and in another year by a survey S_2. In this case the information in S_2 clearly provides a valuable input when survey S_1 is evaluated.

As outlined in Section 3, the calculation of distribution of errors is based on a sample of relative differences between FBS and HBS shares. When drawing the distribution (the histogram of Step 1 in Figure 3) for the survey S_1, the sample observations of the survey S_2 are given a higher sample weight than the other surveys. In particular, S_2 is given a frequency weight which is equal to *G*/*d* (where *G* is the number of surveys, i.e. 64, and *d* is the absolute year difference between surveys S_1 and S_2) compared to the weights of 1 given to surveys from other countries. Using the latter sample weights, data from the same country are given a superior importance, but the weight decreases with increasing time distance of this information.

B.2 Error terms for biased items as a function of food away from home

First of all, an a priori judgment has to be made on whether the share of the particular item group is potentially biased or not. In this paper it will be assumed that HBS shares of all item groups, except sugar and meat, are unbiased (the reasons are discussed in Section 2). In Section 2 it is also mentioned that the bias of HBS shares of sugar and meat may depend on the quantity of food consumed away from home. For this reason, the error term is designed as a function of the ratio of food away from home. For each country, a ratio of food consumed away from home on the total food is calculated. Subsequently, countries are ordered on the basis of the latter ratio by percentiles. The error distributions are then drawn for each survey as outlined in Figure 3 (without Step 3), but using the sample consisting of the 25% of surveys which have lower ratio and the 25% which have a higher ratio.[15] In that way the error distribution is calculated using a sample of countries which are 'similar' (in terms of food consumption away from home) to the country under investigation.

B.3 Error terms for unbiased items

In the hypothetical case that all HBS shares are assumed to be unbiased, all prior error terms should have mean 0. In this case, due to constraint (2), the algorithm starts with the assumption that the updated FBS share equal H_i, because

[15] For surveys between the 25th and the 75th percentile, the distributions will be therefore drawn with 50% of the sample. For the survey below (above) the 25th (75th) percentile, the sample size is obviously smaller.

$$F_i = H_i + \sum_l^L \widetilde{w}_{i,l}\, \bar{v}_{i,l} \quad = \quad H_i \quad \forall i \in K.$$

Only by updating the error weights w, the new distribution F is enabled to move away from the HBS distribution.

However, meat and sugars are assumed to be biased, and their error terms have means which are different from 0. In order to explain the implication of having biased items, let's first define that the biased items, meat and sugars, are elements of the set P and that the other item groups are in set *T*. Furthermore, assume that the sum of mean prior errors of meat and sugars is

$$\sum_i^P \sum_l^L \tilde{\varepsilon}_{i,l} = \theta, \text{ with } \theta \neq 0.$$

If all prior error of all other item groups in *T* would have mean 0, such that

$$\sum_i^T (\sum_l^L \tilde{\varepsilon}_{i,l}) = 0,$$

then the sum of expected HBS shares would unequal 1:

$$\sum_i^K (H_i + \sum_l^L \tilde{\varepsilon}_{i,l}) = 1 + \theta \neq 100\%.$$

This violates a very basic condition, namely that the sum of all shares have to be equal 100%. To solve this problem, the location of the distribution is shifted along the x-axes (see Step 3 of Figure 3). This will be done by adding to the elements of $\bar{v}_i$ a scalar γ_i, such that

$$\sum_{i,l}^L w_{i,l}\left(\bar{v}_{i,l} + \gamma_i\right) = -\theta \frac{H_i}{\sum_i^T H_i} \quad \forall i \in T.$$

Therefore, the bias θ is cancelled out by shifting the mean prior error of each unbiased item slightly away from 0 towards the opposite direction of θ. The magnitude of the shift γ_i is proportional to item group *i*'s share size.

Appendix C: R-code of the reconciliation procedure

The optimization procedure is performed by the statistical software R (package version 3.0.1) (R Development Team, 2013). The optimization algorithms are accessed by the package lnoptr (see Ypma, 2013), which is an interface to the free/open-source library for nonlinear optimization *NLopt* (see Johnson, 2013). The optimization procedure firstly searches a rough global optimum, using the algorithm *ISRES* (Improved Stochastic Ranking Evolution Strategy) (Runarsson et al., 2005). To refine the estimates the local search algorithm *SQP* (Sequential Quadratic Programming) (Kraft, 1994) is employed.

The following script represents a schematic illustration of the reconciliation procedure for one single HBS-FBS comparison. The script focusses only on the optimization procedure, omitting the part in which the prior error distributions are constructed.

```
# # #   R E C O N C I L A T I O N   P R O C E D U R E   # # #

# S E T   V A L U E S

# Old FBS shares
F <- DATA$fbs_share
# HBS

# HBS shares
H <- DATA$hbs_share

# Number of item groups
K <- length(F)

# Size of the support set
L <- 5

# Define the domain of the error distributions
# It is a K x (K*L) matrix
vb <- vb.empiric

# Define the error weights of the support set
# It is a K x (K*L) matrix
wb <- wb.empiric

# D E F I N E   T H E   M O D E L

# All updated parameters are in vector u:
#   u[1],     ..., u[K]     are the updated FBS shares
#   u[K+1],   ..., u[2*K] are the updated weights of error support set for H[1]
#   u[2*K+1], ..., u[3*K] are the updated weights of error support set for H[2]
#   u[3*K+1], ..., u[4*K] are the updated weights of error support set for H[3]
#   u[4*K+1], ..., u[5*K] are the updated weights of error support set for H[4]
#   u[5*K+1], ..., u[6*K] are the updated weights of error support set for H[5]

# OBJECTIVE FUNCTION

eval_f0 <- function( u , F, H, wb, vb ) {
  return(
    sum(u[1:K]*log(u[1:K]/F))+
    sum(u[(K+1):(K+K*L)]*log(u[(K+1):(K+K*L)]/wb))
  )
}

# CONSTRAINTS

#Create a multiplicator matrix with the form:
# 1 1 1 1 1 0 0 0 0 0 ....
# 0 0 0 0 0 1 1 1 1 1 ....
# . . .
# (it sums the error terms in constraint (2))

sumMatrix <- matrix(c( rep( c(rep(1, L), rep(0,(L*K))),(K-1)),
                       rep(1, L)), K, byrow = T)

# Define the constraint function
# (Equality constraints are built by combining 2 inequality constraints).

    eval_g_ineq0 <- function( u, F, H, wb, vb ) {
        return(c(

#   Constraint (2): New FBS = H + error term
```

```
        u[1:K] - H - sumMatrix %*% matrix(t(u[(K+1):(l+K*L)]*vb[1:(K*L)]))
       ,-u[1:K] + H + sumMatrix %*% matrix(t(u[(K+1):(l+K*L)]*vb[1:(K*L)]))


#   Constraint (3): Updated shares to sum up to 1

       , sum(u[1:K])-1
       ,-sum(u[1:K])+1

#   Constraint (4): Updated weights to sum up to 1

      , sumMatrix %*% matrix(t(u[(K+1):(l+K*L)])) -1
      ,-sumMatrix %*% matrix(t(u[(K+1):(l+K*L)])) +1

  ) )
}

#   Contraint (5): Define lower and upper bounds for the algorithm.

    LB = c(pmin(F[1:l],H[1:K]),rep(0,(K*L)))
    UB = c(pmax(F[1:l],H[1:K]),rep(1,(K*L)))

# DERIVATIVES

#   Gradient of the objective function (1)

eval_grad_f0 <- function( u, F, H, wb, vb ){
  return(
    log(u/c(F,wb))+1
  )
}


# Jacobian of constraints (2)-(4)
# (Jacobian matrices of constraints (2)-(4) have to be
#  collected in one function.)

# Predefine Jacobian matrices of constraint (2) and (3)
stochConstraint<-c()
weightSumConstraint <-c()

for (i in 0:(K-1)) {

# Jacobian of constraint (2)
  stochConstraint= rbind(stochConstraint,
                        c(rep(0,i),1,rep(0,K-1-i) ,
                          rep(0,i*L),-vb[(L*i+1):(L*i+L)],
                          rep(0,(K-1-i)*L )
                        )
  )

# Jacobian of constraint (3)
  weightSumConstraint= rbind(weightSumConstraint,
                            c(rep(0,K) ,
                              rep(0,i*L),rep(1,L),rep(0,(K-1-i)*L )
                            )
  )
}

# Jacobian of constraint (4)
uFBSSumConstraint= c(rep(1,K),rep(0,K*L))

# Collect Jacobian matrices in a function
eval_jac_g0 <- function(u, F, H, wb, vb ){
  return( rbind(stochConstraint,-stochConstraint,
                uFBSSumConstraint,-uFBSSumConstraint,
                weightSumConstraint,-weightSumConstraint
```

```
  )
  )
}


# S O L V E

# First calculate a rough global optimum using ISRES algorithm.

 # Take the mean of the HBS and FBS shares as starting values.

  starting.shares=(H+F)/2

  results.isres <- nloptr( x0=c(starting.shares,wb)
                      ,eval_f=eval_f0
                      ,lb = LB
                      ,ub = UB
                      ,eval_g_ineq = eval_g_ineq0
                      ,opts = list("algorithm"="NLOPT_GN_ISRES",
                                   maxeval=1.0e+10,
                                   "xtol_rel"=1.0e-2,
                                   "ftol_rel"=1.0e-4 )
                      ,F = F
                      ,H = H
                      ,wb = wb
                      ,vb = vb
  )

# Save the global optima.
    new.starting.values=results.isres$solution

# Take the results of the global optimization as starting values
# for a more precise local optimization.
# Local optima are calculated by the SQP algorithm.

   results <- nloptr( x0=c(new.starting.values)
                      ,eval_f=eval_f0
                      ,eval_g_eq=NULL
                      ,eval_g_ineq = eval_g_ineq0
                      ,eval_grad_f=eval_grad_f0
                      ,eval_jac_g_eq = NULL
                      ,eval_jac_g_ineq = eval_jac_g0
                      ,lb = LB
                      ,ub = UB
                      ,opts = list("algorithm"="NLOPT_LD_SLSQP",
                                   maxeval=1.0e+10,
                                   "xtol_rel"=1.0e-7,
                                   "ftol_rel"=1.0e-14            )
                      ,F = F
                      ,H = H
                      ,wb = wb
                      ,vb = vb
  )

#End
```

Contact:

Statistics Division (ESS)
The Food and Agriculture Organization of the United Nations
Viale delle Terme di Caracalla
00153 Rome, Italy

www.fao.org/economic/ess/workingpapers

FAO-statistics@fao.org

I4315E/1/01.15