Estimating Supply Utilization Accounts (SUA)

Reconstruction of Tables with Fixed Marginal Totals

Natalia Golini - Marco Garieri - Luca Pozzi

email: natalia.golini@fao.org
marco.garieri@fao.org
luca.pozzi@fao.org

Roma, 9th December 2013

Aim

The balancing of Food Balance Sheets

Questions:

- What is a food balance sheet?
- Why and when a food balance sheet is unbalanced?
- How to solve this problem?

Food Balance Sheet (FBS)

- Presents a comprehensive picture of the pattern of a country's food supply during a specified reference period
- Shows the source of supply and utilization for each food item
- Total supply (TS) and total utilization (TU) are defined for each food item (i) in a given country (c) during the period (t) as follows:

$$TS_{i,c,t} = Production_{i,c,t} + Imports_{i,c,t} + Stock_{i,c,t-1} \quad (1)$$

$$\begin{aligned} TU_{i,c,t} &= Food_{i,c,t} + Seed_{i,c,t} + Feed_{i,c,t} + IndUse_{i,c,t} \quad (2) \\ &+ OtherUse_{i,c,t} + Losses_{i,c,t} + Exports_{i,c,t} \\ &+ Stock_{i,c,t} \end{aligned}$$

Closed Food Balance Sheet

A FBS is balanced when the following equality holds:

$$TS_{i,c,t} = TU_{i,c,t} \quad \forall i \tag{3}$$

Problems:

- A FBS is assembled from a variety of sources (both official and unofficial)
- The quality of a balance sheet and its coverage vary considerably among countries and items (or commodities)
- Inaccuracies and errors are introduced at each stage of a balance sheet's construction

As a consequence we have unbalanced FBS:

$$\Longrightarrow TS_{i,c,t} \neq TU_{i,c,t}$$

How to solve the *balancing* problem?

Starting point: some terms of equations $TS_{i,c,t} = TU_{i,c,t}$ are consolidated, that is released by official sources and therefore considered as accurate values

$$\begin{aligned} Production_{i,c,t} + Imports_{i,c,t} + Stock_{i,c,t-1} &= Food_{i,c,t} + Seed_{i,c,t} + Feed_{i,c,t} \\ &+ IndUse_{i,c,t} + OtherUse_{i,c,t} \\ &+ Exports_{i,c,t} + Losses_{i,c,t} \\ &+ Stock_{i,c,t} \end{aligned}$$

Then we can rearrange the equation as follows:

$$\begin{aligned} Production_{i,c,t} + Imports_{i,c,t} - Exports_{i,c,t} &= Food_{i,c,t} + Seed_{i,c,t} + Feed_{i,c,t} \\ &+ IndUse_{i,c,t} + OtherUse_{i,c,t} \\ &+ Losses_{i,c,t} - StockVar_{i,c,t} \end{aligned}$$

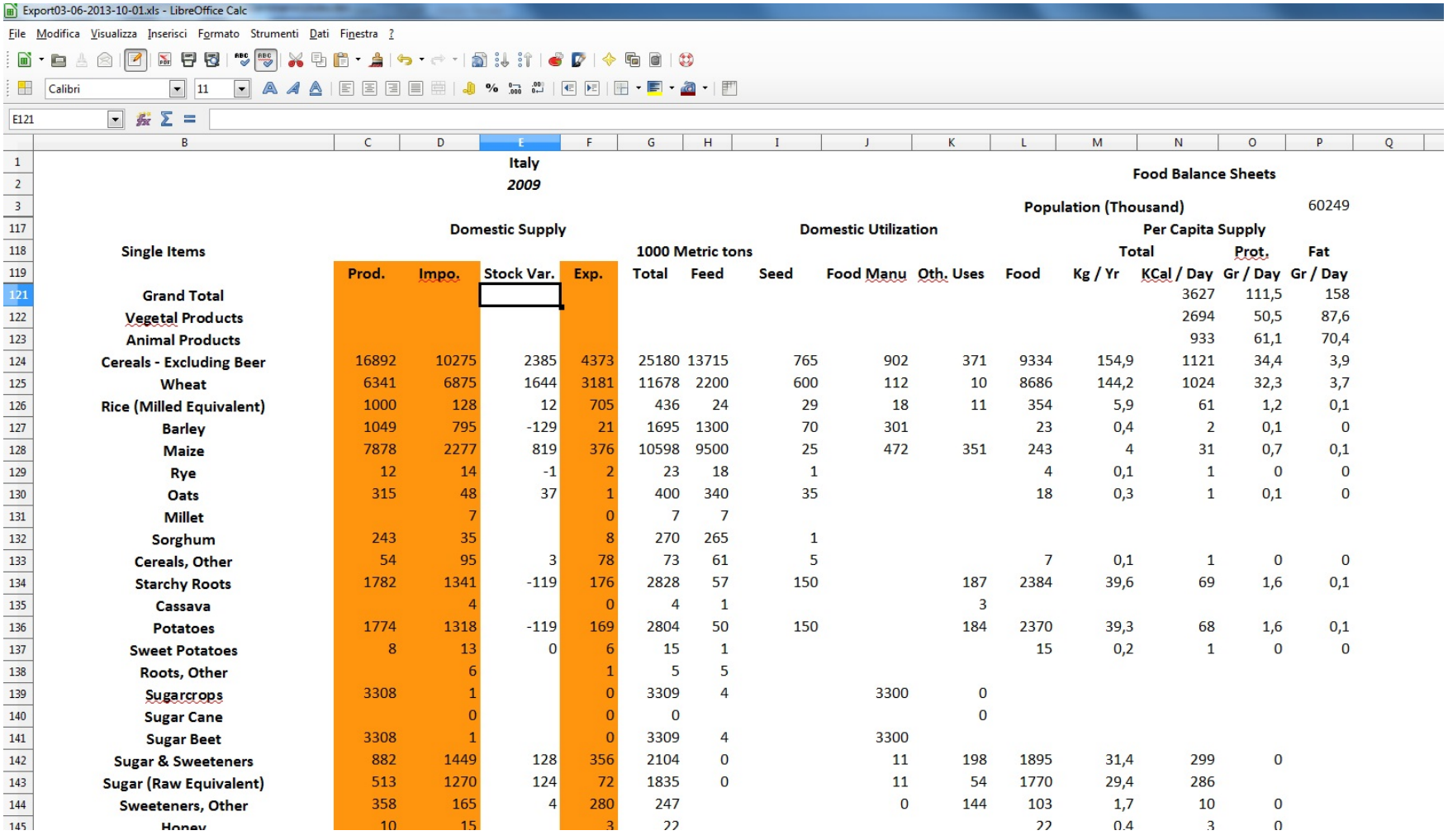
where $StockVar_{i,c,t} = Stock_{i,c,t-1} - Stock_{i,c,t}$

Estimating Supply Utilization Accounts (SUA)

Aim: to estimate the SUAs in such a way that the sum of consolidated balancing terms adds up to the sum of not consolidated balancing items

$$
\begin{aligned}
Production_{i,c,t} + Imports_{i,c,t} - Exports_{i,c,t} \quad &= \quad Food_{i,c,t} + Seed_{i,c,t} + Feed_{i,c,t} \\
&+ \quad IndUse_{i,c,t} + OtherUse_{i,c,t} \\
&+ \quad Losses_{i,c,t} - StockVar_{i,c,t}
\end{aligned}
$$

Example of FBS



Export03-06-2013-10-01.xls - LibreOffice Calc

File Modifica Visualizza Inserisci Formato Strumenti Dati Finestra ?

Calibri 11

E121

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1				Italy									Food Balance Sheets			
2				*2009*												
3												Population (Thousand)			60249	
117			Domestic Supply					Domestic Utilization					Per Capita Supply			
118	Single Items					1000 Metric tons							Total		Prot.	Fat
119		Prod.	Impo.	Stock Var.	Exp.	Total	Feed	Seed	Food Manu	Oth. Uses	Food	Kg / Yr	KCal / Day	Gr / Day	Gr / Day
121	Grand Total												3627	111,5	158
122	Vegetal Products												2694	50,5	87,6
123	Animal Products												933	61,1	70,4
124	Cereals - Excluding Beer	16892	10275	2385	4373	25180	13715	765	902	371	9334	154,9	1121	34,4	3,9
125	Wheat	6341	6875	1644	3181	11678	2200	600	112	10	8686	144,2	1024	32,3	3,7
126	Rice (Milled Equivalent)	1000	128	12	705	436	24	29	18	11	354	5,9	61	1,2	0,1
127	Barley	1049	795	-129	21	1695	1300	70	301		23	0,4	2	0,1	0
128	Maize	7878	2277	819	376	10598	9500	25	472	351	243	4	31	0,7	0,1
129	Rye	12	14	-1	2	23	18	1			4	0,1	1	0	0
130	Oats	315	48	37	1	400	340	35			18	0,3	1	0,1	0
131	Millet		7		0	7	7								
132	Sorghum	243	35		8	270	265	1							
133	Cereals, Other	54	95	3	78	73	61	5			7	0,1	1	0	0
134	Starchy Roots	1782	1341	-119	176	2828	57	150		187	2384	39,6	69	1,6	0,1
135	Cassava		4		0	4	1			3					
136	Potatoes	1774	1318	-119	169	2804	50	150		184	2370	39,3	68	1,6	0,1
137	Sweet Potatoes	8	13	0	6	15	1				15	0,2	1	0	0
138	Roots, Other		6		1	5	5								
139	Sugarcrops	3308	1		0	3309	4		3300	0					
140	Sugar Cane		0		0	0				0					
141	Sugar Beet	3308	1		0	3309	4		3300						
142	Sugar & Sweeteners	882	1449	128	356	2104	0		11	198	1895	31,4	299	0	
143	Sugar (Raw Equivalent)	513	1270	124	72	1835	0		11	54	1770	29,4	286		
144	Sweeteners, Other	358	165	4	280	247			0	144	103	1,7	10	0	
145	Honey	10	15		3	22					22	0,4	3	0	

Food Balance Sheet in tabular form

Item	Food	Feed	Seed	IndUse	OthUse	StVar	Tot.
Cereals	9334	13715	765	902	371	2385	22702
Wheat	8686	2200	600	112	10	1644	9964
Rice	354	24	29	18	11	12	424
...	...	...	...	...	...	...	...
.. ...	...	...	...	...	...	...	...
Oats	18	340	35			37	356
.. ...	...	...	...	...	...	...	...
Potatoes	2370	50	150		184	-119	2873
Sweet Pot.	15	1				0	16
.. ...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...

Tot = Food + Feed + Seed + IndUse + OtherUse - StockVar

Food Balance Sheet in tabular form

Item	Food	Feed	Seed	IndUse	OtherUse	StVar	Tot.	Tot.
Cereals	9334	13715	765	902	371	2385	22702	22794
Wheat	8686	2200	600	112	10	1644	9964	10035
Rice	354	24	29	18	11	12	424	423
...	...	...	...	...	...	...	...	...
.. ...	...	...	...	...	...	...	...	...
Oats	18	340	35			37	356	362
.. ...	...	...	...	...	...	...	...	...
Potat.	2370	50	150		184	-119	2873	2947
Sw. Pot.	15	1				0	16	15
.. ...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
Tot.	110203	30772	1981	34294	6104	5636	177718	183329

Tot = Food + Feed + Seed + IndUse + OtherUse - StockVar

Tot = Production + Import - Export

Note that "StockVar" can assume positive and negative values

Proposed Methodological Approach

Let us assume we fix a particular country and a particular year
The table has for each row a commodity C, each column is a different levels L (Food, Feed, Seed, Losses, StVar, IndUse and OtherUse), for each commodity C
For each commodity the total of the row is R and for each level the total of the column is T

	L_1	L_2	...	L_j	...	L_s	Tot_rows
C_1	x_{11}	x_{12}	...	x_{1j}	...	x_{1s}	R_1
C_2	x_{21}	x_{22}	...	x_{2j}	...	x_{2s}	R_2
...	...	...	...	...	...	...	...
C_i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{is}	R_i
...	...	...	...	...	...	...	...
C_r	x_{r1}	x_{r2}	...	x_{rj}	...	x_{rs}	R_r
Tot_cols	T_1	T_2	...	T_j	...	T_s	

Prior information

- Totals of the rows R_i, given as fixed number
- Distribution of the cells: $x_{ij} \sim \mathcal{TN}(\mu_{ij}, \sigma^2_{ij})$, where μ_{ij} is the estimated mean and a fixed standard deviation σ_{ij}
- The range of the possible outcomes of the columns' totals T_j, given as interval $(t_{j,min},\ t_{j,max})$

Why $\mathcal{TN}$ distribution?

The shape of the distribution will change dependently with the prior information we have for that particular cell. Some estimates of particular levels L are more accurate than others, then if it is more accurate it will be closer to a normal distribution, otherwise it will be like a uniform distribution within the bounds

Sequential sampling steps

- For each row independently, sample each cell from their distribution except the last one (VarStock), which is given by difference from the total of the row R
- For each row independently, check if the value for the last column (VarStock) falls within the given distribution for that cell
- The column's totals T are calculated, and check if all T_js fall inside the given intervals $(t_{j,min},\ t_{j,max})$
- If not successful in the last step the table is rejected and it starts again
- If successful we choose the table with minimize/maximize our objective function

Algorithm

1. For each row R_i (say commodity) of length s:
   1. Set the last cell of the row x_{is} as StockVar, otherwise as the one with the biggest bounds
   2. Sample all cells beside the last one, x_{is}
   3. Compute x_{is} as difference from the totals minus all previous values: $x_{is} = R_i - \sum_{j=1:s-1} x_{ij}$
   4. Check if x_{is} falls inside $x_{is} \sim \mathcal{TN}(\mu_{is}, \sigma^2_{is})$, if not, sample again from the first cell of the row
2. Once all rows R_i are sampled:
   1. Compute the column totals T_j
   2. Check for all if T_j, $t_{j,min} \leq T_j \leq t_{j,max}$ is respected
      - If previous step succeed, the table is accepted as a solution
      - If previous step did not succeed, the algorithm starts from the beginning

Simulation on sample table

This is just an example of a little FBS

Expected Value	Food	Feed	Losses	Seed	IndUse	StVar	Tot	Tot2
Cereals	9210	12940	122	624	833	-344	24073	24150
Starchy Roots	2274	191	129	150	0	-175	2919	2975
Oilcrops	177	310	26	24	5169	277	5429	5451
Vegetable Oils	1527	12	402	0	4	65	1880	1882
Vegetables	12430	930	0	12	0	0	13372	13411
Fruits	9000	0	6	0	6965	90	15881	15874
Meat	5218	0	0	0	16	0	5234	5238
Tot Col	39836	14383	685	810	12987	-87	68788	68981

Scenario I

In this scenario, the bounds given for each cell are really tight. In the following table both percentage and absolute value of gap from the expected values are shown for each cell

±% (absolute)	Food	Feed	Losses	Seed	IndUse	StVar
Cereals	2 (184)	5 (647)	10 (12)	2 (12)	2 (17)	10 (-34)
Starchy Roots	2 (45)	5 (10)	10 (13)	2 (3)	0	10 (-18)
Oilcrops	2 (4)	5 (16)	10 (3)	10 (2)	2 (103)	10 (28)
Vegetable Oils	2 (31)	5 (1)	10 (40)	0	10 (0)	10 (7)
Vegetables	2 (249)	2 (19)	0	10 (1)	0	0
Fruits	2 (180)	0	10 (1)	0	2 (139)	10 (9)
Meat	2 (104)	0	0	0	10 (2)	0
Tot Col	20 (7967)	20 (2877)	20 (137)	20 (162)	20 (2597)	20 (-17)

Scenario II

The bounds, in this case, has a almost double size than in Scenario I

±% (absolute)	Food	Feed	Losses	Seed	IndUse	StVar
Cereals	5 (461)	10 (1294)	20 (24)	5 (31)	2 (17)	20 (-69)
Starchy Roots	5 (114)	10 (19)	20 (26)	5 (8)	0	20 (-35)
Oilcrops	5 (9)	10 (31)	20 (5)	20 (5)	2 (103)	20 (55)
Vegetable Oils	5 (76)	10 (1)	20 (80)	0	20 (1)	20 (13)
Vegetables	5 (622)	5 (47)	0	20 (2)	0	0
Fruits	5 (450)	0	20 (1)	0	2 (139)	20 (18)
Meat	5 (261)	0	0	0	20 (3)	0
Tot Col	20 (7967)	20 (2877)	20 (137)	20 (162)	20 (2597)	20 (-17)

Scenario III

In this scenario, the prior bounds have an huge size comparing to the Scenario II

±% (absolute)	Food	Feed	Losses	Seed	IndUse	StVar
Cereals	10 (921)	10 (1294)	30 (37)	5 (31)	5 (42)	30 (-103)
Starchy Roots	10 (227)	10 (19)	30 (39)	5 (8)	0	30 (-58)
Oilcrops	10 (18)	10 (31)	30 (8)	30 (7)	5 (258)	30 (83)
Vegetable Oils	10 (153)	10 (1)	30 (121)	0	30 (1)	30 (20)
Vegetables	10 (1243)	5 (47)	0	30 (4)	0	0
Fruits	10 (900)	0	30 (2)	0	5 (348)	30 (27)
Meat	10 (522)	0	0	0	30 (5)	0
Tot Col	20 (7967)	20 (2877)	20 (137)	20 (162)	20 (2597)	20 (-17)

Results

In the following table the execution times for each Scenario for 100 iterations are shown

Scenario	user	system	elapsed
I	27.00	1.45	55.91
II	11.04	0.39	11.07
III	11.82	0.03	11.87

In all the three Scenarios and for different number of iterations (100, 1000, 10000) no table has a frequency more than one, thus all the sampled table are different from each other
As a summary for the results of the different iterations for each Scenario, the Root-Mean-Square-Error (RMSE) and the Relative-RMSE (RRMSE) have been calculated. It is important to remark that this step makes sense just in a simulation study and not when the real tables will be sampled

Other examples?

Case a single cell has a true value, how to resample

If after the sampling particular cell values are given as fixed, an additional function of the method is given in order to resample the table given that value as true

- The best table obtain from the sampling procedure is taken
- The new cell bounds are given (a point distribution is given for that cell)
- Resample the row where the cell is present
- Calculate the new column totals
- Take the best new table

Critical Points

- Objective function (maximum of the Food for developed countries, minimize of the VarStock for not-developed countries), probably another one is better
- Need of the prior information for each cell from your side
- ...

Homeworks

For each country and for each year **you** need to provide:

1. Prior distribution of **each cell** (mean and percentage of difference from the mean)
2. Intervals for **column totals**
3. **Row totals** exact values

Question: if we produce a closed FBS at time t, can we use this information for the balancing of the FBS at time $t + 1$?

Thanks!