Imputation of Loss Ratios for SUA/FBS Commodities A Technical Report*

Last update: December 31, 2013

Food losses have been on the agenda of policy makers for decades. However, few efforts have been made to measure the amount of food lost along the production chain. Losses occur for almost all types of food crops and they occasionally surpass 50% of total supply (Gustavsson et al., 2011). The relevance of food losses for the worldwide food security stands in sharp contrast to its poor data availability.[1] In particular, loss figures used to compile the Food Balance Sheets (FBS) reveal many inconsistencies, and the empirical foundation of these are often questionable. The review of Shapouri et al.(2013) highlights many of these inconsistencies in current assessments of losses in the FBS. Hence, there is a pressing need to update the current methodology for measuring food losses.

This report will present a method for estimating food losses by making an efficient use of the available information. The following approach assesses loss figures for all primary commodities (except live animals) and all countries listed in the SUA Working System[2]. The estimated losses are based on a statistical model which incorporates information about commodity- and country-characteristics.

The biggest portion of the dataset is taken from a sub-sample of the SUA Working System that is regarded as reliable. Additional data have been collected in order to increase the regional coverage of the dataset. The main objective of the task is to produce loss rates based on empirical evidence, and for which the calculation method is transparent and well-documented.

The proposed method should not be considered a final product, but rather as a procedure open to improvements. In the case that improved data on food losses is provided in the future, the method should be adjusted in order to exploit the potential of these new data.

Data

Within the framework of SUA/FBS, food losses (often also defined as "Waste") are an element of domestic utilization. Relevant for SUA/FBS are only losses from the time at which production is recorded until food reaches the final consumer (FAO, 1972). Losses at harvest or household waste are therefore excluded by definition. Under this definition, the quantity of losses is considerably

*Klaus Grünberger (ESS). E-mail address: klaus.grunberger@fao.org. Any views or opinions expressed are solely those of the author and do not necessarily represent those of ESS.

[1] Parfitt et al. (2012) summarize the few efforts which have been made in the last decades in measuring food losses.

[2] SUA(Supply Utilization Account) is a database on which basis the FBS is constructed. It contains more food items than the FBS, as in the FBS some SUA items are summarized into more general food groups.

lower than losses accumulated from the field until the food arrives on the plate.
Since losses can vary considerably between commodities and countries, a large amount of information is needed to quantify commodity- and country-specific losses. In the SUA Working System, there are more than 100 primary food items and almost 200 countries and territories. This totals approximately 20,000 different loss figures that must be assigned each year. But this high demand of information stands in contrast to the scarcity of available data. Unlike production, food losses are not explicitly requested in country questionnaires, and are often only reported in National FBSs. As a consequence, official data on loss quantities are often missing. Thus far, ESS has used various methods to assess these unknown loss quantities.
The loss data in the SUA Working System are categorized into different types of sources, where the origin is indicated by flags. Table 1 and 2 show the frequencies and mean loss rates by data source, food group and geographical region. There are 5 major types of data sources:

1. Data with an empty flag: These are "official" loss data from national FBS or national statistic institutes. Even if there might not always be a genuine research effort behind these loss data, one can expect them to provide substantial country- and commodity-specific information. Unfortunately the share of "official" loss observations in the period from 1971 to 2009 is only 4.4%.

2. Flag "C": Losses calculated with fixed loss ratios from the TCF Table (see FAO, 1972). The latter ratios are in large part unchanged since the 1960s and they require updates. With a share of 86% they represent the vast majority of the data.

3. Flag "F": Loss data provided by ESS staff and rely on their personal assessments. Although the quality of personal assessments is not called into question, this method of data generation may be criticized due to its lack of transparency.

4. Flag "B" (where the "B" stands for "balanced"): In the case that all other elements of the FBS are given, losses simply represent the residual of all other elements. The mean loss rate of 18.4% is substantially higher than the mean of all other types of sources, which raises some strong doubts about its reliability.

5. Flag "*": These are data which should be provided by NGOs. Unfortunately sources were rarely given. Even when there was a note, it did not necessarily refer to a research institution, but rather to a trend which had been followed.

We consider loss data with the empty flag as "reliable" and we use only the latter for this analysis. Data from the other sources should be updated for the reasons stated previously (maturity of the data and intransparancy).[3]

Data Selection

For the estimations we use only data with an empty flag, i.e. "official" information. However, we recognize that the quality of these data is questionable. The definition of losses in some National FBS might not be the same as the definition in the SUA/FBS.[4] Furthermore their methodology

[3] In principle we could use data with flag "*", but as mentioned above some doubt arose about the sources of these data. Since the quantity of these data was also low, we decided not to consider them for our analysis.

[4] See Appendix B for the definition of losses.

Table 1: Loss data from the SUA Working System

Flag	Freq.	Percent	Mean
	11,788	4.4	7.2
C	232,025	85.8	7.4
F	9,742	3.6	6.5
B	12,824	4.8	18.4
*	1,511	0.6	5.1
Other	2,471	0.9	.
Total	270,361	100.00	7.8

Source: SUA Working System 1970-2009 for the following primary commodity groups: cereals, roots and tubers, sugar crops, pulses, tree nuts, oil crops, vegetables, fruits, stimulants and spices, milk, eggs, offals, slaughter fats and meat.

Table 2: Mean of all loss rates (in %) by food group and geographical region.

	SSA	Northern Africa & Middle East	Latin America	Western Industrialized Countries	Eastern Eur. & Central Asia	Asia & Pacific	Total
Cereals	7	5	4	4	12	5	5.9
Roots & Tubers	11	7	10	7	12	8	9.3
Sugar crops	6	7	3	1	1	6	4.4
Pulses	7	4	4	4	5	4	4.8
Tree nuts	5	4	13	6	4	4	5.4
Oil crops	6	6	6	3	4	3	4.6
Vegetables	10	10	11	9	13	8	9.8
Fruits	9	9	11	8	11	8	9.1
Stimulants & Spices	5	5	4	7	6	3	4.3
Milk	5	4	4	3	3	5	4.4
Eggs	10	5	5	3	7	6	6.3
Meat	10	8	8	5	8	10	8.7
Total	8.4	7.8	8.8	6.7	9.8	6.7	7.8

Source: SUA Working System data from 1970 to 2009. Loss rates of all provenance are included.

Note: The category of Western industrialized countries includes Northern America, Western Europe, Australia and New Zealand.

of loss assessments is unclear. In some cases, countries report loss quantities that are evidently calculated by using fixed ratios. Since these ratios remain often unchanged over years, it makes the research behind these loss assessments questionable.

The ideal dataset for our analysis should consist of results from scientific food loss surveys, based on a transparent methodology. Unfortunately, these surveys exist only for some single crops in a few single countries (or even single provinces). This makes it impossible to extrapolate valid predictions for losses of all crops in all countries.[5]

In light of the scarcity of data on food losses, the SUA Working System represents a valuable source of information. These data cover countries from all over the world and many types of food commodities. Even if the loss measurements are not perfectly accurate, they clearly contain real country- and crop-specific information. The latter information will be used to better understand the determining factors behind food losses, and to predict losses for cases where no direct data are available.

As mentioned before, only 4.4% of loss data comes from official sources. Nonetheless the sample size remains acceptable. Since 1970, we count the substantial amount of 11,788 observations for all primary crops in all countries. There are many countries in which data is lacking. The African continent in particular is poorly represented by the data. For this reason, we decided to use observations starting with 1970 for our analysis[6] and additionally we utilize the few data from 2010 to 2012.

Data Extension

Because of the low data availability for Africa and for some commodity groups, additional data on food losses have been collected. Most of the additional data have been taken from national FBS found on the internet. The collection[7] was concentrated on African countries so that an adequate number of Sub-Saharan African countries could enter into the sample. South-Eastern Asia, another region that was poorly represented in the initial data, has been populated by precious data points. Unfortunately the sample could not be sufficiently augmented with data from Northern African countries.

The Final Dataset

Table 3 and 4 show a summary of the loss data which will be utilized in the analysis. The empty cells in the tables indicate that not a single data point is available for the particular item group and geographical region. In Table 3 repeated observations are not considered. E.g., if we have data on wheat in a country for ten years, we count it just as one observation.

Ignoring repeated observations the sample size falls from 11,772 to 879. This means that on average

[5] And even the use of the few available surveys is not unproblematic. As the survey methods are heterogeneous, the results obtained from different survey methodologies are difficult to compare.

[6] There is clearly a trade-off in selecting the time period between the size of the dataset and the up-to-dateness of the data. Keeping also the relatively old data from the 1970's was a compromise in order gain data. In the 1974 the first World Food Conference identified the reduction of post-harvest losses as part of the solution in addressing world hunger food losses, whereas in more recent years less attention was given to food losses (Parfitt et al., 2010). This might be an explanation why for certain commodities and single regions only few data are available in more recent years. This made is inevitable to access older data.

[7] The successful extension of the data set is owe to Rachele Brivio's efforts in collecting data during her internship at ESS.

Table 3: Number of observations in the data used for estimations (without considering repeated observations).

	SSA	Northern Africa & Middle East	Latin America	Western Industrialized Countries	Eastern Eur. & Central Asia	Asia & Pacific	Total	Incl. repeat. obs.
Cereals	24	29	37	140	41	29	313	4,233
Roots & Tubers	13	5	17	19	10	14	80	1,228
Sugar crops		2	2	3	4		11	66
Pulses	5		17	9	2	2	36	470
Tree nuts			4	4	2	2	12	224
Oil crops	7	3	14	14	6	8	52	573
Vegetables	5	2	29	49	19	3	108	1,421
Fruits	10	9	61	46	23	3	154	2,031
Stimulants & spices			8			2	10	108
Milk	2	2	3	14	6	2	29	340
Eggs	2	3	3	15	7	2	32	427
Meat	10	5	16	4	7		42	651
Total	78	60	211	317	127	67	879	
Incl. repeated obs.	164	584	3,418	4,778	1,557	920		11,772

Source: Data from SUA Working System (period of 1970-2012) with an 'empty flag' and additional data from National FBS. Loss rates of value zero are not considered.

we have 13 repeated observations for each country and food item. The largest portion of the data is represented by cereals, followed by fruits and vegetables. For sugar crops, tree nuts, and milk, we have fewer observations. Most observations in the sample come from Latin America and Western industrialized countries. Sub-Saharan Africa, Northern Africa and the Middle East are not as well represented by the data.[8] In theory the dataset constitutes an unbalanced panel, as we observe for many countries commodity-specific losses over several years. However, the time-variation of the countries' commodity-specific losses is small, which indicates that loss rates have not been updated by NSOs each year. Therefore, we treat the dataset like cross sectional data and we control for repeated observations in the estimation.

Table 4 reports the mean loss rates of our using data, where the overall mean is 7.4%. It is important to note that the mean loss figures for food groups or geographical regions are not representative. E.g. the very high loss rate of Latin America (15.9%) is due to high share of fruits and vegetables in the data of this region.

[8]The observations for Sub-Saharan Africa are almost all obtained by the additional data collection. Usually only one observation per crop and country have been collected. The informational content of the data from Sub-Saharan Africa is therefore relatively higher than for other regions, as it contains few repeated observations.

Table 4: Mean loss rates (in %) in the data used for estimations.

	SSA	Northern Africa & Middle East	Latin America	Western Industrialized Countries	Eastern Eur. & Central Asia	Asia & Pacific	Total
Cereals	4	4	6	2	2	4	3.0
Roots & Tubers	6	5	12	7	7	9	8.3
Sugar crops		2	1	1	3		1.9
Pulses	7		4	4	1	1	3.8
Tree nuts			26	3	4	1	10.8
Oil crops	4	4	16	2	1	4	7.6
Vegetables	7	4	31	6	6	5	13.4
Fruits	10	3	25	6	4	5	14
Stimulants & spices			1			0	1.4
Milk	8	1	2	1	0	8	1.7
Eggs	11	6	8	1	1	2	2.5
Meat	4	12	12	5	5		9.3
Total	5.2	4.5	15.9	3.8	3.5	5.0	7.4

Source: Data from SUA Working System (period of 1970-2012) with an 'empty flag' and additional data from National FBS. Loss rates of value zero are not considered.

Methodology

There are many ways to build a statistical model of food losses. Estimation results are dependent on the choice of the functional form and the estimation technique. In this section we propose a rather simple model for estimating food losses. The model is a work-in-progress and may be subject to improvements in the future.

Rationales behind the Model

We assume that each type of food commodity has an idiosyncratic vulnerability of being wasted along the value chain. A perishable food crop like tomato is clearly more sensitive than a non-perishable crop like wheat. The crop's characteristics, such as moisture content, tissue system, etc. strongly influence the risk of losses at storage or transportation. Furthermore, there may be considerable differences in losses across countries. A country's climate, technology, and infrastructure are factors which can influence food losses. The extent of influence may vary across groups of commodities. The quality of transport route is probably more critical for perishable crops and less important for non-perishable crops. The impact of climate also varies across commodities. Some food items might be sensitive to heat, others to humidity or the combination of both. All of these mechanisms require consideration when building a model of food losses.

The Model

We construct an econometric model which takes into account both, crop-specific and country-specific factors that might influence post-harvest losses. The estimation equation has the following form:

$$\begin{aligned} ln(LR_{ijt}) &= f(X_{ijt} + u_{ijt}) \\ &= c + \alpha T_t + \beta RE_j + \delta C_i + \gamma_1 CG_i \times PR_{jt} + \gamma_2 CG_i \times GDP_{jt} + u_{ijt} \end{aligned} \tag{1}$$

The (logarithmized) independent variable LR_{ijt} is the loss ratio of item i in country j at the year t. The loss ratio is calculated according to the definition used for SUA/FBS

$$LR_{it} = \frac{Lossquantitiy}{(Production + Imports + Stockwithdraws)}.^{9} \tag{2}$$

u_{ijt} represents the error term and the explanatory variables in X_{ijt} are the following:

- T_t is a linear annual time trend,

- RE_j is a $r \times 1$ vector of r regional dummy variables (captures regional effects). We categorized the world into 8 regions: 1) Sub-Saharan Africa, 2) Northern Africa, Middle East, Iran, Afghanistan and Pakistan 3) Eastern Asia, 4) South and South-Eastern Asia (incl. Pacific island), 5) Eastern Europe (including Balkan States) and Central Asia, 6) Europe (without Eastern Europe), 7) Latin America and 8) Northern America, Australia and New Zealand.
  This rough regional categorization is driven by data availability. Regions with very low number of observations, like Northern Africa, must be assigned to the most related region. The criterion for merging different regions was mainly the similar socio-economic background.

- C_i is a $c \times 1$ vector of c individual commodity dummy variables. It captures idiosyncratic vulnerability to losses of each single item. This implies the assumption that a commodity has common attributes in each country, and ignores the possibility that a food item can have some specific local features. E.g., maize may be substantially different across the world and therefore loss rates may be different as well.
  The commodity dummies are particularly influential because their estimates affect the prediction of losses for all countries. If only very few data for a commodity are available, these few observations drive the prediction of this commodity for the whole world. The model is therefore particularly sensitive to outliers or measurement error in the loss data.

- CG_i is a $g \times 1$ vector of g dummies which indicate commodity groups. We use the 4 categories: animal products (eggs, milk, meat, fats, etc.), perishable crops (fruits and vegetables), semi-perishable crops (roots and tubers, sugar crops) and non-perishable crops (cereals, pulses, treenuts, oil crops, coffee and cocoa).
  This allows us to capture item group-specific effects of technology or infrastructure. Ideally, we would estimate commodity-specific effects for each item, but we do not have enough observations.

- PR_{jt} is the percentage of paved roads in country j at time t (which reflects transportation infrastructure)[10] ,

[9] However, this way of calculation can be criticized. See Appendix B for a broader discussion of this issue.
[10] Data available at http://data.worldbank.org/indicator/IS.ROD.PAVE.ZS

- GDP_{jt} is country's Gross Domestic Product per capita (in 2005 Dollars)[11] and it should mainly reflect countries technological level.

We allowed the vector CG_i, i.e. the commodity groups, to interact with GDP_{jt} and PR_{jt}. In this way, we allowed the specification to capture crop-specific effects of the countries' infrastructure. This makes the specification more flexible and accounts for the fact that different types of crops are differently vulnerable to standards of storage capacity and transportation. Unfortunately we could not add interactions between region and commodity groups because some regions (like SSA) have no data for certain commodity groups. Furthermore, we do not account for countries' climate.[12]
Estimates are performed simply by OLS using clustered standard errors. The latter clusters are constructed for each commodity and country. In this way standard errors are adjusted for repeated observations. Loss rates of value 0 are omitted. The logarithmic dependent variable assures that outcomes will be strictly positive. After performing the regression, we use the estimation parameters to predict the conditional expectations of losses and their standard errors. We do this for all items for which we have at least one observation in the dataset.

Predictions

The predicted loss rates are simply calculated by using the estimated parameters from equation 1:

$$\widehat{ln(LR_{ijt})} = \hat{f}(X_{ijt}) = \hat{c} + \hat{\alpha}T_t + \hat{\beta}RE_j + \hat{\delta}C_i + \hat{\gamma_1}CG_i \times PR_{jt} + \hat{\gamma_2}CG_i \times GDP_{jt} \tag{3}$$

The values $\widehat{ln(LR_{ijt})}$ are conditional means depending on time, country characteristics and item-specific vulnerability of losses. When the logarithmic predictions are converted to levels we correct for a possible retransformation bias.[13]. Furthermore it is important to note that shocks (like wars, presence of pests, extreme weather phenomena, excess production overcharging storage capacities, etc.) do not enter into this calculation. To estimate these phenomena, data of higher quality are needed.
We also acknowledge that the model might be improved by using more sophisticated estimation techniques.[14] The missing upper bound does not exclude infeasible outcomes, i.e. losses over 100%. However, by trimming prediction at the 98% percentile of the item-specific distribution of predicted values, all predictions remain within a reasonable range.

Results

Table 5 summarizes results. We predicted loss ratios for 123 primary food commodities in 166 countries (20,418 loss parameters). The geographical regions with the lowest loss rates are Western industrialized countries and the countries in transition (both 4.1%). In Latin America (13%) and

[11]Data available at http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD.

[12]In an earlier approach we integrated also countries temperature and rainfall data in equation 1. The prediction produced by this specification had some extreme outliers.

[13]When predicting a logarithmic variable which has to be retransformed into levels one have to take into consideration the error term. Simply exponentiating the logarithmic predictions will produce predictions which can be subject to a considerable large retransformation bias. We use the method proposed by Duan (1983) to adjust for this retransformation bias.

[14]We performed also estimations running several GLM regressions, but results have shown no improvements in terms of goodness of fit.

Table 5: Mean predicted loss rates (in %) by food group and geographical region.

	SSA	Northern Africa & Middle East	Latin America	Western Industrialized Countries	Eastern Eur. & Central Asia	Asia & Pacific	Total
Cereals	7	4	8	3	4	5	5.2
Roots & Tubers	7	6	11	6	4	6	6.8
Sugar crops	2	2	4	2	1	2	2.3
Pulses	5	3	5	2	2	3	3.5
Tree nuts	14	8	16	5	7	10	11
Oil crops	7	4	7	2	3	5	5
Vegetables	19	9	19	6	6	14	13
Fruits	15	7	15	5	5	11	11
Stimulants & Spices	1	0	1	0	0	1	.53
Milk	3	1	3	1	1	2	1.9
Eggs	3	1	4	1	1	2	2.1
Meat	11	6	15	4	3	8	8.6
Total	12	5.8	13	4.1	4.1	8.4	8.4

Source: Own predictions for the year of 2010. Loss rates are predicted for all countries and food items, even if a country does not produce or consume a food item.

Sub-Saharan Africa (12%) predicted loss rates are 3 times higher.
We predict the highest losses for fruits, vegetables and tree nuts, and for stimulants & spices the lowest. The mean loss rate of cereals is 5.2%, ranging from 2.5% in Western Countries and 7.9% in Latin America. The overall mean of loss rates is 8.4%.

How would Losses Change?

Table 6 reports the difference between the predicted loss rates and the loss rates that have been actually assigned in SUA data of 2009. Overall, predicted loss rates are only 0.5% higher than the SUA loss rates. Even if the overall difference is small, losses of certain food groups and geographical regions diverge considerably. Latin America's losses raised by more than 4%. For all other regions the difference in losses is smaller than 2%. The loss rate of Sub-Saharan Africa raised by 1.5%, but losses for cereals, and roots and tubers are lower in the latter region according to our estimates. The prediction for fruits and vegetables are remarkably higher than losses from SUA data in SSA and Latin America. In all regions loss predictions for stimulants & spices, milk and eggs are lower than the current SUA loss rates.

Validation of the Results

Figure 1 plots the predicted loss rates against the observed loss data. On the red line predicted losses equal observed losses. Although one sees that the points are moving upwards along the equality line, data points are very disperse.
The main reason for the dispersion of the predicted data is that the specification does not capture all factors influencing losses in a country. A second reason is the measurement error in loss rates. Even if we consider our data to be reliable, there might be a considerable measurement error in the loss rates reported in National FBS. If this measurement error is not systematically higher or

Table 6: Differences between predicted loss rates and loss rates assign in SUA data of 2009.

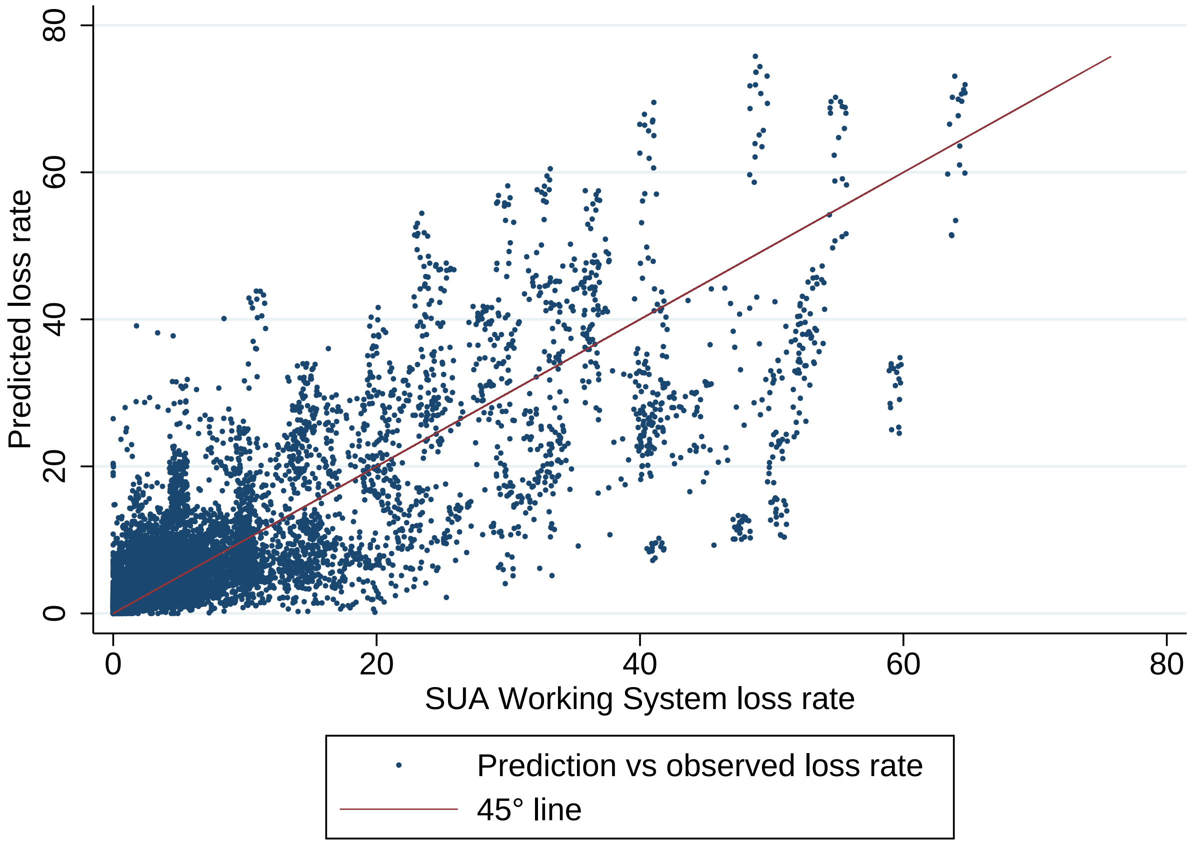
	SSA	Northern Africa & Middle East	Latin America	Western Industrialized Countries	Eastern Eur. & Central Asia	Asia & Pacific	Total
Cereals	-1.5	-2.2	2.0	-2.0	-0.4	-2.2	-1.1
Roots & Tubers	-2.1	1.2	3.7	3.3	-0.7	0.3	0.6
Sugar crops	-4.9	0.7	-1.0	1.1	1.1	-3.2	-2.1
Pulses	0.3	-1.5	1.4	-1.8	-2.3	0.9	-0.5
Tree nuts	5.9	2.5	1.0	-2.3	-0.4	-0.6	0.1
Oil crops	0.7	-0.6	4.3	-1.8	-0.8	1.2	0.6
Vegetables	9.5	0.2	9.8	-1.5	-0.3	3.1	2.5
Fruits	5.2	-1.7	4.5	-2.2	-2.2	0.9	0.8
Stimulants & Spices	-3.1		-2.7	-9.5		-2.0	-2.9
Milk	-2.7	-3.2	-1.0	-0.7	-0.6	-3.0	-2.2
Eggs	-6.8	-3.3	-0.3	-1.7	-1.4	-4.3	-3.5
Meat		1.1	-0.6	2.1	-0.4	-42.4	-1.4
Total	1.5	-1.0	4.5	-1.6	-1.0	0.1	0.5

Source: SUA data of 2009 and own predictions for the year of 2009. Differences are in absolute terms, i.e. they are calculated by subtracting loss percentage of SUA from the prediction's loss percentage. An empty cell means that no losses have been assigned to the particular food group and geographical region in the SUA data.

The very high difference of -42.4% for meat in Asia is due to fact that loss rates in the SUA data for Myanmar are very high. We can not find a explanation for this very high values in the SUA data.

lower than the real value, our points estimates should not be biased. Only the standard errors will be overestimated.

Figure 1: Predicted loss rates vs SUA loss rates



Conclusion and Perspectives

This paper presents a first approach to assess commodity and country-specific loss rates. It is a well known fact that the available knowledge of food loss has some major data gaps (Gustavsson et al., 2011; Parfitt, 2012). In this respect the loss rates estimated by this method provide a valuable reference.
The data used for the estimation consist mainly of loss rates from national FBS, whose quality is not always perfect. Therefore, the results should be interpreted with caution.
The approach should not be considered as concluded but open for both methodological refinements and improvements of the data quality. Several initiatives to gather data on losses are developing at the present time[15] and future attempts in estimating food losses should make use of the updated data.

References

Aulakh, J. and Regmi, A. 2013. "Post Harvest Food Losses Estimation - Development of Consistent Methodology " mimeo.
http://www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/...

Duan, N., 1983. "Smearing estimate: A nonparametric retransformation method." *Journal of the American Statistical Association* 78, 605-610

FAO, 1972. "Technical Conversion Factors For Agricultural Commodities." *FAO Statistics Division.* http://www.fao.org/fileadmin/templates/ess/documents/methodology/tcf.pdf

Gustavsson, J., Cederberg, Ch., Sonesson, U., van Otterdijk, R. and Meybeck, A. 2011."Global Food Losses and Food Waste: Extent, Causes and Prevention." *FAO*, Rome.

Parfitt, J., Barthel, M., and S. Macnaughton, 2010. "Food waste within food supply chains: quantification and potential for change to 2050" *Philosophical Transactions of the Royal Society* 365, 3065-3081

Rembold. F., R. Hodgex, M. Bernard, H. Knipchild, and O. Leo. 2011. "The African Postharvest Losses Information System (APHLIS)." *JRC Scientific and Technical Reports.* Luxemburg.

Shapori, S., and H., Khaira, 2012. "Waste Estimation in FAOSTAT Food Balance Sheets" *Paper prepared for the FAO CapEx project.*

[15] Regional initiatives of loss assessments like APHLIS (Rembold et al., 2011).

Appendix A - Establishing a working routine for ESS

Guide for ESS Staff Compiling the FBS

The output of the procedure is a table that contains predicted loss ratios, together with lower and upper 95%-confidence bands. We added also the loss ratio from FAO (1972) if it is available for the particular crop and country (click here to open the table).
The width of the confidence bands provides the accurateness of the prediction. However, in our case most confidence bands are very large and therefore almost uninformative. They are very large because forecast errors big. The reason for this may be a measurement error in the food losses data. In the latter case forecast errors are overestimated.
It is recommended to statisticians who compile the FBS to choose the point estimate of the prediction. However, in case of doubt, statisticians can choose a value within the confidence bands. Furthermore, if there are doubts about the validity of all values within the whole confidence interval, statisticians might deviate from the recommended output.
Unrealistic predictions might arise for two reasons. First, shocks might occasionally occur in a country. As mentioned above the model does not take these shocks into account. Second, since predictions are just conditional means, they depend mainly on country characteristics. An unusual combination of a country's characteristics might produce unrealistic predictions. Furthermore, for some commodities only few data are available and estimates for these crops are strongly affected by these few observations. As a result, the predictions of the latter are more prone to producing outliers.

Appendix B - Discussion about the Definition of Loss Rates

As mentioned above, within the FAOSTAT system loss rates are defined as

$$LR_{it} = \frac{Lossquantitiy}{(Production + Imports + Stock\,withdraws)}, \tag{4}$$

where $Production$ contains also $Exports$ and $Stock\,increases$.
Losses in the FBS are defined as the amount of food lost after the moment that production is recorded until it reaches consumers (FAO, 1972). For various reasons formula 4 can be only a proxy for the real loss rate.

Losses along the Value Chain

Food losses occur at various steps along the production chain (for a broader review of this issues see Aulakh, 2013). The three most relevant value chain steps for losses are: 1) at farm level, 2) at marketing (transportation to and storage at market) and 3) at the storage of stocks. Exports, imports and stocks withdraws pass through very different steps of the value chain.

Double Counting of Losses

Formula 4 implies that stocks, exports and imports suffer losses twice:

- Stocks: From stocks losses are discounted 1) when the commodity designated for stocks is produced (or imported) and later put to stocks (as they were originally part of *Production* or *Imports*) and 2) when stocks are withdrawn losses are detracted another time.
- Exports and imports: Losses for exports and imports are detracted 1) from exports in the country of origin (as it is part of *Production*) and 2) at another time in the country of destination (where it is part of *Imports*).[16]

Assigning double losses for stocks might not be too far from the reality, but the current loss formula is likely to overestimate losses of imports and exports. The value chain of imports is clearly shorter in the destination country than for commodities which are produced and distributed in the same country. When imported goods reach their destination, they have already passed through several steps of the value chain. Furthermore, imported goods might be better packed than commodities traded interior, which will additionally lower their propensity to losses. A similar argument holds for exports. Losses of exported goods occurring in the country of origin might be lower than for commodities traded interior.

The Role of Own Production

An additional factor which is ignored in the SUA/FBS loss framework is own production. Crops cultivated at home or gathered wild have a shorter value chain than food sold at the market. Losses in countries with an elevated share of own produced food should therefore be lower.

[16] Assuming that the commodity are traded just once.