Statistical Working Paper on Balancing Methodology for Food Balance Sheet (FBS)

Marco Garieri, Natalia Golini, Luca Pozzi
Food and Agriculture Organization
of the United Nations

Abstract

In this paper we describe a sampling strategy to generate balanced FBSs. A FBS is a collection of information from different sources (official and unofficial) prone to measures errors and uncertainties. Our method tries to solve the problem of balancing the FBSs in a flexible way, trusting reliable information and re-allocating the uncertainty over the non-reliable information. Considering agro-economical information in the form of minimal constrains and objective functions, we are able to produce a unique solution (balanced FBS).

Keywords: Food Balance Sheet, FBS, Balancing, Structural Zero.

Disclaimer

This Working Paper should not be reported as representing the official view of the FAO. The views expressed in this Working Paper are those of the author and do not necessarily represent those of the FAO or FAO policy. Working Papers describe research in progress by the authors and are published to elicit comments and to further discussion.

This paper is dynamically generated on February 26, 2015 and is subject to changes and updates.

1. Introduction

The Food Balance Sheet presents the overall description of the food supply and utilization for a specific country in a defined time period. The information provided by the FBSs can be used to analyze the agro-economical condition of a country, in particular to assess food security trends over time for specific regions (1). Due to the importance of this information, FBSs must be precise and reliable.

As described in "Food Balance Sheets: A handbook" (2), very often the quality and coverage of data is not optimal, varying considerably from country to country. FBSs are assembled from a variety of sources, official or unofficial, and inaccuracies and errors may also be introduced at every level. In case of non-reliable information the data are often estimated or adjusted while, in case of missing value, imputed. As a result, the FBSs are unbalanced, i.e., for each commodity in a specified country of the reference period the Total Supply (TS) is not equal to the Total Utilization (TU). Then, the balancing of FBSs is a primary problem within FAO.

In this paper we introduce a very flexible algorithm to solve the problem of balancing FBSs through sampling steps. The method proposed respects the "balance equation" for each rows (as explain in Section 3) and takes in consideration of possible constrains, meaningful from a agro-economical point of view, and choose a unique solution that optimizes an appropriate objective function. The approach considers all prior information on measurement or estimate uncertainties available, in the form of mean and variability around the mean ($\pm$ sd) per cell or column, in case only few information is available.

The paper is structured as follows. Section 2 gives an overview of existing methods, section 3 describes the FBSs and the methods used to estimate its components. Section 4 shows the new method to balance the FBSs. Section 5 presents a case study and the conclusions are in Section 6.

2. Background and Review of Literature

The balancing data problem is well-known in economics literature. Often census-based Input-Output (IO) tables or social-accounting matrixes (SAM) are cases of not balanced tables. Some of the most used approach are the Generalized Maximum Entropy (GME) and the Generalized Cross Entropy (GCE) techniques (3), (4), (5). In (3) they state «SAM estimation is not a statistical model where the issue is specifying a random error generating process, but a problem of estimation in the presence of measurement error». This means that standard assumptions in regression analysis are extremely constraining in estimating SAM, due to the few information on errors and structure, and additionally to the scarse dataset. For this reason, they present a flexible "cross entropy" approach to estimate consistent SAMs starting from data estimated with errors. In this approach the error is a weighed average of known constants. Assumption on the weight, considered as prior, has to be given. However, in GME and GCE techniques, the interpretation of the prior information remains an unsolved problem. For this reason Bayesian approaches have been proposed to overcome to this problem. In a Bayesian framework the prior information held by the researcher has a direct and interpretable formulation (6), (7). However, in this case, a considerable amount of prior information is needed: for example it is common to assume normal prior distribution as prior in balancing data problem (i.e., SAM or IO tables), but still numerically caveats still persist in large scale problems (sparse constraint matrices with real data).
The balancing data problem can be associated with the solution of the contingency tables (two-way or multi-way tables of no-negative values) with fixed margins. In this case, the problem is reformulated in question of inference (Fisher's exact test of independence in two-way or multi-way tables) of model's parameters associated with imputation for the individual cell entries (see (8), (9) (10), (11), (12)) In (8), the authors propose a combination of a sequential importance sampling (SIS) with a sequentially updated normal proposal distribution. And thus, the maximum entropy distribution is sequentially approximated. The algorithm, has to sample, sequentially, all the cell conditionally to the constrains, for this reason is very memory demanding and the speed depends on the table dimension. Moreover perform diagnostics on the accuracy of the results is not trivial, and the algorithm is dependent on the cells order. For these reason, and for the assumption of non-negative cell value, the method is not suitable for our problem of balancing FBSs, and to find an unique solution.

3. FBS structure

As mention before, a FBS represents the TU and the TS of food for a given country in a specified period. Every row is a single (or aggregated) commodity and every column is each element of the Domestic Supply (Production, Imports, Exports and Stock Changes) and the

Domestic Utilization (Food, Food Manufacture, Feed, Seed, Waste, and Other Uses). Each cell is a value expressed in thousands of metric tons (see http://faostat3.fao.org/download/FB/FBS/E).
In Figure 1, the first lines of the FBS for Italy in 2011 are presented.

Italy
2011

Food Balance Sheets

FILTER | EXPORT DATA | FULL SCREEN

Population (Thousand) 60729.0

Single Items	Domestic Supply					Domestic Utilization						Per Capita Supply			
	1000 Metric tons											Total		Prot.	Fat
	Prod.	Impo.	Stock Var.	Exp.	Total	Food	Food Manu	Feed	Seed	Waste	Oth. Uses	Kg / Yr	KCal / Day	Gr / Day	Gr / Day
Cereals - Excluding Beer	19025	12031	-688	5041	25329	9450	815	14082	493	81	410	155.6	1124	34.5	4
Wheat and products	6642	7732	233	3687	10920	8833	74	1636	317	56	3	145.4	1033	32.6	3.7
Rice (Milled Equivalent)	994	134	-2	745	382	315	4	18	31	1	13	5.2	54	1.1	0.1
Barley and products	949	1105	-435	45	1574	24	279	1192	78	2		0.4	2	0.1	0
Maize and products	9753	2834	-419	460	11707	252	458	10560	32	12	394	4.2	32	0.7	0.1
Rye and products	14	13	-3	2	23	4		18	1	0		0.1	1	0	0
Oats	297	61	-63	7	288	14		240	28	7		0.2	1	0	0
Millet and products		9		1	9			9							
Sorghum and products	300	47		4	343			338	1	3					

Figure 1: First lines of the FBS for Italy in 2011.

3.1. Balancing equation

The term "balancing table" refers to the attempt to balance accounts in each country for each commodity. This is done by matching supply (Production, Imports, Stock Changes and Exports) and uses (Food, Food Manufacturing, Feed, Seed, Waste, and Other Uses).
For each commodity i in a given country c at time t the Total Supply (TU) and the Total Domestic Utilization (TU) are described as following:

$$
\begin{aligned}
TS_{c,t,i} &= Production_{c,t,i} + Imports_{c,t,i} + StockVar_{c,t,i} - Exports_{c,t,i} \\
TU_{c,t,i} &= Food_{c,t,i} + FoodManu_{c,t,i} + Feed_{c,t,i} + Seed_{c,t,i} + Waste_{c,t,i} + OtherUses_{c,t,i}
\end{aligned}
$$

Where with $StockVar_{c,t,i}$ we consider the difference between the level of stock at time t and time $t-1$. In order to have a "balanced" or "closed" FBS the following balance identity needs to hold:

$$TS_{c,t,i} = TU_{c,t,i} \qquad \forall i \tag{1}$$

The identity 1 is often not respected, due to the uncertainty of data, as explained in the previous section, Mahjoubi and Prakash (2012) and FAO (2014).

3.2. Methods for estimate TS and TU

Apart from Production and trade (Imports and Exports), which are collected through questionnaires, all other elements of the FBS are usually estimated and official data are available only for very few developed countries (13), (14).

Production and trade

Due to these elements' reliability they drive the FBS production system. Consequently, data collection efforts, especially in the Production domain, that rely on questionnaire response, remain key. The information that FBS supplies ultimately can only be as good as the core underlying data, and concerted action is required to improve data inflows.

More can be said about Production imputation, reference Michael's paper.

Food

This term represents the total supply of all agricultural and derived products available for human consumption. Food available can be reported in terms of primary product equivalent such as wheat and milk or in the form that the products may actually be consumed, such as bread and cheese. This number is typically the residual of the balance sheet. However, due to uncertainty surrounding many other elements of the balance and because of weak methods and outdated parameters, food is not the only residual element. In order to have an accurate estimate for food measurement a methodological framework for incorporating information from other existing statistical sources of information, i.e. Nationally representative Household Survey (NHS) is under development.

Jim work. Reference as soon as available.

Food Processing

This term, also known as Food Manufacturing, is the amount of a commodity processed for food purposes and for which separate entries are provided in the FBSs either in the same commodity tree or in another food commodity. This helps to maintain the concept of accounting for all foods (once and only once) and maintains the links in the various levels of the balance sheets.

Feed

This term represents the quantity of the commodity available for feeding to livestock and poultry during the year. Less than 15% of the countries respond to FAO's feed questionnaire and most of them are developed countries. For the rest of the world, a new method was developed. It is divided in two steps:

i determine feed requirements for metabolizable energy and protein

ii determine allocation of compound/concentrate feedstuffs to match requirements.

See FAO (2014) for more details, and reference Onno work.

Seed

Among the different categories of utilization, Seed use is one of the few categories which can be modeled a priori according to a deterministic rule, animal feed is possibly another. Seeding rates, and ultimately the demand for seed, can be modeled as a function of target plant density, establishment percentage, and seed weight. Multiplied by area planted, seed use can then be derived. However, with the rise in high-precision commercial farming, many farmers are choosing to use certified seed, purchased from specialized seed farmers. This trend requires the need to capture commercial seed production quantities.

Reference Josh work.

Waste

This term is the amount of the commodity lost through wastage during the year at all stages between farms and the household level in handling, storage and transport, but not including waste in the edible and inedible part of the commodity which occurs after the commodity has entered the household. The quantities lost during processing are also not included under this element because they are implicitly considered in applying the underlying extraction rate. Waste data have been reviewed for some 140 developing countries. It can be seen that the coefficients developed to calculate waste in FAOSTAT (waste is calculated as a percentage of total supply) have been constant for many years. Moreover, the review shows inconsistencies in the waste parameters among countries, and commodities within countries.

Reference Klaus work.

Other Utilization

This term is a miscellaneous category to account for all other uses not identified elsewhere (e.g. the use of maize to produce ethanol). This category also includes consumption by those who are not accounted in the country's population (e. g., tourists).

Reference Jim work on tourism.

Stock Changes

Limited information exists on opening or closing stocks for many commodities in many countries, making stocks estimation complicated and imprecise. As a result, stock changes are often calculated to smooth supply and utilization. In practice, they are used as a balancing factor and they may, in part, be calculated residually by first estimating food available for consumption. As the item is partially and residually derived it then reflects not only stock changes but also other statistical errors in the food balance equation.

4. An alternative approach to balance FBSs

4.1. Problem formulation

A generale FBS looks like the following

Table 1: General scheme of the FBS

	L_1	L_2	$\dots$	L_j	$\dots$	L_s	$TotRows$
C_1	x_{11}	x_{12}	$\dots$	x_{1j}	$\dots$	x_{1s}	R_1
C_2	x_{21}	x_{22}	$\dots$	x_{2j}	$\dots$	x_{2s}	R_2
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
C_i	x_{i1}	x_{i2}	$\dots$	x_{ij}	$\dots$	x_{is}	R_i
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
C_r	x_{r1}	x_{r2}	$\dots$	x_{rj}	$\dots$	x_{rs}	R_r
$TotCols$	T_1	T_2	$\dots$	T_j	$\dots$	T_s	Tot

Each row C_i represents a commodity and each column L_j represents a level, both supply or utilization. The table satisfies the following identities:

$$R_i = \sum_{j=1}^{s} x_{ij} \tag{2}$$

$$T_j = \sum_{i=1}^{r} x_{ij} \tag{3}$$

$$Tot = \sum_{i=1}^{r} R_i = \sum_{j=1}^{s} T_j = \sum_{i=1}^{r} \sum_{j=1}^{s} x_{ij} \tag{4}$$

First of all, we assume that for each commodity C_i, its total row R_i is fixed. This means that we rearrange the balancing identity shown in 1 placing to the left of the equation the terms that come from official sources, hereafter called consolidated terms, and to right those coming from unofficial sources. In our example, Production is the only consolidated term. Then, the balancing identity 1 can be rewritten in the following way:

$$\begin{aligned} Production_{c,t,i} = &- Imports_{c,t,i} - StockVar_{c,t,i} + Exports_{c,t,i} \\ &+ Food_{c,t,i} + FoodManu_{c,t,i} + Feed_{c,t,i} + Seed_{c,t,i} + Waste_{c,t,i} + OtherUses_{c,t,i} \end{aligned}$$

where $StockVar_{c,t,i} = Stock_{c,t,i} - Stock_{c,t-1,i}$. Then the FBS assumes the form of the following table:

Table 2: Scheme of the FBS

	Imps	StockV	Exps	Food	FoodM	Feed	Seed	Waste	Other	Production
C_1	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	R_1
C_2	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	R_2
...	...	...	...	...	...	...	...	...	...	...
C_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{i6}	x_{i7}	x_{i8}	x_{i9}	R_i
...	...	...	...	...	...	...	...	...	...	...
C_r	x_{r1}	x_{r2}	x_{r3}	x_{r4}	x_{r5}	x_{r6}	x_{r7}	x_{r8}	x_{r9}	R_r
$TotCols$	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	Tot

In the same way, if we believe that the consolidates terms are more than one, i.e. Production and trade (Imports and Exports), the balancing identity can be rewritten as follows:

$$\begin{aligned} Production_{c,t,i} + Imports_{c,t,i} - Exports_{c,t,i} = &\ StockVar_{c,t,i} + Food_{c,t,i} \\ &+ FoodManu_{c,t,i} + Feed_{c,t,i} + Seed_{c,t,i} \\ &+ Waste_{c,t,i} + OtherUses_{c,t,i} \end{aligned}$$

In this way, we consider data from official sources as highly accurate and reliable data, while those from non-official sources prone to potential measurement errors. The latter are often estimated or adjusted with a certain degree of error that depends on differing concepts, definitions, and methodologies involved in data gathering and generation among countries (15).

For these reasons, the consolidated terms may change from country to country, and from year to year.
Then, we assume that all amounts x_{ij} are measured with an additive error:

$$y_{ij} = x_{ij} + e_{ij} \qquad \forall i, j \tag{5}$$

where y_{ij} represents the commodity value and e_{ij} is the difference between the measured value and its real value.
As discussed in (3), the classical assumptions made in regression analysis $Cov(x_{ij}, e_{ij}) = 0$ with $e_{ij} \sim \mathcal{N}(0, \sigma^2) \quad \forall i, j$ are extremely constraining when little is known about the error of structure and data are scarse. Even if a normal distribution can be attributed to e_{ij}, the assumption of zero mean seems not realistic in our case.
To take into account the measurement error of the individual cells (x_{ij}), we assume for these values a Normal truncated distribution:

$$x_{ij} \sim TN(\mu_{ij}, \sigma_{ij}^2) \tag{6}$$

where μ_{ij} are the estimates for the elements of the TS and TU and σ_{ij} is the standard deviation reflecting the degree of uncertainty of the estimates.

The shape of the distribution will change depending on the prior information available: for accurate values the distribution will be narrow, while for not well measured values we will have a uniform-like prior. Moreover we have values null with certainty (structural zeros), i.e., quantities of processed seed use are not measured in the amount of eggs. Since, each column totals T_j is not a fixed value, the experts can give the possible range of outcomes as the interval $(t_{j,min}, t_{j,max})$. Since the only fixed values are the rows' totals, the algorithm work independently row by row, sampling each cell from its prior distribution. The last value in the row, x_{is}, is calculated as follows:

$$x_{is} = R_i \sum_{j=1}^{s-1} -x_{ij} \tag{7}$$

If x_{is} falls inside its possible range is accepted, otherwise the row is sampled again. In this way the computation time is indeed reduced all the uncertainty is shared between the less reliable estimates. If the all rows are balanced, the algorithm check the columns' totals, and if they fall within each own range. In this case the table is accepted as a possible solution. In all the space of possible tables, an objective function allow to determine the unique solution to the problem, maximizing or minimizing specific column (or columns) or reducing the variability from the input table. Different objective functions, and combination of them, can be implemented in the algorithm, in order to have a table meaningful from the agro-economical prospective. Other types of constrains can be applied, for example small differences in the short run for the same country, but we want to remind the future user that, in case of multiple constrains and objective functions, the space of possible solution will decrease dramatically.

A general remark, to keep in mind for the user, it that the user will have the possibility to generate several number of possible solution (`n` iterations of the algorithm). In this the optimal table, as a result of the objective function, will be a local optimal in the space of all the possible solutions (`n`). For this reason, a greater number of iterations allows more accuracy on the possible optimal balanced table and additional analyses can be performed to understand the suitable number of iterations needed.

5. Case study: Italy 2011

In this section we present the results obtained by applying this new method to the FBS of Italy in 2011.

In this case, the consolidated term is Production (information taken from FAO experts). So the balancing equation is the following:

$$
\begin{aligned}
Production_{Italy,2011,i} = & -Imports_{Italy,2011,i} - StockVar_{Italy,2011,i} + Exports_{Italy,2011t,i} \\
& + Food_{Italy,2011,i} + Feed_{Italy,2011,i} + Seed_{Italy,2011,i} + Losses_{Italy,2011,i} \\
& + IndUses_{Italy,2011,i}
\end{aligned}
$$

where $StockVar_{Italy,2011,i} = Stock_{Italy,2011,i} - Stock_{Italy,2010,i}$
Since we present a case of the previous FBS format, FoodManu is missing and Waste is replaced by Losses and OtherUses by IndUses. These are the description of the old categories:

Industrial use

Agriculture plays an increasingly important role in the industrial economy, providing feedstocks for the production of liquid fuels, chemicals and advanced materials, such as composites for industry, while the emergence of green industries expands opportunities for the rural sector. A generalized model provides reliable estimates of industrial usage. Important drivers of industrial usage likely include income, the level of industrialization, policy mandates, and a country's comparative advantage in cultivating and processing the crop. A data collection strategies that inform other important industrial activities involving food crops, such as the Aglink-Cosimo framework of the OECD/FAO, is used t o test and refine the model.

Losses

In the accounts of FBS, food losses are considered to be the amount of lost commodities, starting from the moment when production is recorded until it reaches the consumer. Food wasted at the household level is not considered as a loss in the FBS, but rather as "waste". The FBS reports food losses for each relevant type of food, each country and for each year. However, the amount of food loss is not asked in the country's production questionnaire and empirical data on food losses are rare. Using "measured" FBS data, together with data from additional surveys, loss ratios is imputed for cases in which no measured data are available (see (16) for more details).

In the next table the FBS for Italy 2011 is presented partially. The value are expressed in kilocalories per capita per day (kcal/cap/day), and the structural zeros are presented as 0*. In this case StockVar was calculated as a difference from the other values, thus this FBS is already balanced, even if the estimates have uncertainty.

Table 3: FBS Italy 2011

Commodity	Imps	Exps	Feed	Seed	Losses	IndUses	Food	Stock	Prod
1. Butter	21.98	3.34	0*	0*	1.02	0	50.88	-6.39	33.90
2. Barley	166.00	14.79	266.24	6.10	3.22	13.42	3.57	8.44e-15	141.35
3. Cereals	13.81	1.96	21.63	0.08	0.37	1.05	1.30	3.67e-01	12.22
...	...	...	...	...	...	...	...	...	...
73. Veg	37.17	83.37	0	0*	4.41	0*	127.82	31.54	146.88
Tot Cols	4351.34	1768.73	2520.21	86.57	167.93	722.42	4317.65	-152.47	5387.90

While in the next table we have the standard deviation for the trade columns (Imports and Exports). This is another input given by the user in the algorithm. In this table, if the standard deviation is 0, it means that the estimates for that particular commodity per import and/or exports are precise, and they will then considered as consolidated terms.

Table 4: FBS Italy 2011, Imports and Exports standard deviation

Commodity	Imports.sd	Exports.sd
1. Butter	0	0
2. Barley	0	0
3. Cereals	0.02	0.01
...	...	...
73. Veg	0.69	0.57

For the other cells, since each column is calculated with the same method, we give can give, to each column, the level of uncertainty as percentage of the corresponding value. For each column we can set the level j of uncertainty as percentage p_j%, and we will that each cells x_{ij} at the j-column, will be sampled from the range:

$$\mu_{ij} - \mu_{ij} * p/100 \leq x_{ij} \leq \mu_{ij} + \mu_{ij} * p/100 \tag{8}$$

So, for example, if we have p_{Feed} = 20%, for the commodity i = Barley, we will sample from 767.24 $\pm$ 767.24*20/100.

The shape of the distribution, within this range, is given by σ. So, in case the estimates are not accurate and we want to sample uniformly within the rage, σ will have a high value, otherwise if it is smaller we will sample from the truncated bell-shape within the range.
For the column StockVar no variability has been provided by the experts and therefore will be calculated as difference between Production and the remaining terms.

In this case we have an additional information, given by the possible range of Feed total, as presented in the following table.

Table 5: FBS Italy 2011, Feed range

Column	Lower bound	Upper bound
Feed	2497.97	3151.12

In this case, we set up, as objective function, the table with the maximum value of Food. As previously mention, other objective function can be applied, and this is just an example.

5.1. Test

We decide to sample 100 plausible (balanced) tables starting from the FBS for Italy 2011. Because Italy is considered saturated country in term of Dietary Energy Supply (i.e. 3539 Kcal/Cap/Day) by experts, the "optimal balance table" is the one has the lowest total of the Food column. Note that the solution is always unique and the results are reproducible, if the seed is fixed.

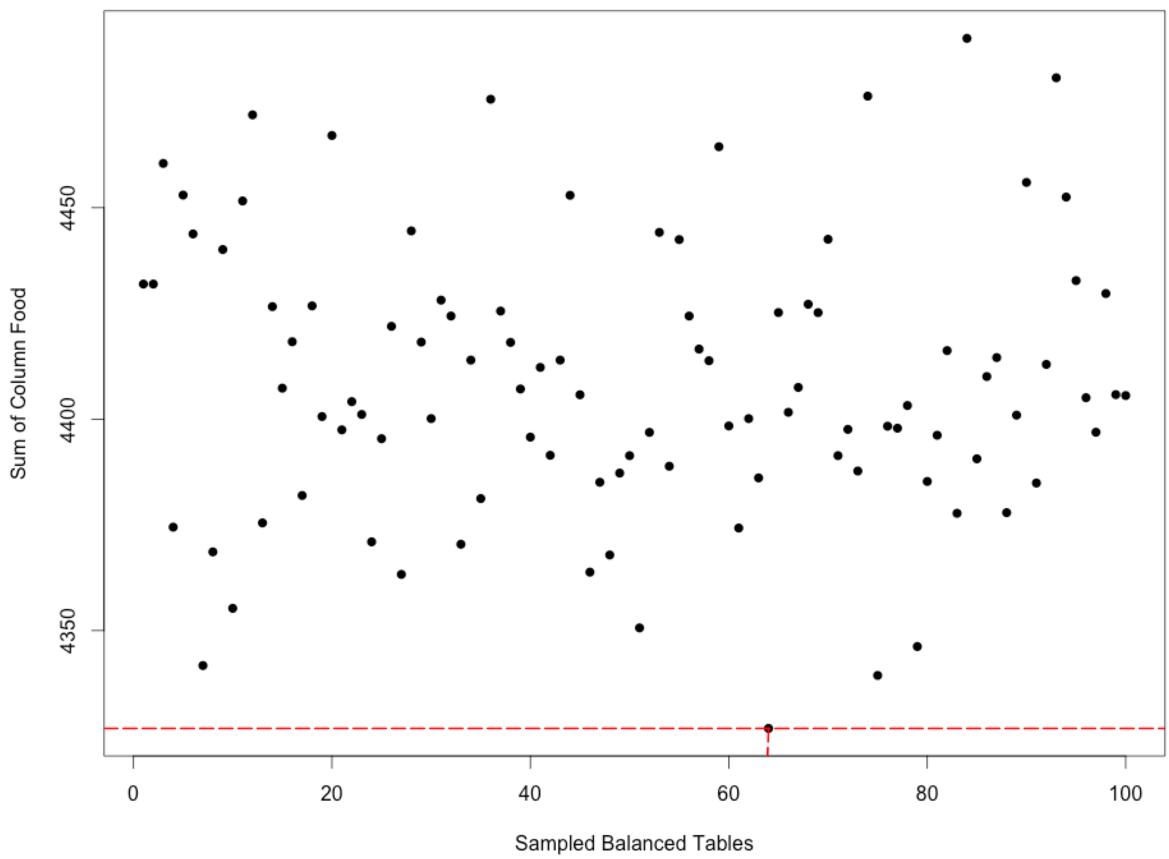

Figure 2: Total of Food for all sampled tables.

In the next table the optimal table, for Italy 2011, is presented

Table 6: FBS Italy 2011, optimal table

Commodity	Imps	Exps	Feed	Seed	Losses	IndUses	Food	Stock	Prod
1. Butter	21.98	3.34	0*	0*	2.97	0	52.42	-2.86	33.89
2. Barley	166.00	14.79	268.60	2.88	4.48	7.90	8.16	0.51	141.35
3. Cereals	13.79	1.96	22.49	2.68	4.17	6.76	9.53	-21.58	12.22
...	...	...	...	...	...	...	...	...	...
73. Veg	37.13	83.68	4.21	0*	1.86	0*	130.43	-36.16	146.88
Tot Cols	4353.29	1767.60	2614.80	130.46	368.21	755.54	4490.04	-385.46	5387.90

In the next Figure 3 we can see the variability of the estimates for "cereals and other products" among all the possible solution (n iteration). The blue dashed line representes the expected value given by the experts, while the green dashed lines indicate the point estimate of the optimal balanced table.

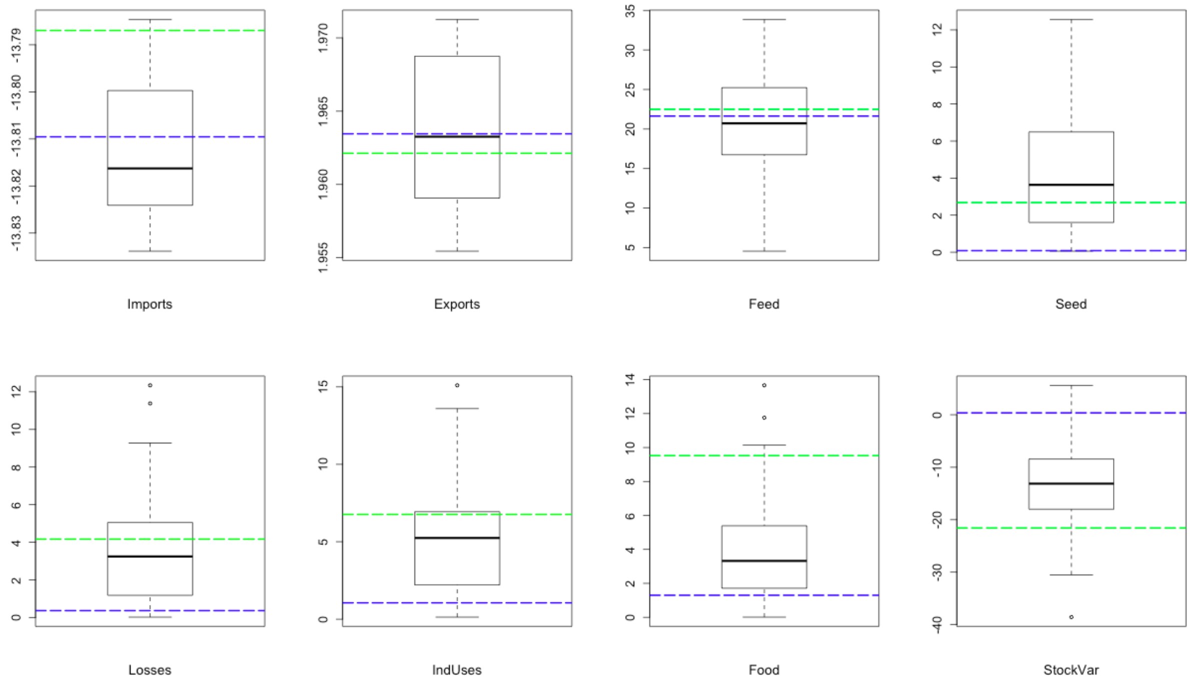

Figure 3: Variability for cereal and other products in the optimal table.

Figure 4 shows the variability of the point estimates of Feed in the sampled balanced tables. The red dashed lines are the upper and lower bounds reported in Table 5.0.2. The blue dashed line is the point estimate of the Feed total given by experts, while the red one is the point estimate reported in the "optimal balanced table". The green dashed line is the value estimated with the new balancing algorithm.

Running different times the algorithm, we noticed that the algorithm always overestimate, but within the ranges, the expected value given by the experts. This suggest further investigations for having a better understanding.

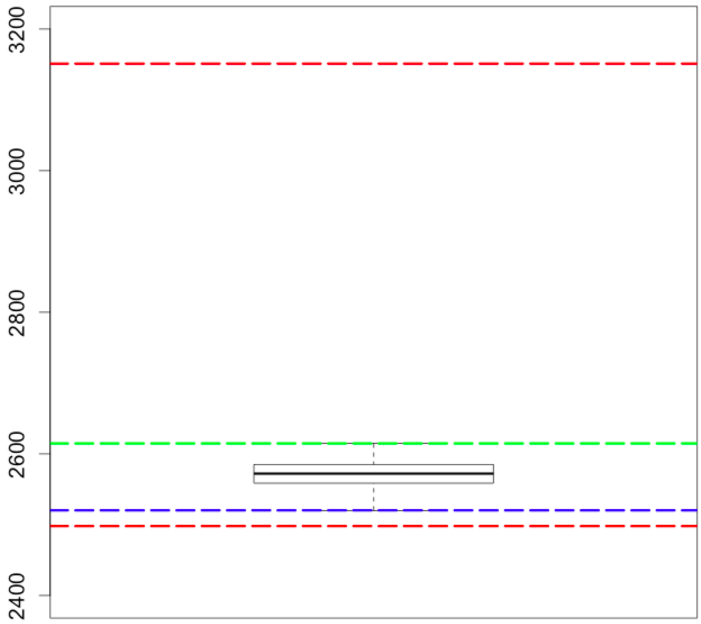

Figure 4: Total of Feed for all sampled tables.

6. Conclusions

The paper shows a flexible algorithm in order to solve the problem of balancing FBS, tables measured with uncertainty. The power of this algorithm is in the ability to find a solution even in the case of incomplete and unreliable information and allowing the user to set several meaningful constrains.
As a note, we want to say that the results will depend on the input given by the users. For this reason it is important that the user provides reasonable information as input, and all the constrains have to be meaningful as well the objective function.

The algorithm is written in the statistical programming language R (17) in the `constable` package. For a user friendly tutorial on `conSTable` package in (18).

Acknowledgement

This work is supervised by Adam Prakash with assistance from Josef Schumidhuber, Onno Hoffmeister, Salar Tayyib whom were crucial in the development of the methodology. The author would also like to thank the team members which participated in the previous discussions providing valuable feedbacks.

A. Supplementary Resources

The data, source code and documentation can all be found and downloaded from https://github.com/mrpozzi/conSTable/tree/develop, the package can also be installed by following the instruction.

B. Pseudo Code

Algorithm 1: Balancing FBS - function `conSTable`

Data: Food Balance Sheet (FBS), Structural Zeros (SZ), Columns Constrains (CC), Objective Function
Result: balanced FBS (bFBS)

```
Initialization;
begin
  forall the cells in i-FBS do
    Apply structural zeros (i-SZ)
  end
end
Sampling;
begin
  forall the commodities C_i do
    x_i. ← TN(μ_i., σ²_i.)
    x_is ← R_i − Σ_j^{s−1} x_ij
    if x_is accetable then
      continue
    end
  end
end
Constrains and Objective Function;
begin
  forall the columns T_j do
    if CC(t_{j,min}) ≤ T_j ≤ CC(t_{j,max})) then
      add sampled FBS as possible solution
    end
  end
  bFBS ← OF(all possible FBSs)
end
```

References

[1] FAO, IFAD and WFP, *The State of Food Insecurity in the World. The multiple dimensions of food security*, http://www.fao.org/docrep/018/i3434e/i3434e.pdf, 2013.

[2] FAO *Food Balance Sheets: A handbook.* http://www.fao.org/docrep/003/x9892e/x9892e00.htm, 2011.

[3] Robinson S., Cattanbo A. and El-Said M, *Updating and Estimating a Social Accounting Matrix using Cross Entropy Methods.*, Economic Systems Research, 13, 47-67, 2000.

[4] Britz W. and Wieck C., *Completeness and Consistency in a Multidimensional Data Base using Constrained Simultaneous Estimation Techniques.*, 2002.

[5] Robilliard A.S. and Robinson S., *Reconciling Household Surveys and National Accounts Data using a Cross Entropy Estimation Method.*, Review of Income and Wealth 49, 395-406, 2003.

[6] Heckelei T., Mittelhammer R. and Jansson T., *A Bayesian alternative to generalized cross entropy solutions for undeterminated econometric models.*, Discussion Paper, Institute for Food and Resource Economics, 2008.

[7] Rodrigues J.F.D., *A Bayesian approach to the balancing of sttistical economic data.*, Entropy, 16, 1243-1271, 2014.

[8] Dinwoodie I.H., Chen Y., *Sampling large tables with constrains.*, Statistica Sinica, 21, 1591-1609, 2011.

[9] Chen Y., *Conditional inference on tables with structural zero.*, Journal of Computational and Graphical Statistics, 16(2), 445-467, 2007.

[10] Chen Y., Diaconis P., Holmes S., and Liu J. S., *Sequential Monte Carlo methods for statistical analysis of tables.*, Journal of the American Statistical Association, 100(469), 109-120, 2005.

[11] Dobra A., Tebaldi C. and West M., *Data augmentation in multi-way contingency tables with fixed marginal totals.* , Journal of Statistical Planning and Inference, 136, 355-372, 2006.

[12] Tebaldi C. and West M. *Reconstruction of contingency with missing data.*, ISDS Discussion Paper, Duke University, n.98-01, 1998.

[13] Mahjoubi R. and Prakash A. , *New framework to compile FAO"s Food Balance Sheets.*, Twenty-fourth Session of the Asia and Pacific Commission on Agricultural Statistics. Da Lat, Viet Nam, 2012.

[14] FAO, Statistics Division (ES) *A new approach to food balance sheet methods and design.*, Twenty-fifth Session of the Asia and Pacific Commission on Agricultural Statistics. Ventiane, Lao PDR, February 2014, 2014.

[15] Jacobs K. and Sumner D.A., *The Food Balance Sheets of the Food and Agriculture Organization: a review of potential ways to broden the appropriate uses of the data.*, FAO Review, 2002.

[16] Grunberger K., Statistic Division (ES) FAO *Estimating food consumption patterns by reconciling food balance sheets and household budget surveys.*, Working Paper Series ESS/14-08, 2014.

[17] R Core Team, *A language and environment for statistical computing.*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL http://www.R-project.org/, 2013.

[18] Garieri M., Golini N., Pozzi L, FAO, Statistics Division (ES) *conSTable Tutorial.*, 2015.

Affiliation:

Marco Garieri, Natalia Golini, Luca Pozzi
Economics and Social Statistics Division (ESS)
Economic and Social Development Department (ES)
Food and Agriculture Organization of the United Nations (FAO)
Viale delle Terme di Caracalla 00153 Rome, Italy
E-mail: marco.garieri@fao.org, natalia.golini@fao.org, luca.pozzi@fao.org
URL: https://github.com/mrpozzi/conSTable/tree/develop