Using Uncertainty and Sobol' Sensitivity Analysis Techniques in the Evaluation of a Composite Provincial Level Food Security Index

Christian P. Umali and Felino P. Lansigan
Institute of Statistics, University of the Philippines Los Baños

A composite provincial level food security index can measure food security at the sub-national level which can be helpful in policy making. However, it can be non-robust due to the uncertainties involved in the choice of input factors to be used in its construction, namely: different sources of data, normalization methods, weighting schemes, aggregation systems, and the level of importance placed on the different dimensions of food security i.e. availability, accessibility, utilization, and stability. In this study, uncertainty analysis technique was employed in order to assess the robustness of the index constructed through the conventional approach. Sensitivity analysis was done in order to quantify the influence each factor has in the index building process, identify factors that should be prioritized and can be fixed in the successive development of the index, and determine which levels of factors are responsible for producing the desirable model outcomes. The study had been exploratory in considering the ratio with mean normalization technique and a combination of the additive and geometric aggregation methods as potential inputs in the index construction. In computing for the Sobol' sensitivity indices, the formulas suggested by Jansen et al. (1994) and Nossent and Bauwens (2012) were investigated and compared under varying sample sizes. The results can provide a more appropriate choice of procedure in computing for the Sobol' sensitivity indices at an optimum sample size, and likewise insights in the future development of a uniform and more defensible composite provincial level food security index.

Keywords: uncertainty analysis, sensitivity analysis, Sobol' sensitivity index, Monte Carlo integral, composite index, food security

1. Introduction

The paradox in the treatment of composite indices is that composite indices, simply put, are just single-valued numbers. Yet, it is able to communicate to the readers a meaningful array of interpretations because of its ability to summarize complex or multidimensional issues, reduce the size of a list of indicators, and at the same time, provide a big picture by including more information within its existing size limit (Saisana and Tarantola, 2002 as cited in Saisana et al., 2005).

Composite food security indices are found in the same vein as they examine food security comprehensively across many viewpoints that look not just at the singular problem of hunger but equally on other similarly important underlying factors. Food security, being a holistic concept (Epp, 2010), is compounded by many factors and related issues that range from persistent poverty, undernourishment, climate change, water scarcity, etc. (Cargill, Inc., 2014).

While the convention in constructing the index is based on a simplified approach (use of min-max normalization technique, imposition of equal weights on the food security dimensions, and application of an additive aggregation method), in general and for most composite indices, there is no standard way of constructing the index. A composite index calls for subjective judgments on the many stages or input factors involved in its construction, like in the selection of sub-indicators or data to use, data editing methods, normalization, choice of weighting scheme, selection of aggregation system, choice of dimension or sub-indicator weights, choice of experts, etc. (OECD, 2008; Saisana et al., 2005). The construction of a composite index can actually be viewed as a model (Saltelli, 2002; Chan et al., 1997), where the output is the index, and the inputs are the stages involved or factors.

The uncertainties or arbitrariness involved in index construction puts into issue the robustness or biasedness, and transparency of constructed indices. Consequently, with this problem, composite indices can send misleading and non-robust policy messages (Saisana and Tarantola, 2002 as cited in Saisana et al., 2005). Any model's quality depends on the soundness of its assumptions. Thus, an assessment of the uncertainties and subjective choices taken along the modeling process of an index is of great importance (OECD, 2008).

Understanding the mechanics behind the construction or model of a particular index is equally important as it can aid in the index's further development thru a guided selection and prioritization of model inputs responsible for producing robust results, and model simplification.

In this study, uncertainty analysis (UA) technique was employed in order to assess the robustness of a composite food security index. Sensitivity analysis (SA) was also applied in order to measure the amount of influence each input factor exerts on the index model, and interpreted within the frameworks of factor prioritization, factor fixing, and factor mapping settings.

The analysis was framed on the performance of a composite provincial level food security index constructed via the conventional approach using the Philippines data, and carried out in an almost similar fashion as Saisana et al. (2005) in their evaluation of the United Nation's Technology Achievement Index (TAI).

The sensitivity analysis adopted here, in particular, is based on the calculation of the Sobol' sensitivity indices (first-order and total-effect sensitivity indices) which require a Monte Carlo integration procedure and huge number of model evaluations to realize.

Several approaches or formulas in computing for the Sobol' sensitivity indices exist. We considered the formulas by Nossent and Bauwens (2012), which were basically a modification of the formulas by Saltelli (2002), and an earlier formula presented in Bilal (2014) which we attributed to Jansen et al. (1994).

In the estimation of the Sobol' indices using Monte Carlo integration, there are documented cases of inaccurate or incoherent computed sensitivity indices and slow convergence. In some instances, negative sensitivity indices were calculated when ideally they should be non-negative (see Glen and Isaacs, 2012; Saltelli et al., 2004; MOEA Framework, ____). Nossent and Bauwens (2012) and Mokhtari and Frey (2005) encountered unrealistic results and slow convergence before a suitable transformation was introduced to their model outputs. Similarly, we obtained unexpected results (e.g. first-order sensitivity index greater than its counterpart total-effect sensitivity index) when we initially applied the formula by Saltelli (2002). An investigation of the different formulas at varying sample sizes will help to settle the appropriate approach and optimum sample size required in the estimation of the Sobol' indices.

In continuously improving the design of the index, other ways or methods of construction which could potentially produce results that are more reflective of the actual state of food security should be explored. In this line, we put forward an aggregation technique that is a hybrid of an additive and geometric aggregation methods. The proposed method intends to regard the utilization dimension of food security more indispensable over the other dimensions. Likewise, the ratio with mean method was considered and examined as a potential normalization technique in index construction.

The objectives of the study are: (a) to explore other possible ways or methods that can be used as inputs in the construction of a composite food security index; (b) to assess the robustness of the provincial level food security index constructed via the conventional approach; (c) to account for the amount of variation in the food security index model that is due to the specific input factors, to identify which factors demand more priority for future exploration and study, which factors can be fixed in subsequent research or analyses, and which levels of factors are responsible for producing the desirable model outcomes; and (d) to identify a more appropriate approach and optimum sample size in computing for the Sobol' sensitivity indices.

2. Concepts

2.1. Uncertainties in Constructing Food Security Indices

The literature is host to a lot of studies that demonstrate the variation in the approaches in constructing food security indices owing mainly to the subjective decisions on many factors taken in the process. A generalization that can be drawn in the review is that no standard procedure exists that tell how the index should be calculated.

The variation in data to use can be attributed to the evolving notion of food security (see Iqbal and Amjad, 2010; Ahlawat and Kaur, 2013), variation in the identification of the food security dimensions or sub-indicators, level of food security being measured, availability of data in a country or area, and issues arising from correlations of data or redundancy.

The popular definition of food security is:

Food security exists when all people, at all times, have physical and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life (World Food Summit, 1996 as cited in Maxwell, 2013).

But FAO again in 2001 (see Ahlawat and Kaur, 2013), in 2002 (see Wineman, 2014) and again in 2004 (see Magombeyi, et al., 2013) issued almost very similar yet not exact definitions. The version in 2002 has the word "social" added to it (see FAO, 2002 as cited in Wineman, 2014). There are also definitions or notions of food security at the community level (see Hamm and Bellows, 2003 as cited in Epp, 2010), society (see Epp, 2010), household (see Gross, 2002 as cited in FAO et al., 2010), and individual (see FAO et al., 2010).

The levels of food security indices done and encountered in the literature were at the global, regional, country level (see EIU, 2014; Iqbal and Amjad, 2010; PSA, 2012), district, municipality level (see Ahlawat and Kaur, 2013; Alemu, 2015; FAO et al., 2010; Magombeyi et al., 2013), and household level (see Ndhlevel et al., 2012; Sajjad and Nasreen, 2014; WFP and GSS-MFA, 2012; FNRI-DOST, 2015). Depending on the level of assessment, the required indicators and sources of data to come up with the indicators may vary. There is a wide discrepancy between food security at the national, community, and household level (Anderson, 1990; Du Toit, 2011 as cited in Ndhlevel et al., 2012).

Food security is a multifaceted condition, thus, it contains many dimensions (Maxwell et al., 2013; Wineman, 2014). There are four popularly known dimensions of food security as far as the literature is concerned, namely: (1) *availability* of food - the regular availability in sufficient quantities of foodstuffs of appropriate quality in accordance with tastes and preferences of the people; (2) *accessibility* of food - depends on factors like incomes, sources of income including remittances, income disparities, real food prices, landlessness, gender, literacy, and employment status; (3) food *utilization* - involves the effective biological utilization through adequate food, clean water, sanitation and healthcare for attainment of nutritional well-being; and (4) *stability* - implies that people have all time access to adequate food without involving any risk of losing physical availability and economic access to it as a result of economic shocks and resulting higher prices, natural disaster and wars (Iqbal and Amjad, 2010; Magombeyi et al., 2013; WFP and GSS-MFA, 2012).

Some studies recognize or use only as few as two dimensions or indicators in determining food security status, for example, Suharyanto et al., (2014) utilized only food expenditure and energy consumption adequacy, Alemu (2015) used per capita calorie availability and households' per capita calorie consumption needs, while others avail of as many as five dimensions: *quantity*, *quality*, *acceptability*, *safety* and *stability* (see Coates, 2013 as cited in Maxwell et al., 2013). Still, in a significant number of studies, the *stability* dimension is not recognized (see Ahlawat and Kaur, 2013; Sajjad and Nasreen, 2014; PSA, 2012; Suharyanto et al., 2014; Barrett, 2002 as cited in Wiesmann et al., 2009), or the exact terminology or concept of *stability* was not mentioned (see EIU, 2015; Smith and Subandoro 2007).

The Food and Nutrition Research Institute (FNRI), the primary government agency concerned with the research and study of food and nutrition in the country published in 2015 estimates of the prevalence of food secure households per province based on households' responses to nine (9) key food insecurity questions from the Household Food Insecurity Access Scale (FNRI-DOST, 2015).

The correlation of data can be viewed as a redundancy of information that is something to correct for (OECD, 2008) and so in some situation this would call for data selection or reduction.

Non-participatory assignment of weights on food security indicators includes the imposition of equal weights on the dimensions, which is a common practice observed, and the employment of Principal Component Analysis (PCA) technique (see Wineman, 2014; Cavatassi, 2004 as cited in Wineman, 2014). But these methods are argued to lack the benefits of the participatory approaches which are able to cater for the stakeholders' contribution and expectations in the design of the index and the associated policies (OECD, 2008).

Summing of the dimension scores is the conventional method of aggregation in the food security index construction. However, this method is argued to have a full compensability feature as opposed to the geometric aggregation method which possesses some non-compensatory attributes (OECD, 2008). In this study, an aggregation method that is a combination of additive and geometric methods was proposed which aims to highlight the importance of the *utilization* dimension and likewise to explore other options for aggregation and not just limit the problem within the conventional. The proposal is an attempt to model the hypothesis or belief that a first-hand evidence of food security is the population's nutritional status. The proposed method was designed to not allow the *utilization* dimension to be easily compensated by the other dimensions.

2.2. Uncertainty and Sensitivity Analysis

The process of computing a composite index can be treated as a model, where the output (y) is considered as a random variable (which can be the composite index or any model output that can be derived from the computation of the index, such as the corresponding rank of a country or province, shifts, etc.), and x_1 – e.g. selection of sub-indicators, x_2 – data selection, x_3 – data imputation method, x_i, … x_k are the input factors (OECD, 2008; Saltelli, 2002; Chan et al., 1997). That is,

$$y = f(x_1, x_2, \dots x_k)$$

Uncertainty analysis (UA) aims to quantify the overall uncertainty in the model as the result of the uncertainties in the inputs. (OECD, 2008; Saisana et al., 2005). The Monte Carlo approach to UA proceeds first by assigning a probability distribution to each input factor $\boldsymbol{x}_i$. Next, n randomly generated combinations of the independent input factors $\boldsymbol{x}^l$, l = 1,2,…n (a set $\boldsymbol{x}^l = x_1^l, x_2^l, \dots x_k^l$) are used in the model and evaluated to produce the scalar output $\boldsymbol{y}^l$, l = 1,2,…n. The sequence $\boldsymbol{y}^l$ gives the empirical probability distribution of the output $\boldsymbol{y}$.

Given a computational model (see first equation), where the input factors are uncorrelated, y can be seen as the realization of a stochastic process obtained by sampling each of the $\boldsymbol{x}_i$ from its marginal distribution. The joint probability distribution of the input variables or the model output $\boldsymbol{y}$ is (Saltelli, 2002):

$$P(x_1, x_2, \dots, x_k) = \prod_{i=1}^{k} p_i(x_i)$$

Sensitivity analysis (SA), on the other hand, aims to quantify how much each individual uncertainty factor contributes to the output variance (OECD, 2008; Saltelli, 2002; Saisana et al., 2005). Normally, the important and practical statistics produced by the Sobol' SA are the first order sensitivity indices ($\boldsymbol{S}_j$) and the total effect sensitivity indices (S_j^T). The first order sensitivity index $\boldsymbol{S}_j$ tells about the contribution of the input $\boldsymbol{x}_j$ to the output variance on its own or in solo. The total effect sensitivity index, S_j^T, tells about the contribution of the input $\boldsymbol{x}_j$ to the output variance on its own and in cooperation with the other input factors. For a given factor $\boldsymbol{x}_j$, a significant difference between S_j^T and $\boldsymbol{S}_j$ conveys an important role of interactions for that specific factor and other possible factors in $\boldsymbol{y}$.

The Sobol' sensitivity indices are basically ratios of variances of $\boldsymbol{y}$ and are commonly estimated via numerical approximation or Monte Carlo integration.

Nossent and Bauwens (2012) studied the convergence of the Sobol' indices computed based on the different formulas for $\hat{E}^2(y)$ and $\hat{V}(y)$ in combination with the formulas of Saltelli (2002). The results showed that for certain combinations of $\hat{E}^2(y)$ and $\hat{V}(y)$ applied to the first-order and total effect sensitivity indices formulas of Saltelli (2002), convergence and stability of the computed indices are demonstrated at sample sizes beginning at 4000 to 6000 (Nossent and Bauwens, 2012).

Let M_1 and M_2 be two randomly generated matrices of input factors where M_1 can be thought of as the "sample" matrix and M_2 as the "re-sample" matrix:

$$\boldsymbol{M}_1 = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & & & \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, \quad \boldsymbol{M}_2 = \begin{bmatrix} x'_{11} & x'_{12} & \dots & x'_{1k} \\ x'_{21} & x'_{22} & \dots & x'_{2k} \\ \dots & & & \\ x'_{n1} & x'_{n2} & \dots & x'_{nk} \end{bmatrix}$$

Both M_1 and M_2 consist of n randomly generated combinations of the levels of the k input factors. This can be accomplished by repeatedly sampling from each of the 's.

Another matrix N_j is defined as:

$$\boldsymbol{N}_j = \begin{bmatrix} x'_{11} & x'_{12} & \dots & x_{1j} & \dots & x'_{1k} \\ x'_{21} & x'_{22} & \dots & x_{2j} & \dots & x'_{2k} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x'_{n1} & x'_{n2} & \dots & x_{nj} & \dots & x'_{nk} \end{bmatrix}$$

where N_j is a matrix where all factors except x_j are re-sampled, the column **j** in Nj comes from M1 and the rest come from M2.

Let $\boldsymbol{a}_{i_1 i_2 \ldots i_s}$ denote the vector of length n containing the model evaluations corresponding to the rows of the input factor matrix $\mathbf{N}_{i_1 i_2 \ldots i_s}$. Similarly, the matrix $\mathbf{N}_{i_1 i_2 \ldots i_s}$ is obtained as the matrix M1 with all the columns substituted except $i_1, i_2, \ldots, i_s$ by the corresponding columns of matrix M2.

For the Monte Carlo integral computation of the first-order sensitivity indices in Saltelli (2002), for $\boldsymbol{k} = 8$ as the case in this study, Nossent and Bauwens (2012) recommended using the following formulas:

$$\hat{S}_j = \frac{\left(\frac{1}{n-1}\right) \boldsymbol{a}_{12345678}{}' \boldsymbol{a}_j - \hat{E}^2(y)}{\hat{V}(y)}$$

$$\hat{E}^2(y) = \left(\frac{1}{n}\right) \boldsymbol{a}_{12345678}{}' \boldsymbol{a}_0$$

$$\hat{V}(y) = \left(\frac{1}{2(n-1)}\right) \boldsymbol{a}_c{}' \boldsymbol{a}_c - \left(\frac{1}{2n} \boldsymbol{a}_c{}' \boldsymbol{J}\right)^2$$

For the Monte Carlo integral computation of the total effect sensitivity indices in Saltelli (2002), for $\boldsymbol{k} = 8$ as the case in this study, Nossent and Bauwens (2012) recommended using the following formulas:

$$\hat{S}_j^T = 1 - \frac{\left(\frac{1}{n-1}\right) \boldsymbol{a}_0{}' \boldsymbol{a}_j - \hat{E}^2(y)}{\hat{V}(y)}$$

$$\hat{E}^2(y) = \left(\frac{1}{2n} \boldsymbol{a}_c{}' \boldsymbol{J}\right)^2$$

$$\hat{V}(y) = \left(\frac{1}{n-1}\right) \boldsymbol{a}_0{}' \boldsymbol{a}_0 - \left(\frac{1}{n} \boldsymbol{a}_0{}' \boldsymbol{J}\right)^2$$

where $\boldsymbol{a}_c$ is $\boldsymbol{a}_{12345678}$ and $\boldsymbol{a}_0$ combined.

An alternative Monte Carlo approach to computing the sensitivity indices was detailed in Zheng & Rundell (2006) (Bilal, 2014) and applied by Bilal (2014). The integral representation used however was a slight generalization of the formulas that were utilized as early as in Jansen et al. (1994) and Šaltenis and Dzemyda (1982) (Sobol', 2001). The proof to this approach was provided in Sobol' (2001). For the sake of uniform reference, we referred to this alternative method or approach as the method attributed to "Jansen et al. (1994)" just for labeling purposes. The formulas is given as follows:

$$E(y) = f_0 = \int f(\boldsymbol{x})\, d\boldsymbol{x} \approx \left(\frac{1}{n}\right) \sum_{r=1}^{n} f(\boldsymbol{x}_r)$$

$$V(y) = \int f^2(\boldsymbol{x}) d\boldsymbol{x} - f_0^2 \approx \left(\frac{1}{n}\right) \sum_{r=1}^{n} f^2(\boldsymbol{x}_r) - f_0^2$$

$$S_j = \frac{V(y) - \frac{1}{2} \int \left[f(\boldsymbol{x}) - f\left(x_j, \boldsymbol{x}'_{-j}\right)\right]^2 dx_j d\boldsymbol{x}'_{-j}}{V(y)} \approx \frac{V(y) - \frac{1}{2n} \sum_{r=1}^{n} \left[f(\boldsymbol{x}_r) - f\left(x_{jr}, \boldsymbol{x}'_{-jr}\right)\right]^2}{V(y)}$$

$$S_j^T = \frac{\frac{1}{2} \int \left[f(\boldsymbol{x}) - f\left(x'_j, \boldsymbol{x}_{-j}\right)\right]^2 dx' d\boldsymbol{x}_{-j}}{V(y)} \approx \frac{\frac{1}{2n} \sum_{r=1}^{n} \left[f(\boldsymbol{x}_r) - f\left(x'_{jr}, \boldsymbol{x}_{-jr}\right)\right]^2}{V(y)}$$

Variance-based sensitivity analyses such as the calculation of the Sobol' indices are computationally intensive. The formula by Jansen et al., (1994) requires $2\boldsymbol{n}(\boldsymbol{k+1})$ model evaluations to compute for the full set of first-order and total effect sensitivity indices, while the formula by Nossent and Bauwens (2012) requires $\boldsymbol{n}(\boldsymbol{k+2})$ model evaluations to accomplish the same task.

2.3. Assumptions and Computational Issues

The formulas for the sensitivity indices described earlier were derived anchored on the assumption that the input factors are orthogonal or uncorrelated. In the case of correlated factors, the first-order and total effects sensitivity indices calculated under an orthogonal inputs case still remain as valid measures and can be interpreted under the different "defensible settings" for sensitivity analysis (OECD, 2008; Saltelli, 2002; Saisana et al., 2005; Saltelli et al., 2004). These "defensible settings" which provide interpretations of the sensitivity indices under different frameworks are as follows: Factor Prioritization (FP) setting; Factor Fixing (FF) setting; and Factor Mapping (FM) setting. We refer the readers to Saltelli et al. (2004) where these are amply described.

There are cases of sensitivity indices that were computed to be negative (see Glen and Isaacs, 2012; Saltelli et al., 2004; MOEA Framework, ____). Such cases often happen when the true or analytical indices being computed are close to zero or the input parameters are uninfluential. This problem was deemed as a sampling error and can be addressed by increasing the sample size (Saltelli et al., 2004).

The accuracy and convergence of the computed index can likewise be compromised when f_0 or the mean of the output is large (Sobol', 2001; Nossent and Bauwens, 2012; Mokhtari and Frey, 2005). A remedy is to subtract from $f(x)$ an approximate value c_0 of f_0 (Sobol', 1990 as cited in Sobol', 2001). Transformation or normalization of $f(x)$ before computing the sensitivity indices can likewise correct the error (see Nossent and Bauwens (2012); Mokhtari and Frey (2005)).

3. Data and Methodology

3.1. Input Factors

Input factors that reflect the uncertainties involved in the construction of the food security index were acknowledged. These include the following: X_1 – inclusion/exclusion of sub-indicators or dimensions; X_2 – normalization method; X_3– weighting scheme; X_4 – aggregation system; and X_5, X_6, X_7 and X_8 – for the choice of weights of the dimensions of food security, namely: *availability*, *accessibility*, *utilization* and *stability*, respectively. The input factors are shown in Table 1.

Table 1. The input factors acknowledged in the construction of the food security index (PCA = Principal Components Analysis; AHP = Analytic Hierarchy Process; BAP = Budget Allocation Procedure)

Input factor	*Definition*	*Assigned PDF*	*Range*
X_1	Inclusion/exclusion of sub-indicators	Uniform, discrete	1–Complete sub-indicators 2–Complete sub-indicators with *stability* removed 3–Regional level sub-indicators removed 4–PCA-Correlation based sub-indicators 5–PCA-Correlation based sub-indicators with *stability* removed
X_2	Normalization method	Uniform, discrete	1–Min-max 2–Z-score 3–Ratio with mean
X_3	Weighting scheme	Uniform, discrete	1–Equal weights 2–AHP 3–BAP
X_4	Aggregation system	Uniform, discrete	1–Additive 2–Geometric 3–Combination of additive and geometric
X_5	Weights for *Availability*	Uniform, discrete	[1,2,…,11]
X_6	Weights for *Accessibility*	Uniform, discrete	[1,2,…,11]
X_7	Weights for *Utilization*	Uniform, discrete	[1,2,…,11]
X_8	Weights for *Stability*	Uniform, discrete	[1,2,…,11]

X_1 is composed of 5 dataset versions representing the variations in the selection of which sub-indicators or data to use in the index computation. The *PCA-Correlation based sub-indicators* were arrived at by utilizing PCA and correlation analysis in order to remove some redundancy in the data. The details of this procedure was not shown here to save space.

Descriptions of the sub-indicators or data per food security dimension utilized in this study are presented in the Appendix in Tables A-1, and A-2. These sub-indicators comprising more than 70 variables are deemed the latest, relevant, exhaustive, and available data that came from official statistics or estimates produced by various government and non-government agencies that spans from 2006 to 2015. Weights of the sub-indicators per dimension and dataset version were assigned subjectively but judiciously. Whenever a sub-indicator is eliminated in a dimension under a dataset version, its weight is set to zero and the remaining sub-indicators are assigned the weights found under the *Complete sub-indicators* (X_1=1) then re-scaled to sum to unity. The weights assigned to the sub-indicators per dimension and dataset version were not shown here in order to save space.

X_2 is composed of *min-max*, *z-score*, and *ratio with mean* normalization methods.

Min-max: $$\frac{G_i - Min(G)}{Max(G) - Min(G)}$$

Z-score: $$\frac{G_i - \bar{G}}{\sqrt{V(G)}}$$

Ratio with mean: $$\frac{G_i}{\bar{G}}$$

where G_i is the raw data or sub-indicator score of a province; $V(G)$ and $\bar{G}$ are the corresponding variance and mean of the sub-indicator respectively.

X_3 represents the weighting scheme which includes the assignment of *equal weights* to the dimensions, weighting based on *Analytic Hierarchy Process* or *AHP* (Saaty, 1990; University of Siena, _____), and weighting based on *Budget Allocation Procedure* (*BAP*) which were employed in conjunction with the weights of the food security dimensions and elicited from interviews of identified experts.

X_3 covers the purely *additive*, purely *geometric*, and a *combination of additive and geometric* aggregation systems. If we let $N_{q,p}$ and w_q to be the normalized score and assigned weight respectively to a food security dimension q of a province p, and let q=u for the *utilization* dimension, then the composite index (CI_p) will be computed as follows under the various aggregation systems:

Additive: $$CI_p = \sum_{q=1}^{Q} w_q \cdot N_{q,p}$$

Geometric: $$CI_p = \prod_{q=1}^{Q} (N_{q,p})^{w_q}$$

Combination:
$$CI_p = \left\{(N_{u,p})^{w_u}\right\} \cdot \left\{(\sum_{q=1}^{Q-1} (\frac{w_q}{\sum_q^{Q-1} w_q})(N_{q,p}))^{1-w_u}\right\}$$

X_5, X_6, X_7 and X_8 were each composed of 11 possible weights elicited from 11 food security experts or resource persons. The resource persons interviewed are attached to research, agricultural, environmental advocacy, and academic institutions. The elicited weights are presented in Table 2.

Table 2. Weights of the dimensions of food security (in percentage) elicited from the interviewed experts

Expert	*Budget Allocation Procedure*				*Analytic Hierarchy Process*				
	Availability	*Accessibility*	*Utilization*	*Stability*	*Availability*	*Accessibility*	*Utilization*	*Stability*	*Consistency Index*
1	15	15	10	60	13.23	13.82	4.27	68.67	0.10
2	20	30	30	20	9.92	24.40	24.40	41.29	0.02
3	40	30	15	15	62.10	23.55	10.77	3.59	0.42
4	30	20	20	30	12.80	12.80	4.23	70.18	0.12
5	30	20	30	20	20.26	22.64	40.84	16.25	1.69
6	40	30	15	15	48.42	27.15	10.63	13.80	0.04
7	50	20	20	10	46.07	30.66	20.00	3.28	0.67
8	30	25	25	20	52.96	28.19	13.19	5.66	0.28
9	30	30	10	30	31.71	31.71	4.88	31.71	0.00
10	25	25	25	25	25.00	25.00	25.00	25.00	0.00
11	20	60	10	10	22.65	59.34	4.60	13.40	0.21

3.2. Computation of Food Security Indices and Model Outputs Investigated

Food security indices were computed based on the combination of the levels of the input factors generated in UA and SA. The series of steps in computing the composite food security index per province p (CI_p) given a particular set l ($l = 1,2,\dots n$) of input factors ($\boldsymbol{x}^l = x_1^l, x_2^l, \dots x_8^l$) is illustrated in Figure 1. The raw data or sub-indicators, whenever applicable, were first deflated by the provincial population and inverted so that they will reflect a positive or direct relationship to food security. Different kinds of data were dealt with different methods of data inversion. Next, sub-indicators were chosen which is dependent on the version of the dataset that was selected. Sub-indicators were then normalized and their weighted sum derived to come up with the dimension scores. Dimension scores were then combined together into just one score (per province), which was dependent on the choice of weighting scheme, weight values of the dimensions, and aggregation system selected. Recalculation of the dimension weights was done whenever necessary to maintain the important property that the weights should all sum to unity. Finally, since different normalization methods were utilized, the raw food security scores were re-scaled for comparability purposes. The final re-scaling procedure done was a version of min-max but utilizes different potential minimum and maximum scores which are dependent on the initial method of normalization applied to an index.

There are cases when no index can be computed simply because the combination of the levels of the input factors selected is not conformable. The z-score normalization can produce negative scores and therefore is not compatible with the geometric and combination of additive and geometric aggregation systems.

The conventional or traditional way of constructing the food security index (I_c) entails utilizing a *min-max* normalization technique, an imposed equal weighting scheme (non-participatory), and a simple additive aggregation system (compensatory). Further, in this study, we used the complete set of sub-indicators in order to build it.

Aside from the indices, model outputs that were derived and observed in this study were the corresponding provincial ranks, relative shift in the position of the entire system of provinces ($\bar{R}_s$) with respect to a reference point; see Saisana et al. (2005)), and the difference between ranks of two provinces (D_R):

$$\bar{R}_s = \frac{1}{M}\sum_{p=1}^{M}\left|RefRank_p - Rank(CI_p)\right|$$

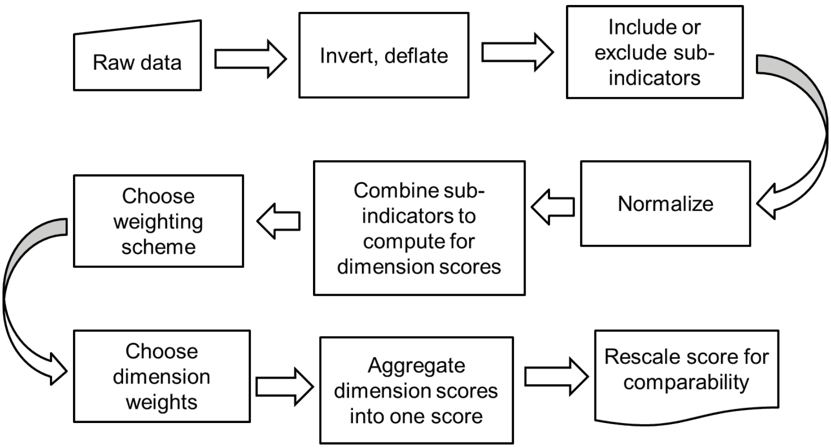


Figure 1. Process of computing the food security indices

$\bar{R}_s$ captures the relative shift in the position of the entire system of provinces from a reference index or rank in a single number, where *RefRank*$_p$ were the resulting rankings of provinces induced by the conventional index I_c or the Monte Carlo *Median Rank*. $\bar{R}_s$ relative to I_c reflects the stability or "robustness" of the conventional index when different inputs are used to construct the food security index.

Analysis of ranks was given more prominence in this study because ranks tend to be resilient over many and different data transformations which are broadly present in this study. Further, by parallelism, the combination of methods that yields the desirable ranks can be the same set of methods that governs the desirable indices.

3.3. Stages of Analysis

The flow of analysis and discussions in this study was designed and presented in three sequential parts: (1) a *Calibration Stage Analysis*; (2) a *Post-Calibration Stage Analysis*; and (3) a *Factor Mapping Stage Analysis*.

Under the *Calibration Stage Analysis*, the approach was more exploratory. The conventional index I_c for all provinces was computed. Uncertainty analysis of the food security indices was performed using a base sample of 7,000 which is cognizant of the results in Nossent and Bauwens (2012). In all stages of analysis, the "sample" and "re-sample" matrices M_1 and M_2 were generated using simple random sampling. The extreme values produced by the RWM normalization technique was inspected. Sensitivity analysis of $\bar{R}_s$ with respect to the conventional index I_c was done employing three formulas (Nossent and Bauwens (2012) with and without adjustment in $f(x)$, and Jansen et al., (1994)) at different sample sizes: 4,000; 5,500; 7,000; 9,000; and 11,000. Subtraction of an estimate of the mean of the model output ($\hat{f}_0$) from $f(x)$ was the adjustment done to $f(x)$ for the sensitivity analysis as the resulting mean of the new function was effectively reduced and its variance maintained.

Under the *Post-Calibration Stage Analysis*, UA and SA were again performed. A new set of the "sample" and "re-sample" matrices, M_1 and M_2 was generated as the ratio with mean (RWM) normalization method was now dropped from the input factors and the sample size was pegged at 9,000. Sensitivity analysis of three model outputs, i.e. $\bar{R}_s$ with respect to the conventional index (I_c), difference between ranks of two provinces (D_R), and $\bar{R}_s$ with respect to the *Median Rank*, was done using only the formula by Jansen et al., (1994). These model outputs were subjected to SA in the quest for consistent results. The results of the sensitivity analyses were interpreted, with more emphasis, within the frameworks of factor prioritization (FP), factor fixing (FF), and factor mapping (FM) settings. The *Median Rank* is the Monte Carlo estimate of the provinces' food security ranks and was estimated using the median and mean ranks derived from the UA. This was regarded as the revealed true population value of the provinces' rankings on the aspect of food security.

The *Factor Mapping Stage Analysis* was concerned with knowing what combinations of levels of the input factors are responsible for the realization of the desirable model outputs. This goal was accomplished by first categorizing the shifts ($\bar{R}_s$) from the Monte Carlo *Median Rank* and then identifying the levels of the input factors responsible for producing the different or desired levels of shifts while being guided by the results of the sensitivity analysis. For factors with high first-order effects or main effects, the effects of their levels taken singularly were investigated. For factors with high interaction effects, the effects of their levels in combination with the levels of other interacting factors were observed.

Syntax program was done in R 3.2.2 and was used to implement the UA and SA in this study.

4. Results and Discussion

4.1. Calibration Stage Analysis

Graph of the computed food security indices based on the conventional approach and indices based on the ratio with mean normalization technique (combined with the same set of the other input factors found in the conventional approach except for the normalization method) for the top performing provinces is shown in Figure 2 below. The provinces were arranged in descending order of the conventional index. In this result, it is clear to see the remarkable difference in the food security indices and corresponding ranks of provinces produced using RWM as compared to those generated using the conventional approach.

For the re-scale of the food security indices or scores derived via RWM normalization, the minimum and maximum RWM-score values used were 0.442 and 1.752 respectively. On the other hand, for the re-scale of the index scores derived from z-score normalization, the minimum and maximum z-score values used were -2.201 and 2.753 respectively.

For the uncertainty analysis (UA), indices and ranks were computed across the 7000 samples (actual final sample was 5406 due to conflicting input factors), which are actually 7000 combinations of the different levels of the input factors. For a very few provinces (Batangas, Bukidnon, Isabela, Palawan, Tarlac, and Tawi-tawi, some of the computed indices exceeded 100, while for Cavite alone, one or two indices were calculated to be slightly negative.

These unusually extreme indices were investigated although we did not show here the full details to save space. The RWM normalization method was identified to be particularly responsible for the indices' excessive magnitude and high variability in these provinces that can exceed the normal range. Upon inspection, it was noticeable that these provinces have very high production, either singly or combinations, of palay, corn, banana, cassava, mongo, chicken, duck, eggs, and fish.

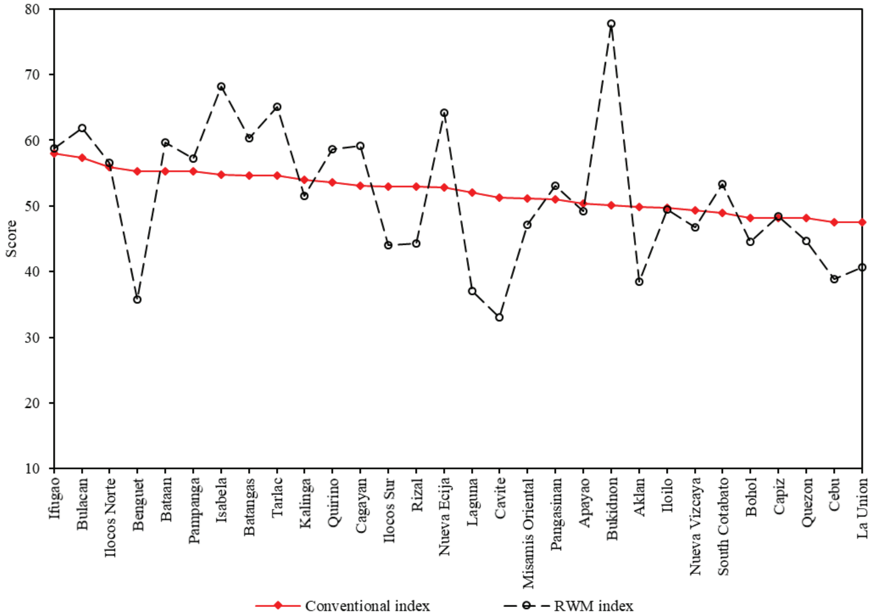


Figure 2. Computed food security indices using the conventional approach and ratio with mean (RWM) normalization technique of the top provinces

For example, Tawi-tawi has a RWM score for fish production of 13.74, which means that Tawi-tawi produces fish that is approximately 14 times what an average or regular province in the country does. In this context, the very high food security index computed for Tawi-tawi via RWM can be viewed not just exhibiting a compensatory mechanism, but more so exhibiting a "be-all" mechanism as the high fish production of Tawi-tawi can make it the most food secure province regardless of the province's failures on some other things.

Even with a low sub-indicator weight, a high RWM sub-indicator score could bring a significant impact in the result. This was illustrated in the case of Benguet and Bukidnon where based on the conventional index Benguet was outperforming Bukidnon, but under the RWM normalization, this scenario is reversed.

This reversal was identified to have stemmed from the calculation of the *stability* dimension scores using RWM for the two provinces where the area harvested with corn (in hectares) is a component. This area is 38 ha. for Benguet, 190,752 ha. for Bukidnon, and the average across all the provinces is 31,437.5 ha. For this sub-indicator, Bukidnon gained a 6.07 RWM score, and when multiplied by the allotted weight, Bukidnon practically yielded 0.18 which is extremely in excess of the approximately 3% (0.03) allotted weight for this sub-indicator under the *stability* dimension. This case cannot happen in Min-Max. In Min-Max even if a province has a 100 Min-Max score (this is likewise equivalent to 1.0) for this sub-indicator, the highest score that that province can gain contributory to its final *stability* score is only 3, or strictly 3%, no more. This 0.18 sub-indicator score of Bukidnon is approximately 15% of Bukidnon's total un-rescaled *stability* dimension score which is 1.15.

The RWM scores for Cavite are low and many of its high-scored sub-indicators are in the *stability* dimension. Thus, it is fairly understandable that the very few negative indices computed under this province occurred under the combination of the methods RWM normalization, AHP weighting scheme, geometric aggregation (which is relatively a penalizing aggregation system), and dataset where the *stability* dimension was removed. It can be argued that the limits used for re-scaling the RWM indices be changed to be able to accommodate these unusually extreme values, but this will not have any bearing in the provinces' ranks, and increasing the score ceiling will only pull the indices of the other provinces farther down.

The study finds the results of the RWM normalization interesting amidst trivial. We opined that the RWM method offers a normalization procedure that allows a single aspect of a problem or model (in this case, a single or a group of sub-indicators, or a single dimension of *food security*) to dominate or determine the outcome of the whole model (i.e. a "be-all" or "answers-all" mechanism), regardless of failures in some aspects. At this stage of analysis, we allowed this kind of mechanism and supposed that such a condition exists in the real world. In the subsequent analyses, the RWM normalization was dropped, avoiding trivial results as much as possible as one reason.

The shifts in the position of the entire system of provinces ($\bar{R}_s$) away from the conventional index (I_c), which were derived by varying the levels of the input factors (UA), range from 0.00 or no change, to as high as 21.04 positions or ranks with a mean of 10.98. This means that, on the average, a province's rank on food security according to the conventional approach may improve or slip down by about 11 ranks when one recalculates the indices using another approach, say by changing the method of normalization or interviewing a different expert. In the succeeding analyses under the *Post-Calibration Stage Analysis* where the RWM normalization was excluded, the shifts ($\bar{R}_S$) have notably decreased.

In the sensitivity analyses of $\bar{R}_s$ with respect to the conventional index (I_C), the estimated mean ($\hat{f}_0$) used for the adjustment was 10.98 coming from the earlier UA which utilized a sample of 7,000. The results of the SA are shown in Table 3 below (for $n = 9{,}000$), and in the Appendix in Table B. The SA result for $n = 9{,}000$ was regarded as the most appropriate choice of results from among the varied SA's.

Large negative sensitivity indices were observed especially when using the formula by Nossent and Bauwens (2012) without adjustment and this was unexpected. Negative sensitivity measures were likewise documented in other studies (see Saltelli et al., 2004). Small computed negative sensitivity indices happen in cases when the analytical values are close to zero or the input factors are not influential. Negative Sobol' sensitivity indices were attributed to sampling error in the estimation and can be overcome by increasing the sample size. The variance analytically is always positive, but in the case of the estimation of Sobol' variances which is done via numerical approximation and via the use of two independent and different sample matrices, the computation of negative variances can be possible.

Some first-order sensitivity indices were observed to be greater than their counterpart total-effect sensitivity indices. These occur mostly when using the approach by Nossent and Bauwens (2012), more so in the unadjusted case. These results can be regarded as the same biased and unrealistic results alluded to by Nossent and Bauwens (2012), Sobol' (2001), and Mokhtari and Frey (2005) that can happen when f_0 is large.

The practical similarity in the conclusions that can be derived from the results between the approach by Jansen et al. (1994) and by Nossent and Bauwens (2012) with adjustment as the sample size increases was noted. Likewise, the results with Jansen et al. (1994) were within expectations (i.e. no large negative sensitivity measures and first-order sensitivities were smaller than the total-effect sensitivities). For these reasons, the approach by Jansen et al. (1994) was recommended for the succeeding analyses.

Table 3. Sobol' sensitivity measures for the average shift in provinces' ranks ($\bar{R}_s$) with respect to the Conventional Index (I_C) estimated using three formulas, n=9000

Input factor	*Jansen et al. (1994)*			*Nossent and Bauwens (2012) without adjustment*			*Nossent and Bauwens (2012) with adjustment*		
	S_i	S_{Ti}	S_{Ti} - S_i	S_i	S_{Ti}	S_{Ti} - S_i	S_i	S_{Ti}	S_{Ti} - S_i
Dataset	0.185	0.208	0.023	0.171	0.226	0.056	0.176	0.214	0.038
Normalization	0.606	0.631	0.025	1.032	0.813	-0.219	0.595	0.590	-0.005
Weighting scheme	0.041	0.098	0.057	0.043	0.111	0.067	0.032	0.100	0.068
Aggregation	0.053	0.056	0.003	-0.199	-0.320	-0.122	0.085	0.058	-0.027
Availability weight	0.022	0.052	0.030	0.017	0.078	0.061	0.017	0.056	0.039
Accessibility weight	-0.006	0.011	0.017	0.000	0.023	0.023	-0.003	0.012	0.015
Utilization weight	0.013	0.037	0.024	0.021	0.050	0.029	0.015	0.039	0.024
Stability weight	-0.006	0.008	0.014	0.003	0.023	0.020	-0.001	0.006	0.007
Sum indices	0.908	1.101		1.089	1.003		0.917	1.075	

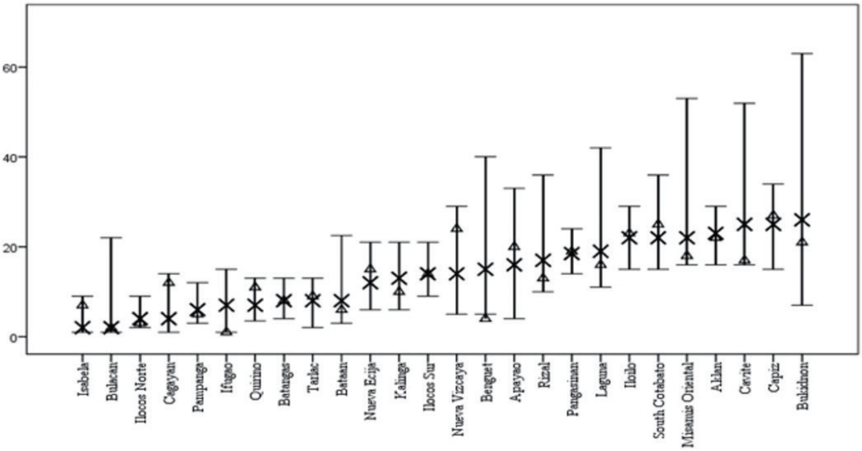
The result of the approach by Jansen et al. (1994) was further reinforced with the use of a large sample size. At an initial sample n = 9,000, the effective sample realized was 6,946, which is within the sample size demonstrated by Nossent and Bauwens (2012) that convergence in the Sobol' sensitivity measures starts to happen. Further at n = 9,000, the practical conclusion or results that can be derived from that based on Jansen et al. (1994) was just similar to that with using Nossent and Bauwens (2012) with adjustment. While better results can be gained based on using n = 11,000 samples, results based on the sample size n = 9,000 just gives practically the same conclusion at a lesser computational cost. For these reasons, a sample size of n = 9,000 was recommended to be used for the follow-up analyses.

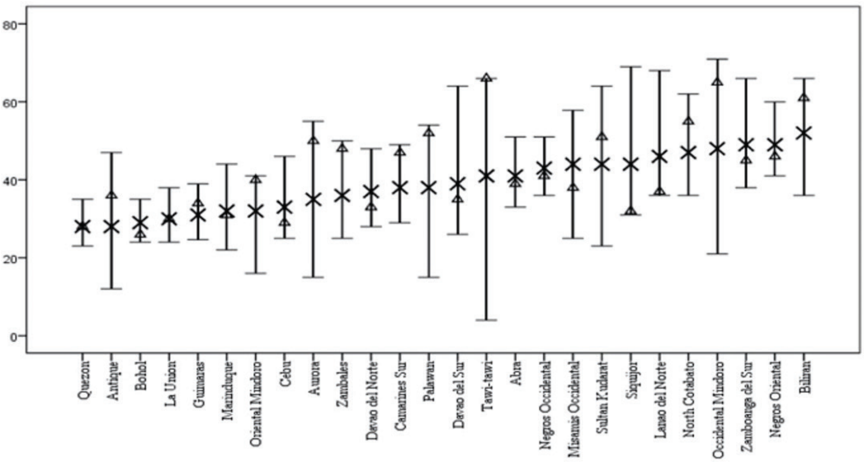
The discussion of many of the results of the SA for n = 9,000 (see Table 3) was restrained. The variance of the model output $\bar{R}_s$ is 22.37, which the SA partitioned among the input factors. About 60% of the variation in $\bar{R}_s$ was due to the singular effect of the normalization method used dominating all other factors. This sensitivity measure for the normalization method is only a quantification and confirmation of what we have already observed earlier: the RWM normalization causes too much fluctuation in the outcome of the food security indices. This finding is not true in some previous studies encountered (see OECD (2008) and Saisana et al. (2005)) as the normalization method usually was non-influential. For the succeeding analyses, we therefore recommended to exclude the RWM method in the list of the normalization procedures to avoid radical results as much as possible.

4.2. Post-Calibration Stage Analysis

Based on the recommendations from the previous stage of analysis, UA and SA were again undertaken, but this time excluding the RWM normalization method, fixing the sample size at 9,000, and adopting the method by Jansen et al. (1994) in the computation of the Sobol' indices. A new set of the matrices of the sample input triggers (M_1 and M_2) were generated, each composed of n = 9,000. The old matrices were no longer applicable at this stage as the RWM normalization method was now dropped down from the analysis.

The corresponding ranks of the provinces based on the computed conventional index were plotted together with the per province median, 5th, and 95th percentiles of the ranks derived from the variation of the input factors (UA). These results are shown in Figure 3. In the graphs, the provinces were arranged in ascending order of the median rank.





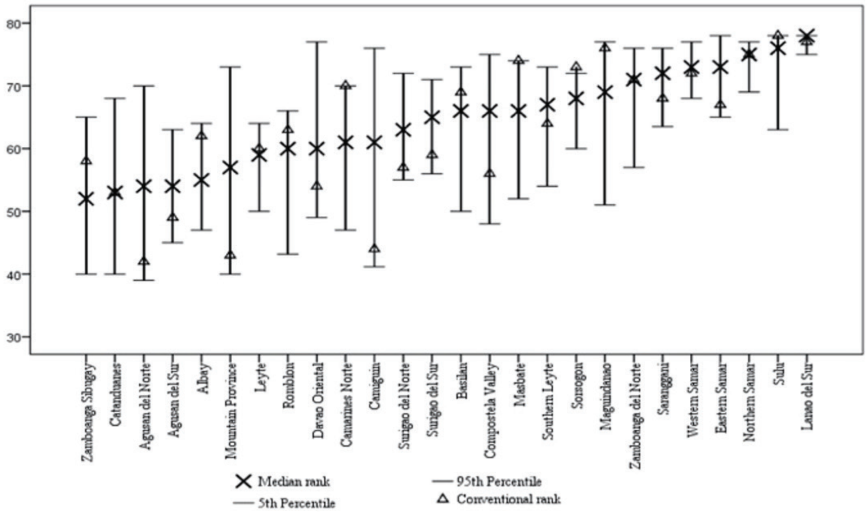


Figure 3. Rank of the provinces based on the food security indices computed from the variation of the input factors. Ratio with mean normalization method was not included. Initial sample= 9000, actual sample = 5988 due to conflicting input factors

Generally, the graphs show the non-concordance of the ranks of the conventional index and the median ranks, the same as when previously the RWM normalization method was not yet dropped. If the median values were to be treated as the true values and if the uncertainties acknowledged in this study were an accurate reflection of the existing uncertainties involved in the construction of the food security index, then the conventional approach in computing for the food security index provide a bias picture of the measurement of food security of the provinces. In bootstrapping or resampling problems, the median value is normally treated as the revealed true value of the unknown population parameter. The implication of this is that the conventional index (I_c), for example, can rate a province to be the most food secure whereas the reality is that that province is not the most food secure, save only that the combination of methods enveloped in I_c was just "biased" or favorable to that particular or group of provinces.

The shift in the provinces' ranks ($\bar{R}_s$) relative to the conventional index were again calculated per combination of the different levels of input factors generated. These shifts ($\bar{R}_s$) have now reduced compared to when previously the RWM normalization was not yet dropped. These range from 0.00 (no change) to 17.21, with a mean of 7.74.

The differences in the food security ranks of two provinces, D_R (i.e. rank of Benguet minus rank of Cagayan) were calculated using the variation in the input factors for an initial sample of n = 9,000. The same sample input matrices used in this stage of the analysis were used in the UA and SA of D_R. D_R ranges from -44 to 57, with an average of 12.69 and a standard deviation of 13.3. There is no particular basis with regards to the choice of two specific provinces to compare under SA, although we observed that the food security rankings of Benguet and Cagayan under the conventional approach and the *Median Rank* were not consistent and captured our interest. Any choice of provinces will do.

The overall Monte Carlo estimate of the rank of the provinces on food security (*Median Rank*) was determined based on the median and mean of the ranks per province derived from the uncertainty analysis based on 9,000 samples. The medians of the ranks per province can be the same. For example, for the provinces of Isabela and Bulacan, the median of their ranks based on UA were both 2. However, the mean of their ranks were: 3.2 for Isabela and 7.2 for Bulacan. Based on these median and mean of the ranks, Isabela was re-ranked first (rank 1) over Bulacan (rank 2). Whenever some provinces have similar median rank in the UA, their mean ranks were used to further determine their exact order or position and break the tie. This new ranking system that allows only one province in a rank position was referred hereto as the *Median Rank*. The *Median Rank* was treated or assumed as the revealed true population value of the provinces' rankings on the aspect of food security. The *Median Rank* for all the provinces is presented in Table 4.

Table 4. Rank of provinces on food security based on the median and mean ranks (Median Rank) derived from the variation of the input factors. Ratio with mean normalization method was excluded. Initial sample=9000, actual sample=5988 due to conflicting input factors

Median rank	*Province*	*Median rank*	*Province*	*Median rank*	*Province*	*Median rank*	*Province*
1	Isabela	21	South Cotabato	41	Tawi-tawi	61	Davao Oriental
2	Bulacan	22	Misamis Oriental	42	Abra	62	Camarines Norte
3	Ilocos Norte	23	Aklan	43	Negros Occidental	63	Camiguin
4	Cagayan	24	Capiz	44	Misamis Occidental	64	Surigao del Norte
5	Pampanga	25	Cavite	45	Sultan Kudarat	65	Surigao del Sur
6	Quirino	26	Bukidnon	46	Siquijor	66	Basilan
7	Ifugao	27	Quezon	47	Lanao del Norte	67	Masbate
8	Tarlac	28	Antique	48	North Cotabato	68	Compostela Valley
9	Batangas	29	Bohol	49	Occidental Mindoro	69	Southern Leyte
10	Bataan	30	La Union	50	Negros Oriental	70	Sorsogon
11	Nueva Ecija	31	Guimaras	51	Zamboanga del Sur	71	Maguindanao
12	Kalinga	32	Oriental Mindoro	52	Biliran	72	Zamboanga del Norte

Median rank	Province	Median rank	Province	Median rank	Province	Median rank	Province
13	Ilocos Sur	33	Marinduque	53	Zamboanga Sibugay	73	Saranggani
14	Nueva Vizcaya	34	Cebu	54	Catanduanes	74	Eastern Samar
15	Benguet	35	Aurora	55	Agusan del Sur	75	Western Samar
16	Apayao	36	Zambales	56	Agusan del Norte	76	Northern Samar
17	Rizal	37	Davao del Norte	57	Albay	77	Sulu
18	Pangasinan	38	Palawan	58	Mountain Province	78	Lanao del Sur
19	Laguna	39	Camarines Sur	59	Leyte		
20	Iloilo	40	Davao del Sur	60	Romblon		

$\bar{R}_s$ with respect to the *Median Rank* were computed by varying the input factors. The distribution of this $\bar{R}_s$ is shown in Figure 4.

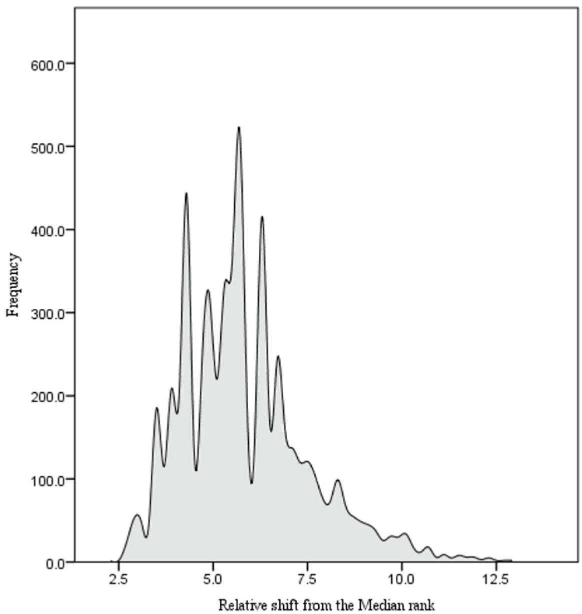


Figure 4. Shift in the provinces' ranks relative to the median rank ($\bar{R}_s$) due to the variation in the input factors. Ratio with mean normalization method was not included. Initial samples = 9000, actual samples = 5988 due to conflicting input factors

The results of the sensitivity analysis on the three model outputs are shown in Table 5. Notwithstanding the numerical discrepancies in the results of the SA's, there are generally major similarities or consistency that can be deduced across the model outputs.

Since there is no clear rule on what magnitude of the sensitivity index makes an input factor "influential", a rough guide is to regard a factor influential if its sensitivity measure equaled or exceeded the value $1/k$ (Saisana et al, 2005). In the present study, a factor can be considered influential if it is able to explain $1/8 \approx 13\%$ of the total variation in the model output.

Table 5. Sobol' sensitivity measures for three model outputs estimated using the formula by Jansen et al. (1994). Ratio with mean normalization was not included.

Input factor	*$\bar{R}_s$ with respect to the Conservative Index (I_C)*			*Difference in rank (D_R) between Benguet and Cagayan*			*$\bar{R}_s$ with respect to the Median Rank*		
	S_i	S_{Ti}	$S_{Ti} - S_i$	S_i	S_{Ti}	$S_{Ti} - S_i$	S_i	S_{Ti}	$S_{Ti} - S_i$
Dataset	0.43	0.47	0.04	0.13	0.24	0.11	0.27	0.40	0.13
Normalization	0.00	0.00	0.00	0.00	0.00	0.00	-0.02	0.02	0.04
Weighting scheme	0.16	0.27	0.11	0.06	0.24	0.19	0.17	0.44	0.27
Aggregation	0.14	0.19	0.05	0.35	0.39	0.04	0.16	0.27	0.11
Availability weight	0.05	0.09	0.05	0.16	0.29	0.12	0.01	0.25	0.24
Accessibility weight	0.00	0.02	0.01	0.00	0.03	0.03	0.04	0.08	0.04
Utilization weight	0.06	0.12	0.06	0.03	0.09	0.06	0.05	0.20	0.15
Stability weight	0.00	0.02	0.02	0.00	0.05	0.05	0.05	0.07	0.03
Sum indices	0.84	1.18		0.74	1.34		0.74	1.73	

For the results of SA on $\bar{R}_s$ with respect to the Conventional Index (I_C), under a factor prioritization setting, in order of priority, the choice of sub-indicators or dimensions, weighting scheme, and aggregation system should be given priority in future studies. Investments or experiments to know more their nature and learn what their true values are should be given more importance than the other factors, as fixing them singularly to their "true values" would greatly reduce the uncertainty in the outcome of the food security indices.

Under a factor fixing setting, the normalization method, weight for *accessibility*, and weight for *stability* can be pegged to fixed values as their variation will not cause any or significant effect to the output. These factors are uninfluential as reflected by their low total-effect sensitivity indices. The result for the normalization method this time after the RWM normalization was dropped was very different from that previously when normalization dominates the determination of the output model.

The weighting scheme influence the output in its singular capacity, but its effect to the output in cooperation with the other inputs (interaction effect) was given more emphasis in this study. The considerable difference between their total-effect and first-order effect sensitivity indices for the weight for *availability* and weight for *utilization* likewise reveal that these two input factors interact with the other input factors to cause variations in the output. In fact, it is the interaction between the weighting scheme and weights for *availability* and *utilization* that appears prominent when the results of SA for the other model outputs are considered.

The general message that the results of the sensitivity analysis on the average shift in provinces' ranks ($\bar{R}_s$) with respect to the *Median Rank* convey is nothing new when compared to what the results of the SA's on the other model outputs communicate. It likewise reinforces that the weighting scheme, weight for *availability*, and weight for *utilization*, and slightly, the choice of dataset influence the output through their cooperative or interaction effects.

4.3. Factor Mapping Stage Analysis

The histogram of $\bar{R}_s$ relative to the *Median Rank* presented us a natural grouping of the shifts into three categories (see Figure 4). The two deep "valleys" that can be observed in the distribution (observed at $\bar{R}_s = 4.5$ and $\bar{R}_s = 6.0$) were considered as the cut-off points for the categorization of the shifts: *Mild shift* ($0 \leq \bar{R}_s \leq 4.5$); *Moderate shift* ($4.5 < \bar{R}_s \leq 6.0$); *Large shift* ($6.0 < \bar{R}_s < \infty$).

Large shifts in the food security rankings from the "ideal or true ranking", and therefore the undesirable outcomes, are associated with the PCA/Correlation based dataset with the *stability* dimension removed (see Figure 5, A). Mild shifts in the food security rankings from the "ideal or true ranking", and therefore the desirable outcomes, are associated with the dataset where the regional level sub-indicators were removed. Mild shifts are not clearly associated with both the purely additive and purely geometric aggregation systems (see Figure 5, B).

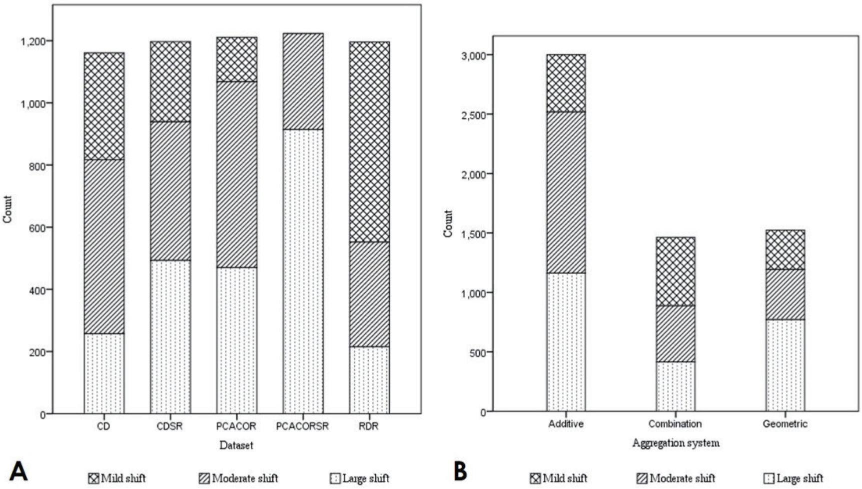


Figure 5. Average shift of provinces' ranks ($\bar{R}_s$) from the median rank due to variation in the input factors (A) by Dataset. (CD-Complete dataset; CDSR-Complete dataset with Stability removed; PCACOR-PCA/Correlation based dataset; PCACORSR-PCA/ Correlation based dataset with Stability removed; RDR-Regional level data removed); (B) by Aggregation system. Ratio with mean normalization method was not included.

The main effect of the weighting scheme was also explored although we are more inclined to believe here that the weighting scheme works more by influencing the model outputs through interaction effects. Table 6 shows the distribution of by weighting scheme. There is a good reason to believe that the huge fluctuations in the provinces' food security rankings are associated with using the Analytic Hierarchy Process as a method for eliciting the experts' opinions on the weights of the food security dimensions.

Table 6. Average shift of provinces' ranks ($\bar{R}_s$) from the median rank due to variation in the input factors by weighting scheme. Ratio with mean normalization method was not included. Figures in parentheses are percentages.

Frequency	*Equal Weighting*		*Analytic Hierarchy Process*		*Budget Allocation Procedure*	
Mild shift	610	(30.0)	266	(13.5)	513	(25.9)
Moderate shift	1096	(53.9)	473	(23.9)	681	(34.4)
Large shift	328	(16.1)	1236	(62.6)	785	(39.7)
Total	*2034*	*(100.0)*	*1975*	*(100.0)*	*1979*	*(100.0)*

Plots of the weights of *availability* and *utilization* by weighting scheme and category of shifts are shown in Figure 6, A and B respectively. Lower weights for *availability,* i.e. ≤ 30%, under the Budget Allocation Procedure is associated with moderate shifts (this comprise about 75% of the cases in this specific cross-classification). A slightly higher weight for *availability* (median=30%) under the BAP weighting scheme is likewise associated with mild shifts in the provincial ranks.

Low weights for *utilization* (median=10%) under AHP caused large deviation of the indices from the ideal robust ranking, while higher weights for *utilization* (median=20%) under BAP caused moderate deviations of the provinces' food security status.

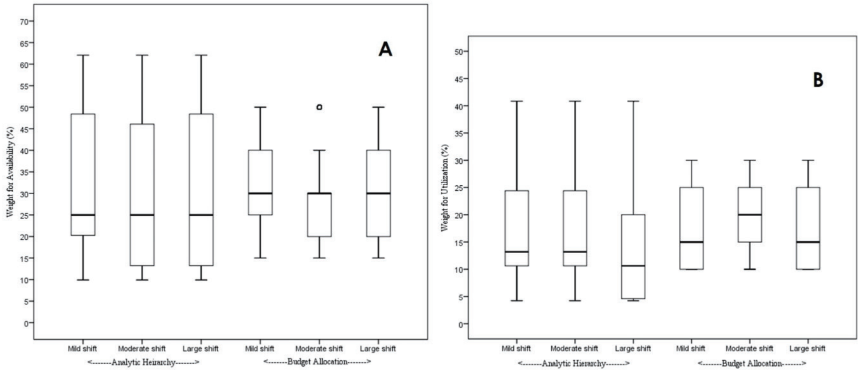


Figure 6. Average shift of provinces' ranks ($\bar{R}_s$) from the median rank due to variation in the input factors (A) by Weighting scheme and Weight for availability; (B) by Weighting scheme and Weight for utilization. Ratio with mean normalization method was not included.

5. Conclusion and Recommendation

Composite indices are useful in benchmarking the performance of organizational units and can communicate significant policy messages. It is important therefore to assess their quality or reliability especially when many uncertainties and subjective judgments are involved in their construction. This study offered a groundwork for exploring the construction of composite provincial level food security indices in the Philippines that lends useful insights which can be relevant to future index developers, policy makers and stakeholders. The study acknowledged the existence of various uncertainties, possibilities, or different options (input factors) that can be used in the construction of the food security index such as: (1) inclusion/exclusion of sub-indicators; (2) normalization method; (3) weighting scheme; (4) aggregation system; and varying weights of the food security dimensions – (5) weight for *availability*; (6) weight for *accessibility*; (7) weight for *utilization*; and (8) weight for *stability*.

The application of uncertainty and sensitivity analysis techniques was demonstrated in the assessment of the robustness, and biasedness of the conventional composite food security index. The application of these techniques in particular was done, with more emphasis, within the context of factor prioritization, factor fixing, and factor mapping settings.

The model outputs – food security indices, corresponding provincial ranks based on the food security indices, average shift in the food security position of the entire system of provinces ($\bar{R}_s$) in relation to the conventional index and the *median rank*, and the difference in two provinces' ranks based on food security status, were the subjects of investigations of the study.

The study likewise had been exploratory and investigative of the various ways or uncertainties involved in the construction and analysis of composite food security indices, by utilizing the *ratio with mean* transformation as a method of normalization, by proposing a method of aggregation that is a *combination of additive and geometric aggregation* systems, and by applying and comparing different formulas, at varying sample sizes, of computing the Sobol' sensitivity indices.

The study has shown the following:

1. If the uncertainties considered in this study were an accurate representation of the true uncertainties involved in the construction of the food security index, and if the calculated median values under UA were assumed to be the real food security status of the provinces, then the food security index being calculated using the conventional approach (I_c), provides a bias picture of the food security status of the provinces given the many reversals noted in the ranks, and the computed provincial food security indices;
2. The main effects of choice of dataset or sub-indicators, aggregation system, and weighting scheme largely influence the outcome of the model. Therefore, their study or consultations about them in the future should be given high priority as understanding them and knowing their real values will lead to the greatest reduction in the output's uncertainty;
3. The method of normalization, weight for *accessibility*, and weight for *stability* are not influential, and therefore can be fixed to their nominal or conventional values without a significant loss of information effected on the model output;
4. The weighting scheme, weight for *availability*, and weight for *utilization*, and slightly, the choice of dataset, influence the output through their cooperative or interaction effects;
5. Combinations of levels of input factors were identified to be responsible for producing the "desirable" (mild shifts in the provinces' food security status from the Monte Carlo estimate of food security rank) and "undesirable" model outputs;
6. The formulas for the calculation of the Sobol' sensitivity indices by Jansen et al. (1994) and Nossent and Bauwens (2012) with adjustment of the model output provide practically the same conclusions on the results of the sensitivity analyses particularly as the sample size increases. The formulas by Jansen et al. (1994) yielded results that are not ambiguous, i.e. no large negative sensitivity measures and the first-order sensitivity indices' magnitudes were less than their corresponding total-effect sensitivity indices, however at a higher computational cost; and
7. The ratio with mean (RWM) normalization method caused much fluctuation in the model outputs compared to the other normalization methods, and that it exhibited a highly compensatory mechanism in the provinces' food security status, which the study dubbed as an "answers all" or "be all" mechanism, that warrants further exploration and serious caution in the method's application.

The weights of the sub-indicators within the dimensions were assigned subjectively and could have been solicited also from experts, similar to the case of the food security dimension weights. Future studies can therefore consider the uncertainty in the sub-indicator weights as another input factor in the food security index model in addition to the other uncertainty input factors already included in this study.

Finally, the findings in this study can be taken into consideration in the future development of a uniform composite provincial level food security index in the country that relies on the FAO's definition of food security. Uncertainty and sensitivity analysis should be incorporated in building the index to increase and guarantee the index' transparency and reliability.

References

AHLAWAT, S. and D. KAUR. 2013, Spatial Patterns of Food Security in Haryana: A District- level Analyis, *Journal of Tropical Resources and Sustainable Science*, 1 (2): 1-11.

ALEMU, Z.G., 2015, Developing a Food Security Map for South Africa, Working Paper Series No. 220, African Development Bank, Tunis, Tunisia.

BILAL, N., 2014, Implementation of Sobol's Method of Global Sensitivity Analysis to a Compressor Simulation Model, 22nd International Compressor Engineering Conference, July 14-17, 2014, Paper 2385, Purdue University, Available at http://docs.lib.purdue.edu/icec/2385

CARGILL, INC., 2014, Food Security: The Challenge, Available at https://www.cargill.com/wcm/groups/public/@ccom/documents/document/na3059573.pdf. Accessed: July 22, 2016

CHAN, K., A. Saltelli, and S. Tarantola, 1997, Sensitivity analysis of model output: variance-based methods make the difference, In: Proceedings of the 1997 Winter Simulation Conference Eds. S. Andradottir, K.J. Healy, D.H. Withers, and B.L. Nelson

EIU (The Economist Intelligence Unit Ltd.), 2014, Global food security index 2014: An annual measure of the state of global food security. DuPont.

EIU (The Economist Intelligence Unit Ltd., 2015, Global Food Security Index, Available at http://foodsecurityindex.eiu.com/, Accessed on April 2015

EPP, S., 2010, Provincial Approaches to Food Security, A Scan of Food Security Related Policies in Canada. Manitoba Food Charter

FAO, UNICEF, and WFP (Food and Agriculture Organization, United Nations International Children's Emergency Fund, and World Food Programme), 2010, Nutrition Security and Food Security in Seven Districts in NTT Province, Indonesia: Status, Causes and Recommendations for Response. Final Report. Food and Agriculture Organization of the United Nations, Rome, Italy

FNRI-DOST (Food and Nutrition Research Institute-Department of Science and Technology), 2015, Philippine Nutrition Facts and Figures 2013: Food Security Survey. FNRI Bldg., DOST Compound, Bicutan, Taguig City, Metro Manila, Philippines.

GLEN, G. and K. ISAACS, 2012, Estimating Sobol sensitivity indices using correlations, *Environmental Modelling & Software,* 37 (2012) 157-166

IQBAL, M. and R. AMJAD, 2010, Food security in South Asia: strategies and programmes for regional collaboration, *Munich Personal RePEc Archive* (MPRA), Paper No. 38077

JANSEN, M.J.W., W.A.H. Rossing, and R.A. Daamen, 1994, Monte Carlo estimation of uncertainty contributions from several independent multivariate sources, In: J. Grasman, G. Van Straten (Eds.), Predictability and Nonlinear Modelling in Natural Sciences and Economics, Kluwer Academic Publishers, Dordrecht, pp. 334–343.

MAGOMBEYI, M.S., A.E. Taigbenu and J. Barron, 2013, Rural poverty and Food insecurity mapping at district level for improved agricultural water management in the Limpopo River Basin, Colombo, Sri Lanka: CGIAR Challenge Program on Water and Food (CPWF). 54pp. (CPWF Research for Development (R4D) Series 6)

MAXWELL, D., J. Coates, and B. Vaitla, 2013, How Do Different Indicators of Household Food Security Compare? Empirical Evidence from Tigray, Feinstein International Center, Tufts University: Medford, USA.

MOEA Framework, _______, Discussions about the sensitivity indices in the analysis feature of the MOEA Framework located in the Comments Section. Available at https://github.com/MOEAFramework/MOEAFramework/issues/43, Accessed on October 16, 2016

MOKHTARI, A. AND H.C. FREY, 2005, Review and Recommendation of Methods for Sensitivity and Uncertainty Analysis for the Stochastic Human Exposure and Dose Simulation (SHEDS) Models, In: Evaluation and Recommendation of Methodology for Conducting Sensitivity Analysis in Probabilistic Models Vol. 2, Alion Science and Technology, Durham, North Carolina

NDHLEVEL, S., L. Musemwa and L. Zhou, 2012, Household food security in a coastal rural community of South Africa: Status, causes and coping strategies, *Journal of Agricultural Biotechnology and Sustainable Development*, 4 (5): 68 –74

NOSSENT, J. and W. BAUWENS, 2012, Optimising the convergence of a Sobol' sensitivity analysis for an environmental model: application of an appropriate estimate for the square of the expectation value and the total variance. In: 2012 International Congress on Environmental Modelling and Software Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany. R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.). International Environmental Modelling and Software Society (iEMSs), Germany

OECD (Organization for Economic Cooperation and Development), 2008, Handbook on Constructing Composite Indicators, Methodology and User Guide. OECD and JRC, Ispra, Italy

PSA (Philippine Statistics Authority), 2012, Philippine Food Security Information System (PhilFSIS), Available at: http://philfsis.psa.gov.ph/, April 2015

SAATY, T.L, 1990, How to make a decision: The Analytic Hierarchy Process, *European Journal of Operational Research*, 48 (1): 9-26

SAISANA, M., A. Saltelli, and S. Tarantola, 2005, Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *J.R. Statist. Soc*. A, 168 (Part 2): 307-323

SAJJAD, H. and I. NASREEN, 2014, Food Security in Rural Areas of Vaishali District, India: A Household Level Analysis, *European Journal of Sustainable Development*, 3 (3): 235-250

SALTELLI, A., 2002, Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145: 280-297

SALTELLI, A., S. Tarantola, F. Campolongo, and M. Ratto, 2004, Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models. John Wiley & Sons Ltd. England

SMITH, L.C. and A. SUBANDORO, 2007, Measuring Food Security Using Household Expenditure Surveys. Food Security in Practice technical guide series. Washington, D.C.: International Food Policy Research Institute (IFPRI).

SOBOL', I.M., 2001, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Mathematics and Computers in Simulation,* 55: 271–280

SUHARYANTO, J.H. Mulyo, D.H. Darwanto and S. Widodo, 2014, Determinants of Food Security Among Rice Farming Households in the Province of Bali: An Ordered Logistic Model. *Journal of Economics and Sustainable Development*, 5 (8): 35-42

UNIVERSITY OF SIENA., _____, Note on the Analytic Hierarchy Process. Department of Engineering and Mathematics, University of Siena, Italy 1240, Available at: www.dii.unisi.it/~mocenni/Note_AHP.pdf, Accessed on November 14, 2015

WIESMANN, D., L. Bassett, T. Benson and J. Hoddinott, 2009, Validation of the World Food Programme's Food Consumption Score and Alternative Indicators of Household Food Security, *IFPRI Discussion Paper* No. 00870, Poverty, Health and Nutrition Division, International Food Policy Research Institute (IFPRI)

WINEMAN, A., 2014, Multidimensional Household Food Security Measurement in Rural Zambia, Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2014 AAEA Annual Meeting, Minneapolis, MN, July 27-29, 2014, Department of Agricultural, Food, and Resource Economics, Michigan State University

WFP and GSS-MFA (World Food Programme and Ghana Statistical Service, Ministry of Food and Agriculture), 2012, Comprehensive Food Security & Vulnerability Analysis. Focus on Northern Ghana, United Nations World Food Programme Headquarters, Rome, Italy

ZHENG, Y. and A. RUNDELL, 2006, Comparative Study of Parameter Sensitivity Analyses of the TCR-activated ERKmapk Signalling Pathway, *Proceedings of Systems Biology*, Vol. 153, no. 4, p. 201-211

Appendix

Table A-1. Sub-indicators under the Availability and Accessibility dimensions

Dimension / Sub-indicator	***Definition***	***Dimension / Sub-indicator***	***Definition***
Availability		***Accessibility***	
PALAY_DIV	Palay production (per capita)	INCOME_2012	Average annual family income 2012
CORN_DIV	Corn production (per capita)	POV_INC_INV	Poverty incidence among population (inverted)
BANANA_DIV	Banana production (per capita)	SUB_INC_INV	Subsistence incidence among population (inverted)
SPOTATO_DIV	Sweet potato production (per capita)	EXPENDITURE	Provincial expenditure
UBI_DIV	Ubi production (per capita)	EMPLOYMENT_RATE	Employment rate
CASSAVA_DIV	Cassava production (per capita)	CCT	Households with children and pregnant women that are CCT beneficiaries
GABI_DIV	Gabi production (per capita)	PRICE_RICE_INV	Retail price of regular milled rice (inverted)
SBEANS_DIV	String beans production (per capita)	PRICE_PORK_INV	Retail price of pork lean meat (inverted)
SQUASH_DIV	Squash production (per capita)	PRICE_EGG_INV	Retail price of chicken egg (inverted)
EGGPLANT_DIV	Eggplant production (per capita)	PRICE_CHICKEN_INV	Retail price of dressed chicken (inverted)

Dimension / Sub-indicator Availability	Definition	Dimension / Sub-indicator Accessibility	Definition
KANGKONG_DIV	Kangkong production (per capita)	PRICE_GG_INV	Retail price of galunggong or roundscad (inverted)
MONGO_DIV	Mongo production (per capita)	PRICE_BANGUS_INV	Retail price of bangus (inverted)
PECHAY_DIV	Pechay production (per capita)	PRICE_TILAPIA_INV	Retail price of tilapia (inverted)
CARABAO_DIV	Carabao production (per capita)	PRICE_TULINGAN_INV	Retail price of tulingan (inverted)
CATTLE_DIV	Cattle production (per capita)	PRICE_SPOTATO_INV	Retail price of sweet potato (inverted)
HOG_DIV	Hog production (per capita)	PRICE_MONGO_INV	Retail price of mongo (inverted)
GOAT_DIV	Goat production (per capita)	PRICE_SBEANS_INV	Retail price of string beans (inverted)
CHICKEN_DIV	Chicken production (per capita)	PRICE_EGGPLANT_INV	Retail price of eggplant (inverted)
DUCK_DIV	Duck production (per capita)	PRICE_SQUASH_INV	Retail price of squash (inverted)
EGG_CHICKEN_DIV	Chicken egg production (per capita)	FMR_LENGTH	Completed farm to market roads, total length
EGG_DUCK_DIV	Duck egg production (per capita)		
FISH_DIV	Fisheries production (per capita)		
UNWANT_FOOD_INV	Households that sometimes ate food they don't want (inverted)		

Table A-2. Sub-indicators under the Utilization and Stability dimensions

Dimension / Sub-indicator	***Definition***	***Dimension / Sub-indicator***	***Definition***
Utilization		***Stability***	
UNDERWT_05_INV	Prevalence of underweight among 0-5 years old (inverted)	WORRY_SOME_INV	Percentage of households who worry sometimes about food (inverted)
UNDERHT_05_INV	Prevalence of stunting among 0-5 years old (inverted)	WORRY_OFTEN_INV	Percentage of households who worry often about food (inverted)
THIN_05_INV	Prevalence of wasting among 0-5 years old (inverted)	DAMAGE_TYPHOON_INV	Average annual damage to rice farming due to typhoon (inverted)
UNDERWT_510_INV	Prevalence of underweight among 5.08-10 years old (inverted)	DAMAGE_FLOOD_INV	Average annual damage to rice farming due to flood (inverted)
UNDERHT_510_INV	Prevalence of stunting among 5.08-10 years old (inverted)	DAMAGE_DROUGHT_INV	Average annual damage to rice farming due to drought (inverted)
THIN_510_INV	Prevalence of wasting among 5.08-10 years old (inverted)	INSURGENCY_INV	Incidence of attacks of insurgencies and terrorisms (inverted)
UNDERHT_1019_INV	Prevalence of stunting among 10.08-19 years old (inverted)	POP_GRATE_INV	Population growth rate (inverted)
THIN_1019_INV	Prevalence of wasting among 10.08-19 years old (inverted)	BIRTH_SPACE_17_INV	Percentage of birth spacing (7-17 months) (inverted)
CED_20_INV	Prevalence of chronic energy deficiency among adults (inverted)	BIRTH_SPACE_23_INV	Percentage of birth spacing (18-23 months) (inverted)
IDD_612_INV	Prevalence of iodine deficiency disorder among 6-12 years old (inverted)	AREA_RICE	Area harvested with palay
MORTALITY_INV	Under 5 mortality rate (inverted)	AREA_CORN	Area harvested with corn

Dimension / Sub-indicator	*Definition*	*Dimension / Sub-indicator*	*Definition*
Utilization		***Stability***	
SAFE_WATER	Proportion of households with safe source of water	NO_HOUSE_INV	Percentage of women 14-59 years old who do not own house (inverted)
EDUC	Women 14-59 years old who attended post-secondary education	NO_LAND_INV	Percentage of women 14-59 years old who do not own land (inverted)
		CPI_RICE_SD_INV	Standard deviation of CPI of rice (inverted)
		CPI_CORN_SD_INV	Standard deviation of CPI of corn (inverted)
		CPI_FRUITS_SD_INV	Standard deviation of CPI of fruits (inverted)
		CPI_VEG_SD_INV	Standard deviation of CPI of vegetable (inverted)
		CPI_MEAT_SD_INV	Standard deviation of CPI of meat (inverted)
		CPI_EGG_SD_INV	Standard deviation of CPI of egg (inverted)
		CPI_FISH_SD_INV	Standard deviation of CPI of fish (inverted)

Table B. Sobol' sensitivity measures for the average shift in provinces' ranks ($\bar{R}_s$) with respect to the Conventional Index (I_C) estimated using three formulas at varying sample sizes

Sample size	*Input factor*	*Jansen et al. (1994)*			*Nossent and Bauwens (2012) without adjustment*			*Nossent and Bauwens (2012) with adjustment*		
		S_i	S_{Ti}	S_{Ti} - S_i	S_i	S_{Ti}	S_{Ti} - S_i	S_i	S_{Ti}	S_{Ti} - S_i
n=4000	Dataset	0.198	0.203	0.004	0.207	0.245	0.038	0.194	0.219	0.024
	Normalization	0.609	0.629	0.020	1.073	0.845	-0.228	0.611	0.602	-0.009
	Weighting scheme	0.040	0.094	0.055	0.046	0.137	0.091	0.037	0.101	0.063
	Aggregation	0.053	0.055	0.003	-0.164	-0.279	-0.115	0.085	0.059	-0.025
	Availability weight	0.015	0.049	0.034	0.024	0.093	0.069	0.017	0.051	0.035

Sample size	Input factor	Jansen et al. (1994)			Nossent and Bauwens (2012) without adjustment			Nossent and Bauwens (2012) with adjustment		
		S_i	S_{Ti}	S_{Ti} - S_i	S_i	S_{Ti}	S_{Ti} - S_i	S_i	S_{Ti}	S_{Ti} - S_i
	Accessibility weight	-0.009	0.011	0.020	0.002	0.053	0.051	-0.004	0.015	0.019
	Utilization weight	0.011	0.035	0.024	0.040	0.069	0.029	0.022	0.039	0.017
	Stability weight	-0.009	0.009	0.018	0.007	0.053	0.046	0.000	0.007	0.008
	Sum indices	0.908	1.086		1.236	1.217		0.961	1.093	
n=5500	Dataset	0.194	0.199	0.005	0.179	0.236	0.057	0.180	0.210	0.030
	Normalization	0.614	0.619	0.005	1.043	0.835	-0.208	0.592	0.603	0.011
	Weighting scheme	0.055	0.095	0.040	0.048	0.117	0.070	0.041	0.098	0.057
	Aggregation	0.063	0.056	-0.007	-0.179	-0.295	-0.116	0.087	0.057	-0.029
	Availability weight	0.032	0.050	0.017	0.017	0.085	0.068	0.019	0.053	0.034
	Accessibility weight	0.005	0.011	0.006	0.000	0.036	0.036	-0.003	0.014	0.017
	Utilization weight	0.025	0.036	0.011	0.030	0.056	0.026	0.018	0.040	0.022
	Stability weight	0.004	0.008	0.004	0.008	0.033	0.025	0.000	0.006	0.006
	Sum indices	0.992	1.074		1.146	1.104		0.934	1.082	
n=7000	Dataset	0.185	0.200	0.015	0.186	0.229	0.043	0.181	0.212	0.032
	Normalization	0.609	0.633	0.024	1.044	0.832	-0.212	0.599	0.598	-0.002
	Weighting scheme	0.039	0.098	0.059	0.044	0.118	0.074	0.035	0.101	0.066
	Aggregation	0.052	0.056	0.004	-0.177	-0.307	-0.130	0.085	0.059	-0.026
	Availability weight	0.016	0.051	0.035	0.012	0.083	0.071	0.016	0.055	0.038
	Accessibility weight	-0.012	0.012	0.023	0.000	0.029	0.029	-0.004	0.012	0.016
	Utilization weight	0.011	0.037	0.026	0.026	0.057	0.030	0.017	0.039	0.021
	Stability weight	-0.010	0.008	0.019	0.010	0.027	0.018	0.001	0.006	0.005
	Sum indices	0.889	1.096		1.146	1.068		0.930	1.081	
n=11000	Dataset	0.178	0.209	0.031	0.168	0.226	0.059	0.171	0.214	0.042
	Normalization	0.606	0.637	0.031	1.043	0.827	-0.216	0.593	0.594	0.002
	Weighting scheme	0.035	0.098	0.064	0.045	0.106	0.061	0.031	0.102	0.071

Sample size	***Input factor***	***Jansen et al. (1994)***			***Nossent and Bauwens (2012) without adjustment***			***Nossent and Bauwens (2012) with adjustment***		
		S_i	S_{Ti}	S_{Ti} - S_i	S_i	S_{Ti}	S_{Ti} - S_i	S_i	S_{Ti}	S_{Ti} - S_i
	Aggregation	0.052	0.056	0.005	-0.190	-0.318	-0.128	0.084	0.062	-0.022
	Availability weight	0.016	0.053	0.037	0.021	0.070	0.049	0.018	0.054	0.036
	Accessibility weight	-0.008	0.011	0.019	0.000	0.022	0.022	-0.001	0.012	0.013
	Utilization weight	0.010	0.038	0.028	0.021	0.044	0.023	0.015	0.038	0.023
	Stability weight	-0.008	0.008	0.016	0.005	0.020	0.015	0.001	0.007	0.006
	Sum indices	0.881	1.111		1.113	0.998		0.912	1.082	

Equality of Test Statistics/Procedures for Independent Samples

Joyce Raymond B. Punzalan, Charlie S. Labina, and Therese Ann G. Capistrano
School of Statistics, University of the Philippines Diliman

In testing the null hypothesis Ho: $\mu_1 - \mu_2 = d_0$, two *t*-tests are available: the pooled and unpooled cases. Almeda et al (2010) stated that *even if the sample sizes are less than 30 but $n_1 = n_2$ use the pooled case and if the population variances are considerably different, the pooled case will still provide a good inference provided that $n_1 = n_2$ and both populations are normal. Therefore, in a planned experiment, one should make every effort to use equal sample sizes to simplify the analysis.* We derive the relationship between n_1 and n_2 that equalizes the two test statistics/procedures. $n_1 = n_2$ or $S_1^2 = S_2^2$ equalizes the two test statistics. $n_1 = n_2$ and $S_1^2 = S_2^2$ equalize the two test procedures.

Keywords: pooled t-test, unpooled t-test, degrees of freedom

1. Introduction

Consider the following problem from Walpole (1997):

A large automobile manufacturing company is trying to decide whether to purchase brand 1 or brand 2 tires for its new models. To help arrive at a decision, an experiment is conducted using 12 of each brand. The tires are run until they wear out. The results are $\bar{x}_1$ = 37,900 km., s_1 = 5100 km., $\bar{x}_2$ = 39,800 km., and s_2 = 5900 km. Test the hypothesis at the .05 level of significance that there is no difference in the two brands of tires. Assume the populations to be approximately normally distributed.

To solve this, we first specify the null hypothesis Ho: $\mu_1 - \mu_2 = d_0 = 0$ and the two-sided alternative hypothesis Ha: $\mu_1 - \mu_2 \neq 0$. Then, we choose among four possible test statistics:

$$z_1 = \frac{(\overline{X}_1 - \overline{X}_2) - d_o}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$$

cannot be used since the population standard deviations σ_1 and σ_2 are not known and

$$z_2 = \frac{(\overline{X}_1 - \overline{X}_2) - d_o}{\sqrt{(S_1^2/n_1) + (S_2^2/n_2)}}$$

is used only if the sample sizes n_1 and n_2 are large. The remaining test statistics are

$$t_1 = \frac{(\overline{X}_1 - \overline{X}_2) - d_o}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

and

$$t_2 = \frac{(\overline{X}_1 - \overline{X}_2) - d_o}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}.$$

The pooled test statistic t_1 is used when $\sigma_1^2 = \sigma_2^2$ is assumed while the unpooled test statistic t_2 is used when $\sigma_2^2 \neq \sigma_2^2$ is assumed.

In our classes, we emphasize that the equality or inequality of the unknown population variances must be stated explicitly in the problem. Since this is not the case for the problem at hand, the logical thing to do is to compare the results for t_1 and t_2. One can verify that $t_1 = t_2 = -.84396$.

The degrees of freedom of t_1 is $v_1 = n_1 + n_2 - 2$ while that of t_2 (due to Welch-Satterthwaite) is

$$v_2 = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}}.$$

For the problem at hand, one can verify that $v_1 = 22$ and $v_2 = 21.54887$, which can be rounded to 22. The p-value of the test is .407775. We get essentially the same results whether t_1 or t_2 is used. Which leads one to think: why is this so? (Some practitioners choose to be conservative by using $v_2 = 21$.)

This paper seeks to derive the relationship between n_1 and n_2 that equalizes the two test statistics/procedures. The derivation, which is a good exercise for students of Inference and mathematically-inclined Elementary Statistics, can be found in Section 2. Section 3 presents SAS outputs to illustrate the results. Finally, Section 4 contains remarks.

2. Derivation

To show that $t_1 = t_2$, it is sufficient to show that their denominators are equal. That is,

$$\frac{n_1 - 1}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \frac{1}{n_1} \quad \text{and} \quad \frac{n_2 - 1}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \frac{1}{n_2}.$$

Or equivalently,

$$n_1(n_1-1)(n_2+n_1) = n_1 n_2(n_1+n_2-2)$$

and

$$n_2(n_2-1)(n_2+n_1) = n_1 n_2(n_1+n_2-2).$$

These simultaneous equations imply that

$$n_1(n_1-1) = n_2(n_2-1)$$

Or equivalently in terms of n_1 alone or n_2 alone

$$(n_1-n_2)[n_1-(1-n_2)] = 0 \text{ or } (n_2-n_1)[n_2-(1-n_1)] = 0.$$

The solution set for n_1 is $\{n_2, 1-n_2\}$ while for n_2 is $\{n_1, 1-n_1\}$. Since 1- n_1 and 1- n_2 are not positive, we get Result 1.

Result 1: $t_1 = t_2$ = if and only if $n_1 = n_2$.

If $S_1^2 = S_2^2 = S^2$, the squared denominator of t_1 becomes

$$\frac{(n_1-1)S^2+(n_2-1)S^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right) = \frac{S^2(n_1-1+n_2-1)}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$$

$$= S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$$

And the squared denominator of t_2 becomes

$$\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2} = S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$$

Result 2: $t_1 = t_2$ if $S_1^2 = S_2^2$.

Moore et al (2012) noted that $\min(n_1 - 1, n_2 - 1) \le v_2 \le v_1$.

If $n_1 = n_2 = n$ then

$$v_2 = \frac{(S_1^2/n+S_2^2/n)^2}{\frac{(S_1^2/n)^2}{n-1}+\frac{(S_2^2/n)^2}{n-1}} = \frac{(S_1^2+S_2^2)^2/n^2}{\frac{S_1^4}{n^2(n-1)}+\frac{S_2^4}{n^2(n-1)}} = \frac{(S_1^2+S_2^2)^2}{(S_1^4+S_2^4)(n-1)}$$

and

$$v_1 = n+n-2 = 2(n-1)$$

So when $n_1 = n_2 = n$, then $v_1 = v_2$ if

$$2(S_1^4 + S_2^4) = (S_1^2 + S_2^2)^2 = S_1^4 + S_2^4 + 2S_1^2 S_2^2 \text{ or if}$$

$$S_1^4 + S_2^4 = 2S_1^2 S_2^2 \text{ or if } (S_1^2 - S_2^2)^2 = 0 \text{ or if } S_1^2 = S_2^2$$

Result 3: The pooled and unpooled t-tests are equivalent when $n_1 = n_2$ and $S_1^2 = S_2^2$

3. SAS Outputs

The tables in this section were generated using SAS.They illustrate numerically what was derived in Section 2.

Let dataset 1 = {1, 2, 3, 4} and dataset 2 = {5, 6, 7}. $n_1 \neq n_2$ so we expect $t_1 \neq t_2$ as can be seen in the fourth column of Table 1.

Table 1 $t_1 \neq t_2$ for $n_1 \neq n_2$

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	5	-3.87	0.0117
Satterthwaite	Unequal	4.9592	-4.04	0.0101

To illustrate Result 1, let dataset 1 = {1, 2, 3, 4} and dataset 2 = {5, 6, 7, 9}. $n_1 = n_2$ so we expect $t_1 = t_2$ as can be seen in the fourth column of Table 2.

Table 2 $t_1 = t_2$ for $n_1 = n_2$

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	6	-3.97	0.0074
Satterthwaite	Unequal	5.5846	-3.97	0.0085

To illustrate Result 2, let dataset 1 = {1, 2, 3} and dataset 2 = {9, 9, 10, 11, 11}. Here, $S_1^2 = S_2^2 = 1$ and we have $t_1 = t_2$ in the fourth column of Table 3.

Table 3 $t_1 = t_2$ for $S_1^2 = S_2^2$

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	6	-10.95	<.0001
Satterthwaite	Unequal	4.339	-10.95	0.0003

To illustrate Result 3, let dataset 1 = {1, 2, 3, 4} and dataset 2 = {5, 6, 7, 8}. and = 5/3 so we expect $t_1 = t_2$ and $v_1 = v_2$ as can be seen in the third and fourth columns of Table 4. The last column gives equivalent p-values. Welch-Satterthwaite might have results 2 and 3 in mind when they derived the degrees of freedom of t2; that the degrees of freedom is approximately that of t1 when $n_1 = n_2$ and $S_1^2 = S_2^2$.

Table 4 $t_1 = t_2$ for $n_1 = n_2$ *and* $S_1^2 = S_2^2$

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	6	-4.38	0.0047
Satterthwaite	Unequal	6	-4.38	0.0047

4. Remarks

In our search for references for this paper, we found the essence of Result 1 in Utts and Heckard (2004) and in Almeda et al (2010). According to Walpole et al (2002), if σ_1^2 and σ_2^2 are considerably different, t_1 can still be used provided that $n_1 = n_2$. This paper further shows that (Result 2) $t_1 = t_2$ if $S_1^2 = S_2^2$ and that (Result 3) the two testing procedures are equivalent when $n_1 = n_2$ and $S_1^2 = S_2^2$.

Although the pooled procedure is reasonably robust against nonnormality and unequal standard deviations when n_1 and n_2 are nearly the same, Moore et al (2012) cautioned that the assumption $\sigma_1^2 = \sigma_2^2$ is hard to verify thus making the procedure risky. When n_1 and n_2 are quite different, the pooled procedure becomes sensitive to unequal standard deviations.

Utts and Heckard (2004) enumerated guidelines for using the pooled t-test. (1) When n_1 and n_2 are very different, the pooled test can be quite misleading unless s_1 and s_2 are similar. (2) If n_1 and n_2 are very different and the smaller standard deviation accompanies the larger sample size, we do not recommend using the pooled procedure. (3) If n_1 and n_2 are very different, s_1 and s_2 are similar, and the larger sample size produced the larger standard deviation, the pooled test is acceptable because it will be conservative. Generally, it's best to use the unpooled procedure unless s_1 and s_2 are quite similar.

References

ALMEDA, J. et al (2010), Elementary Statistics, Quezon City: The University of the Philippines Press

HINES, W. (1996), "Pragmatics of Pooling in ANOVA Tables," The American Statistician, 50(2), 127-139.

MOORE, D. et al (2012), Introduction to the Practice of Statistics, 7th Edition, New York: W. H. Freeman and Company

UTTS, J. AND HECKARD, R. (2004), Mind on Statistics, 2nd Edition, Canada: Thomson Learning, Inc.

WALPOLE, R. (1997), Introduction to Statistics, 3rd Edition, Singapore: Simon & Schuster Asia Pte Ltd.

WALPOLE, R. et al (2002), Probability and Statistics for Engineers and Scientists, 7th Edition, New Jersey: Prentice-Hall, Inc.