Statistical Standardization of the Supply and Utilization Accounts

Michael C. J. Kao

Economic and Social Statistics Division (ESSD)

Food and Agriculture Organization
of the United Nations

June 18, 2014

Outline

1. Introduction
2. Implementation
3. Framework for National Accounts
   - Traditional Accounting Standardization
   - Probabilistic Convolution Standardization
4. Implications
5. Further Work

Outline for section 1

1. Introduction
2. Implementation
3. Framework for National Accounts
   - Traditional Accounting Standardization
   - Probabilistic Convolution Standardization
4. Implications
5. Further Work

Supply and Utilization Account

The Supply and the Utilization Account (SUA) is a detailed national account of the supply and utilization of agricultural products.

The Food Balance Sheet (FBS) is an aggregated summary of the Supply and the Utilization Account.

Using accounting analogy, SUA is the collection of all trasanction took place while FBS is a summary derived from SUA acting like financial reports.

What is standardization

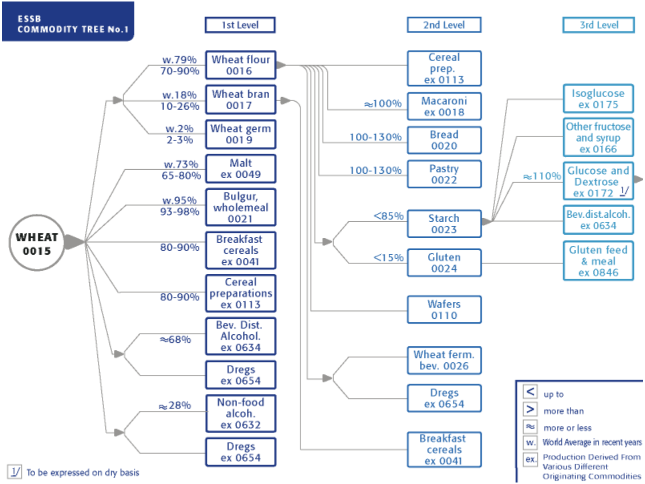
Standardization in this context is the procedure of converting derived commodity such as wheat flour and bread to a comparable standard expression.

For example, instead of how much wheat flour, bread are consumed individually we express it as its wheat equivalent.

The standardized value enables comparison and also reduce the amount of information for comprehension.

Mapping

In order to perform standardization, one must specify the relationship between the derived commodities and their primary commodity. One such instance is the commodity tree shown in the following slide.


ESSB
COMMODITY TREE No.1
1st Level
2nd Level
3rd Level
WHEAT 0015
w.79% 70-90%
Wheat flour 0016
w.18% 10-26%
Wheat bran 0017
w.2% 2-3%
Wheat germ 0019
w.73% 65-80%
Malt ex 0049
w.95% 93-98%
Bulgur, wholemeal 0021
80-90%
Breakfast cereals ex 0041
80-90%
Cereal preparations ex 0113
≈68%
Bev. Dist. Alcohol. ex 0634
Dregs ex 0654
≈28%
Non-food alcoh. ex 0632
Dregs ex 0654
Cereal prep. ex 0113
≈100%
Macaroni ex 0018
100-130%
Bread 0020
100-130%
Pastry 0022
<85%
Starch 0023
<15%
Gluten 0024
Wafers 0110
Wheat ferm. bev. 0026
Dregs ex 0654
Breakfast cereals ex 0041
Isoglucose ex 0175
Other fructose and syrup ex 0166
≈110%
Glucose and Dextrose ex 0172 1/
Bev.dist.alcoh. ex 0634
Gluten feed & meal ex 0846
1/ To be expressed on dry basis
< up to
> more than
≈ more or less
w. World Average in recent years
ex. Production Derived From Various Different Originating Commodities

Outline for section 2

1. Introduction
2. Implementation
3. Framework for National Accounts
   - Traditional Accounting Standardization
   - Probabilistic Convolution Standardization
4. Implications
5. Further Work

Background

The current specification of the commodity tree is technically known as a **rooted tree**.

A rooted tree is a tree which has a designated vertex known as the **root**. In our case, it is the primary commodity group or the target group we would like to standardize to.

The tree is intuitive but restrictive, in this seminar we would like to propose the more general flexible framework of **graph** for standardization.

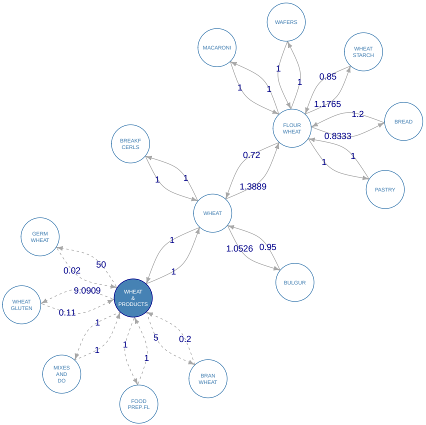
Terminology

Vertex:	A node representing the commodity
Edge:	A line connecting the vertex.
Root:	A specific designated node.
Parent:	The vertex directly above the current vertex.
Child:	The vertex directly below the current vertex.
Leaf:	A node which has has no child.

In a rooted tree, a vertex can have and at most one parent, we would like to relax this and allow multiple inputs for each derivative.

Furthermore, we don't want to fix the selection of the root. We want to be flexible and be able to standardize to any equivalence. To achieve this, we need the vertex to be able to reach every single other vertex.

This is more than a figure, its a mathematical model



Difference in Quantity and Calorie Standardization

Quantity Standardization is performed for all element such as import/export, waste, seed and other utilization; on the other hand, calorie standardization is based solely on food to reflect the consumables.

Certain commodity are standardized in calorie but not in quantity. For example, only wheat flour are standardized to obtain the quantity while wheat germ and bran are not. This is to avoid double counting.

Benefits of a network specification

Traceability:	We can see how it is standardized.
Flexibility:	It can take every single possible form.
Simplicity:	The relationship can be modified easily.

Outline for section 3

Traditional Accounting Standardization

Tradtional accounting standardization system assume that everything is measured without error, and the standardization procedure is merely a matrix arithematic exercise. Furthermore, the frameworkd does not capture the error in model based estimation of elements such as feed and waste.

This is both unrealistic in practice and problematic as we over-state our confidence about the derived statistic.

Probabilistic Convolution Standardization of the SUA

To account for the over-simplified scenario of the accounting system, we propose to use a statistical framework for standardization.

Quantities can be specified as distribution to reflect uncertainty or lack of state of knowledge.

Statistical Standardization of the SUA

Take the commodity tree for example, the extract rate of wheat flour from wheat is between 70 $\sim$ 90% with a global average of 79%. This can be represented as a normal distribution with mean of 0.79 and a standard deviation of 0.035.

Under the assumption of symmetric distributions, the standardization will result in a mean estimate equivalent to the accounting system but a distribution reflecting the state of inadequate information.

Analytical Solution

The next page shows a distributional standardization based on the normal distribution. We have add uncertainty to the quantity of **Wheat Flour** and **Bread**, the standardization shows that the final group **wheat and Products** now has a normal distritbution as well.

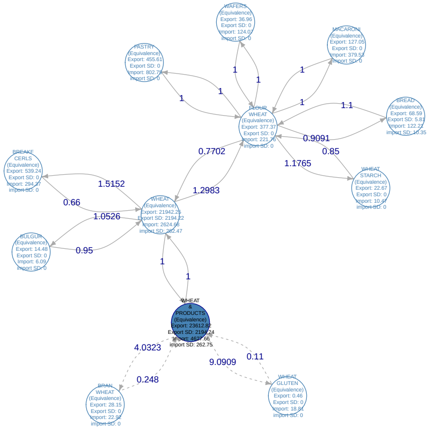
Let:

$$\text{Bread} \sim \mathcal{N}(\mu_{\text{bread}}, \sigma_{\text{bread}})$$
$$\text{Wheat flour} \sim \mathcal{N}(\mu_{\text{wheat flour}}, \sigma_{\text{wheat flour}})$$

then:

$$\text{Wheat and Products} \sim \mathcal{N}\left(\sum_{i \in \mathcal{C}} d_i \mu_i, \sqrt{\sum_{i \in \mathcal{C}} d_i \sigma_i^2}\right)$$

Analytical Solution



MCMC Solution

We don't have to restrict ourself to distribution with analytical mixture, as a matter of fact we can speficy any distribution and obtain the mixture via MCMC.

We can model the dependency between extraction rate and calorie at the same time, and many other joint probability.

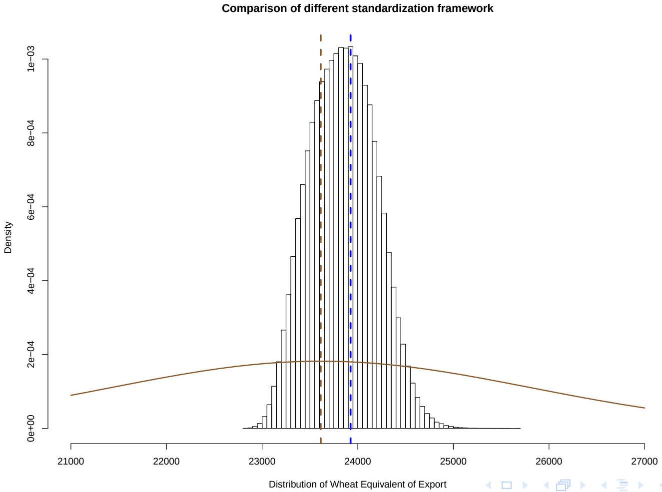
MCMC Solution

Let us assume the following complicated situation:

$$
\begin{aligned}
\text{wheat} &\sim \mathcal{U}\left(E_{a,b}, \sum_{i \in \mathcal{T}} I_{b,a}\right) \\
\text{bread} &\sim \mathcal{U}\left(E_{a,b}, \sum_{i \in \mathcal{T}} I_{b,a}\right) \\
\text{Wheat flour \& Bulgur} &\sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
\text{Pastry} &\sim \Gamma(\alpha, \beta)
\end{aligned}
$$

Then the final standardization will be a mixture distribution and does not have to be restricted to any standard probability distribution.

Mixture distribution of wheat equivalent



Outline for section 4

Removal of unecessary accounts

We no longer need a category called statistical discrepancy.

Integration with other FBS methodology

Errors in model based estimate can be preserved and passed to future analysis and methodology

The new standardization also provides a lower level prior distribution for the multiway-contingency sampling in the new FBS balancing methodology. It give us more insight about the sources of the variability and undertainty rather than guess at the aggregated level.

Bayesian estimate of the Prevalence of Undernourishment

The calorie computed from the standardization and ultimately the dietary energy supply (DES) is used for the calculation of the Prevalence of Undernourishment (PoU).

Bayesian methodology can be developed for PoU to reflect our uncertainty about food availability and PoU.

Outline for section 5

How to set the distributions?

Given the large number of commodity, it may be difficult to set the distribution for all the elements accross all country for the whole system.

Integration with other methodology

Despite the flexibility, the integration of the probabilistic standardization with other methodology will require careful and comprehensive thought for implementation.

- MCMC results will not be replicable, at least not to the same degree of accuracy.
- Passing parameters from different methodology will require special classes to handle ditribution or sample.