Statistics Netherlands

Discussion Paper

Scalar measures of uncertainty in reconciled economic accounts

The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

2016 | 09

Nino Mushkudiani, Jeroen Pannekoek, Li-Chun Zhang

Summary: Many macro-economic figures, for instance arising in the context of national economic and social accounting systems, are connected by known constraints. We refer to such identities as accounting equations. The actual input estimates, often based on a variety of sources, usually do not automatically satisfy the relationship due to measurement and sampling errors. The estimation of an accounting equation involves then an adjustment or reconciliation step by which the input estimates are modified to conform to the known identity. In this paper we consider the measurement of uncertainty in such reconciled estimated accounting equations. The initial input uncertainty can in principle be quantified by a joint covariance or mean squared error matrix. However, this does not take into account the adjustment effects. Moreover, a matrix of uncertainty measures like this is not easy to present or digest. We consider an accounting equation as a single entity and develop scalar uncertainty measures that more directly capture the adjustment effect as well as the relative contribution of the various input estimates to the final estimated account. We develop two approaches for defining these scalar measures. For each approach we define two variants of scalar measures and using a simple simulation example demonstrate the application of these measures.
Keywords: ***macro account; input estimates adjustment; scalar uncertainty measure; permutation invariance.***

1 Introduction

Many macro-economic figures become much more meaningful if they are viewed in relation to other variables rather than in isolation. Often, such variables are presented together to emphasize and depict their relationship, which may even be the main feature of interest. This is especially the case for variables that can be viewed as components of equations. Consider, for instance, the total labor cost for a sector of the economy. This total variable can be decomposed into the product of the number of full-time employees and the average wage. Together these three variables represent the structure of the labor cost. Similarly, the total value of the production of an economy can be decomposed as the sum of the production of a number of categories of goods and services. In this paper, we view estimates of such equations as single entities and we are interested in a scalar measure for the uncertainty in such estimates.

The equations exemplified above are examples of, what we will refer to as, *accounting equations* or *constraints*. The multiplicative and additive constraints can be represented, respectively, as

$$Y_1 Y_2 = Z \tag{1}$$

and

$$Y_1 + \dots + Y_i + \dots + Y_p = Z \tag{2}$$

Note that we have $p = 2$ for the multiplicative constraint and $p > 2$ for the additive one. In general, we allow $p > 2$ for the multiplicative constraint as well, but the case of $p = 2$, as illustrated above, seems the most common with this type of accounting equation.

In practice, the true but unknown values of the variables (Y_i, Z) are often estimated from multiple sources. Let the true values be denoted by (y_{0i}, z_0). Then, for a multiplicative constraint (the development for an additive constraint is similar), we define the *true account* A_0 by

$$A_0 = [y_{01} y_{02} = z_0]$$

which satisfies (1) by definition. In other words, the constraint (1) is a theoretical property of the true account that must be satisfied by any account actually compiled. The notation $A_0 = [\cdots]$ is introduced to emphasize that the account is conceived as a *single* entity. Denote by $(\hat{Y}_i, \hat{Z})$ the initial *input estimates* of the variables appearing in the account. Usually, these estimates will not satisfy the constraint, i.e. $\hat{Y}_1 \hat{Y}_2 \neq \hat{Z}$. Suppose that, after an adjustment of choice, we obtain adjusted values $(\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{Z})$ that do satisfy the constraint, so that we have

$$A = [\widetilde{Y}_1 \widetilde{Y}_2 = \widetilde{Z}]. \tag{3}$$

We shall refer to (3) as the *estimated account*. The estimated account is considered to be a *single* random outcome. Possible different realisations could be, for instance, $[100 \cdot 1.2 = 120]$ or $[80 \cdot 1.5 = 120]$.

The expected account can be defined as the one we would obtain on average, under any well-defined joint distribution of $(\widetilde{Y}_1, \widetilde{Y}_2, \cdots, \widetilde{Y}_p, \widetilde{Z})$, denoted by $f(\widetilde{Y}_1, \widetilde{Y}_2, \cdots, \widetilde{Y}_p, \widetilde{Z})$. While the expected additive account is, straightforwardly, obtained by substitution of the expectations of the component variables, the expected multiplicative account is more involved to be obtained exactly. As an alternative, we could transform the multiplicative account into an additive one by transforming to the log-scale.

An obvious measure of the uncertainty in the input vector $\hat{X} = (\hat{Y}_1, \hat{Y}_2, \cdots, \hat{Y}_p, \hat{Z})$ is the $(p+1) \times (p+1)$ variance-covariance matrix, $\Sigma_{\hat{X}}$ say, of this vector. Similarly, the uncertainty of an account could be measured by the covariance matrix of the component estimates after adjustment, $\Sigma_{\widetilde{X}}$ say of $\widetilde{X} = (\widetilde{Y}_1, \widetilde{Y}_2, \cdots, \widetilde{Y}_p, \widetilde{Z})$. This approach to measuring uncertainty in accounts is investigated by, among others, Stone(1942). A drawback of the covariance matrix as a measure of uncertainty is its multidimensional nature which makes it difficult to interpret and hampers, for instance, to make comparisons of uncertainties of different estimates of the same account. We therefore consider a *scalar* measure of uncertainty. This scalar measure depends on three components: the vector of unadjusted estimates, accounting equation that should be fulfilled and the adjustment method used for obtaining the estimated vector $\widetilde{X}$. We shall consider two different approaches for defining a scalar measure of uncertainty: covariance approach and deviation approach. Respectively by τ and Δ we define scalar measures of the covariance and the deviation approach. These scalar measures would be helpful in at least two respects:

1. Let $\dot{X} = (\dot{Y}_1, \dot{Y}_2, \cdots, \dot{Y}_p, \dot{Z})$ be an alternative adjustment to $\widetilde{X}$, both having the same input estimates $\hat{X}$ with respect to either (1) or (2). Let $\dot{A}$ and A be the corresponding estimated accounts for $\dot{X}$ and $\widetilde{X}$. Let $\tau(A)$ and $\tau(\dot{A})$ be the respective scalar uncertainty measures. The relative efficiency (RE), i.e. $\tau(A)/\tau(\dot{A})$, would facilitate the choice between the two. This holds also for the uncertainty measure Δ.
2. Write $\tau(A; F_{\hat{X}})$ to signify its dependence on $F_{\hat{X}}$, the distribution of $\hat{X}$. Any change in the input uncertainty would result in a different $F'_{\hat{X}}$, and $\tau(A; F'_{\hat{X}})$ accordingly. The RE $\tau(A; F'_{\hat{X}})/\tau(A; F_{\hat{X}})$ could then help us to identify and to assess the changes, or improvements, that are most effective in terms of the final estimated account directly. This holds also for the uncertainty measure Δ.

The first, *covariance*, approach starts with the covariance matrix Σ of $\widetilde{X}$ as a multivariate measure of the expected deviation of $\widetilde{X}$ from its expected value. This multivariate measure is then reduced to a scalar summary. The second *deviation* approach first reduces the deviation of $\widetilde{X}$ from its expected value to a scalar summary measure. The uncertainty measure is then defined as the expectation of this scalar deviation measure. Thus, the first approach considers a scalar summary of the expected value of multivariate deviations whereas the second approach considers the expected value of scalar summaries of multivariate deviations.

Scalar summary measures of multivariate variability based on the covariance matrix, and their properties, have been studied by e.g. Peña and Rodríguez (2000). One such measure is the average total variation (see, Seber (1984)) which is also considered here as a possibility for constructing an uncertainty measure for an account. Scalar measures for the deviation between a realized and expected account, as in the second approach, summarize the component-wise differences for which summary measures based on L_1 and L_2 norms will be considered.

Reducing the information in the covariance matrix, or the component-wise deviations, to a scalar summary will, in general, lead to a loss of information. A common issue in the development is therefore to clarify in which sense a chosen measure summarizes as much as possible the relevant uncertainty.

The rest of this paper is organized as follows. In section 2 the covariance approach and the deviation approach are developed. In Section 3 we illustrate the theory for an additive account. An illustration is given in Section 4 for the so-called Index Problem, which is a multiplicative account. An application to the National Account supply and use (SU) tables is outlined in Section 5, which is currently being developed.

2 Development of the covariance and deviation approaches

In this section we will develop the two different approaches. Before that, however, we would like to clarify a necessary property that any admissible measure should possess.

2.1 Permutation invariance

Under the same additive constraint (2), any of the following *permutation rearrangements* of A can be considered as equivalent to each other:

$$[(-\widetilde{Z}) + \widetilde{Y}_2 + \cdots + \widetilde{Y}_p = (-\widetilde{Y}_1)], \quad [\widetilde{Y}_2 + \widetilde{Y}_1 + \cdots + \widetilde{Y}_p = \widetilde{Z}], \quad \ldots$$

In other words, the appearance of the $\widetilde{Y}$'s and $\widetilde{Z}$ in an additive account should not matter, because we can always rearrange their positions in the account without changing the constraint. Thus, a scalar uncertainty measure $\tau(A)$ or $\Delta(A)$ is said to be *strongly permutation invariant* if it remains the same for any permutation rearrangement of A.

Similarly, permutation rearrangements can be applied to a multiplicative account under the same constraint. But the interpretation equivalence may be less intuitive. For instance, [Value · (1/Quantum) = Price] is possibly just as acceptable as [Quantum · Price = Value], whereas [Quantum · (1/Value) = (1/Price)] may seem somewhat unnatural. We consider two alternatives. Firstly, provided the components are all strictly positive, one may measure the accounting uncertainty on the log scale, where equivalent permutation rearrangements is more readily acceptable:

$$[\log \text{Quantum} + \log \text{Price} = \log \text{Value}], \quad [\log \text{Quantum} + (-\log \text{Value}) = (-\log \text{Price})], \quad \ldots$$

Secondly, a scalar uncertainty measure $\tau(A)$ or $\Delta(A)$ is said to be *weakly permutation invariant* if it remains the same for any permutation rearrangement of the left-hand side of A, i.e. a permutation of only the $\widetilde{Y}$'s given the choice of Z. We consider weakly permutation invariance primarily as an additional possibility for constructing admissible uncertainty measures for the multiplicative account, as discussed in Section 2.2 below.

2.2 Covariance approach

We start with the additive account. In practice we observe the initial input estimates $(\hat{Y}_i, \hat{Z})$ of the true values (Y_i, Z). The true values and the estimated values $(\widetilde{Y}_i, \widetilde{Z})$ satisfy the constraint (2). For notational convenience we will sometimes combine $\widetilde{Y}$ and $\widetilde{Z}$ in one vector $\widetilde{X}$:

$$\widetilde{X} = (\widetilde{Y}_1, \widetilde{Y}_2, \cdots, \widetilde{Y}_p, \widetilde{Z}) = (\widetilde{X}_1, \widetilde{X}_2, \cdots, \widetilde{X}_p, \widetilde{X}_{p+1}).$$

Due to the constraint in (2) the variance-covariance matrix $\Sigma_{\widetilde{X}}$ has the following formal structure

$$\Sigma_{\widetilde{X}} = \begin{pmatrix} \Sigma_{\widetilde{Y}} & \Sigma_{\widetilde{Y}\widetilde{Z}} \\ \Sigma_{\widetilde{Z}\widetilde{Y}} & \Sigma_{\widetilde{Z}} \end{pmatrix} = \begin{pmatrix} \Sigma_{\widetilde{Y}} & \Sigma_{\widetilde{Y}}1 \\ 1^T\Sigma_{\widetilde{Y}} & 1^T\Sigma_{\widetilde{Y}}1 \end{pmatrix}$$

Scalar measure for the covariance approach are defined as follows:

$$\tau_1(A) = 1^T\Sigma_{\tilde{X}}1 = 4(1^T\Sigma_{\tilde{Y}}1) \qquad \text{and} \qquad \tau_2(A) = \text{Trace}(\Sigma_{\tilde{X}}) = \text{Trace}(\Sigma_{\tilde{Y}}) + 1^T\Sigma_{\tilde{Y}}1 \tag{4}$$

where $\tau_1(A) = V(\sum_{k=1}^{p} \tilde{Y}_k + \tilde{Z}) = V(2\tilde{Z})$, and $V(\tilde{Z}) = 1^T\Sigma_{\tilde{Y}}1$, and
$\text{Trace}(\Sigma_{\tilde{X}}) = \sum_{k=1}^{p} V(\tilde{Y}_k) + V(\tilde{Z})$. τ_2 is also called the total variation, see Seber (1984).

Observe that τ_2 is strongly permutation invariant, whilst τ_1 is only weakly permutation invariant. To see this consider again the two permutations of A

$$[\tilde{Y}_1 + \tilde{Y}_2 + \cdots + \tilde{Y}_p = \tilde{Z}] \quad \text{and} \quad [(-\tilde{Z}) + \tilde{Y}_2 + \cdots + \tilde{Y}_p = (-\tilde{Y}_1)].$$

Define $\tilde{X}_1 = (\tilde{Y}_1, ..., \tilde{Y}_p, \tilde{Z})^T$ and $\tilde{X}_2 = ((-\tilde{Z}), \tilde{Y}_2 ..., \tilde{Y}_p, (-\tilde{Y}_1))^T$, with covariance matrices $\Sigma_{\tilde{X}_1}$ and $\Sigma_{\tilde{X}_2}$. Then, since $1^T\Sigma_{\tilde{X}_1}1 = 4V(\tilde{Z})$ and $1^T\Sigma_{\tilde{X}_2}1 = 4V(-\tilde{Y}_1)$ we would have different values for τ_1. This does not hold for τ_2, which remains the same for all permutations.

Due to the constraint (2), τ_1 actually contains `less' of the information in $\Sigma_{\tilde{X}}$, now that $\tau_2 = \text{Trace}(\Sigma_{\tilde{Y}}) + \tau_1/4$.

Moreover, in many practical situations, some of the inputs may be known *and* held fixed in the adjustment, such as when certain fiscal figures are obtained from the tax office for the Structural Business Statistics. The measure τ_1 could be 0 in such cases, if we rearrange the account so that the right-hand side consists only of the fixed components. Then the variables on the left hand side, denoted by $\tilde{Y} = (\tilde{Y}_1, ..., \tilde{Y}_{p^*})^T$, will sum up to a constant z, where p^* is the adjusted number of components on the left-hand side. Our estimated account will be

$$A = [\tilde{Y}_1 + \tilde{Y}_2 + \cdots + \tilde{Y}_{p*} = z]$$

Then $\Sigma_{\tilde{X}}$ is the variance-covariance matrix of $\tilde{Y}$ and we will have that
$\tau_1 = 1^T\Sigma_{\tilde{X}}1 \propto 1^T\Sigma_{\tilde{Y}}1 = V(z) = 0$. In contrast, $\tau_2 = \text{Trace}(\Sigma_{\tilde{Y}}) + V(z) = \text{Trace}(\Sigma_{\tilde{Y}})$ would remain a meaningful measure of the accounting uncertainty.

Consider now the multiplicative account with $p = 2$. As noted previously, an alternative is to measure the accounting uncertainty on the log scale, which is additive. Applying (3) at the original scale is less convenient, we would have to calculate $E\tilde{Z} = E(\tilde{Y}_1 * \tilde{Y}_2) \neq E(\tilde{Y}_1) * E(\tilde{Y}_2)$. On the log scale for $B = [\log(\tilde{Y}_1) + \log(\tilde{Y}_2) = \log(\tilde{Z})]$ we can define

$$\tau_1(B) = V(\log(\tilde{Y}_1) + \log(\tilde{Y}_2) + \log(\tilde{Z})) = 4V(\log(\tilde{Z})) \tag{5}$$

and

$$\tau_2(B) = V(\log(\tilde{Y}_1)) + V(\log(\tilde{Y}_2)) + V(\log(\tilde{Z})). \tag{6}$$

Another more intuitive approach is via weakly permutation invariance. By this approach we define the scalar measure $\tau(A)$ for $A = [\tilde{Y}_1\tilde{Y}_2 = \tilde{Z}]$, which formally is more closely related to $\Sigma_{\tilde{X}}$ and (4) on the original scale. Let $(\tilde{Z}, D)$ be a one-one transformation of $(\tilde{Y}_1, \tilde{Y}_2)$, so that D summarizes all the *conditional* variation in $(\tilde{Y}_1, \tilde{Y}_2)$ given $\tilde{Z}$. Weakly permutation invariance is achieved provided $V(D)$ is invariant for $\tilde{Y}$-permutations. For example, $D(\tilde{Y}_1, \tilde{Y}_2) = \tilde{Y}_1/\tilde{Y}_2$ does not satisfy this requirement, because $D(\tilde{Y}_1, \tilde{Y}_2) = \tilde{Y}_1^2/Z \neq \tilde{Y}_2^2/Z = D(\tilde{Y}_2, \tilde{Y}_1)$. A simple admissible choice is $D(\tilde{Y}_1, \tilde{Y}_2) = \tilde{Y}_1 + \tilde{Y}_2$, which yields $V(D) = 1^T\Sigma_{\tilde{Y}}1$. Recall the total variance formulae, then for D and $\tilde{Z}$ we can write that:

$$V(D) = E_{\tilde{Z}}(V(D|\tilde{Z})) + V_{\tilde{Z}}(E(D|\tilde{Z})) \tag{7}$$

and using the definition of covariance

$$Cov(\tilde{Y}_1, \tilde{Y}_2|\tilde{Z}) = E(\tilde{Y}_1\tilde{Y}_2|\tilde{Z}) - E(\tilde{Y}_1|\tilde{Z})E(\tilde{Y}_2|\tilde{Z}) = \tilde{Z} - E(\tilde{Y}_1|\tilde{Z})E(\tilde{Y}_2|\tilde{Z}). \tag{8}$$

In analogy to (7) we define a scalar measure for $D(\widetilde{Y}_1, \widetilde{Y}_2) = \widetilde{Y}_1 + \widetilde{Y}_2$:

$$\tau(A) \overset{def}{=} E[V(D|\widetilde{Z})] + V(\widetilde{Z})$$

now we can apply (8) to the fist term $E[V(D|\widetilde{Z})] = E[V(\widetilde{Y}_1 + \widetilde{Y}_2|\widetilde{Z})]$

$$\tau(A) = E[V(\widetilde{Y}_1|\widetilde{Z})] + E[V(\widetilde{Y}_2|\widetilde{Z})] + 2E(\widetilde{Z}) - 2E(E(\widetilde{Y}_1|\widetilde{Z})E(\widetilde{Y}_2|\widetilde{Z})) + V(\widetilde{Z}) \tag{9}$$

The measure (9) involves in a way the first two moments of $\widetilde{Y}_1$, $\widetilde{Y}_2$ and $\widetilde{Z}$ on the original scale. It is weakly permutation invariant given the choice of D.

2.3 Deviation approach

Let us again consider the additive account at first. Let $M_X = (M_1, ..., M_p, M_{p+1} = M_Z)^T$ be the expectation of $\widetilde{X}$, such that $\widetilde{X} - M_X$ contains the deviation of all the components of the final estimated account from those of the expected account. We can adopt a suitable scalar measure to summarize the component-wise deviation, generically denoted by $\delta = \delta(\widetilde{X} - M_X)$. Different choices are possible. Two of them are

$$\delta_1 = \sum_{k=1}^{p+1} w_k|\widetilde{X}_k - M_k| \qquad \text{and} \qquad \delta_2 = \sum_{k=1}^{p+1} w_k(\widetilde{X}_k - M_k)^2 \tag{10}$$

where $w_k \geq 0$, to be referred to as the averaging weights. As a scalar uncertainty measure, we use

$$\Delta(A) = E(\delta). \tag{11}$$

Depending on the choice of w_k's, the deviation approach (11) can easily be made strongly permutation invariant. For δ_1 and δ_2 above, this is the case because the w_k's are component-specific rather than position-specific. We can vary the weights given to the different components, including $w_k = 0$ for those that are known and held fixed in the adjustment. Notice that Δ_2 is closely related to τ_2 under the covariance approach: setting $w_i \equiv 1$ yields a measure that is equal to $\tau_2(A)$ by (4). In this way we may consider the deviation approach based on Δ_2 to be a generalization of the covariance approach based on τ_2.

For an additive account, when the sum z is fixed and known, we can view the account as a decomposition of this total z into its components $\widetilde{Y}_1, \dots, \widetilde{Y}_p$, or as a classification of the amount z into p-categories. The differences $\widetilde{Y}_k - M_k$ can be interpreted as classification errors that sum to zero over k, thus the sum of the positive differences equals minus the sum of the negative ones. The sum of the absolute differences is related to the Earth Movers Distance (EMD) which is a measure for the distance between two discrete distributions, it measures the amount of mass that must be redistributed to go from one distribution ($\widehat{Y}_k$) to the other ($\widetilde{Y}_k$) (see e.g. Hitchcock 1941). The measure δ_1 is the weighted sum of the absolute differences and an obvious choice for the weights w_k will be $1/2$, as in this case δ_1 is the amount of z that is classified differently by $\widetilde{Y}_k$ than by M_k. For δ_2 a natural choice for w_k's will be $1/p$, as it then becomes the mean squared classification error.

Consider now the multiplicative account where $p = 2$, and this expected account on the log scale

$$\mu(B) = [E(\log(\widetilde{Y}_1)) + E(\log(\widetilde{Y}_2)) = E(\log(\widetilde{Z}))].$$

For $m_k = E(\log(\widetilde{Y}_k))$ for $k = 1, 2$ and $\widetilde{Z}$, we have that $\mu(B) = [m_1 + m_2 = m_{\widetilde{Z}}] = [e^{m_1}e^{m_2} = e^{m_{\widetilde{Z}}}]$. It seems natural to use the proportional deviation here, i.e. $\widetilde{Y}_1/e^{m_1}$, $\widetilde{Y}_2/e^{m_2}$ and $\widetilde{Z}/e^{m_{\widetilde{Z}}}$. To measure in a symmetric manner the departure on each side of unity, a transformation to the log scale is commonly employed. This yields then

$$\delta_1 = \sum_{k=1}^{3} w_k |\log(\widetilde{X}_k) - m_k| \qquad \text{and} \qquad \delta_2 = \sum_{k=1}^{3} w_k (\log(\widetilde{X}_k) - m_k)^2 \tag{12}$$

and the corresponding $\Delta(A)$ by (11). Again, this amounts to measuring the uncertainty of a multiplicative account A in terms of the corresponding B on the log scale.

3 Illustration: Additive account

In this section we will illustrate the theory developed above for an additive account on a simple example. In practical situations estimated figures are obtained in many different ways, these figures do not necessarily satisfy economic constraints we put on them. Therefore figures are adjusted so that economic constraints hold.

3.1 Assumptions

Suppose we have the unadjusted variables $Y = (Y_1, \dots, Y_p) \sim N_p(\mu, \Sigma_Y)$, where $\mu = (\mu_1, \dots, \mu_p)$ and Σ_Y is a diagonal matrix with the diagonal values $\sigma_k^2 = V(Y_k)$, for $k = 1, \dots, p$. Suppose also that $\widetilde{Z}$ is fixed, $\widetilde{Z} = z$ and the variances of the unadjusted variables σ_k^2's are proportional to the corresponding expected values, μ_k's:

$$\sigma_k^2 = \sigma^2 \mu_k, \quad \text{for } \; k = 1, \dots, p. \tag{13}$$

This will lead to adjustments proportional to the size of the original figures. Smaller figures will be adjusted less and the large figures will be adjusted more, in absolute value.

We want to adjust Y so that the adjusted vector $\widetilde{Y}$ satisfies the following constraint:

$$A = [\widetilde{Y}_1 + \dots + \widetilde{Y}_p = z]. \tag{14}$$

This can be achieved in many different ways. Here we adjust the variables by correcting the total misclassification $(z - \sum Y_j)$:

$$\widetilde{Y}_k = Y_k + (z - \sum_{j=1}^{p} Y_j)\nu_k \quad k = 1, \dots, p, \tag{15}$$

where ν_k, $k = 1, \dots, p$ denote relative weights that sum up to 1. We call these weights the adjustment weights. Here ν_k could depend on Y_k, its variance or expectation. The expectation and variance of $\widetilde{Y}_k$'s for non-stochastic ν_k are derived in the Appendix.

3.2 Adjustment methods

It is not obvious what the best choice is for the adjustment weights in (15). We consider two different adjusted vectors $\widetilde{Y}_1$ and $\widetilde{Y}_2$ for $\nu_{1k} = 1/p$ and $\nu_{2k} = \mu_k / \sum_{j=1}^{p} \mu_j$:

$$\widetilde{Y}_{1k} = Y_k + \frac{1}{p}(z - \sum_{j=1}^{p} Y_j) \quad k = 1, \dots, p, \tag{16}$$

and

$$\widetilde{Y}_{2k} = Y_k + \frac{\mu_k}{\sum_{j=1}^{p} \mu_j}(z - \sum_{j=1}^{p} Y_j) \quad k = 1, \dots, p. \tag{17}$$

In the next subsection we derive uncertainty measures for these adjustment methods.

3.3 Strongly permutation invariant uncertainty measures

As we have mentioned already deviation uncertainty measures for additive account are in principle strongly permutation invariant. For $\widetilde{Y}$ defined in (15), and for the special cases $\widetilde{Y}_1$ and $\widetilde{Y}_2$ we can derive $\Delta_1(A|z)$ and $\Delta_2(A|z)$ analytically. With the notation $(\cdot|z)$ we want to indicate that $\widetilde{Z} = z$ is fixed. First we consider $\Delta_1(A|z)$. Following the deviation approach from (10) and (11) we have that:

$$\Delta_1(A|z) = E(\delta_{1|z}) = \sum_{k=1}^{p+1} E(w_k|\widetilde{X}_k - M_k|) = \sum_{k=1}^{p} E(w_k|\widetilde{Y}_k - M_k|), \tag{18}$$

since $E(z) = z$. Here $M_k = E(\widetilde{Y}_k)$, for $k = 1, \dots, p$ are defined in the Appendix. Note that the random variables $\widetilde{Y}_k - M_k$, $k = 1, \dots, p$ are negatively correlated with each other and have normal distributions $N(0, \widetilde{\sigma}_k^2)$, where $\widetilde{\sigma}_k^2$ is the variance of $\widetilde{Y}_k$ defined in (30) in the Appendix. Then $|\widetilde{Y}_k - M_k|$, for $k = 1, \dots, p$ have half-normal distributions with expectation equal to $\widetilde{\sigma}_k\sqrt{2/\pi}$ and we will have that:

$$\Delta_1(A|z) = \sum_{k=1}^{p} w_k \widetilde{\sigma}_k \sqrt{2/\pi}. \tag{19}$$

Let $w_k = w$ does not depend on k. Define

$$A_j = [\widetilde{Y}_{j1} + \dots + \widetilde{Y}_{jp} = z], \quad \text{for} \quad j = 1, 2.$$

Applying results from (19) and (31) in the Appendix we can derive the uncertainty measures $\Delta_1(A_1|z)$ and $\Delta_1(A_2|z)$ for the adjustments $\widetilde{Y}_{1k}$ and $\widetilde{Y}_{2k}$ defined in (16) and (17):

$$\Delta_1(A_1|z) = \sigma w \sqrt{2/\pi} \sum_{k=1}^{p} \sqrt{\frac{1}{p^2}\sum_{j=1}^{p} \mu_j + \mu_k(1 - \frac{2}{p})}$$

and

$$\Delta_1(A_2|z) = \sigma w \sqrt{2/\pi} \sum_{k=1}^{p} \sqrt{\mu_k - \frac{\mu_k^2}{\sum_{j=1}^{p} \mu_j}}.$$

As we mentioned above these measures are strongly permutation invariant because the w_k's are component-specific rather than position-specific. To compare these measures observe that for the terms under the square root we can show that:

$$\frac{1}{p^2}\sum_{j=1}^{p} \mu_j + \mu_k(1 - \frac{2}{p}) \geq \mu_k - \frac{\mu_k^2}{\sum_{j=1}^{p} \mu_j}.$$

For the second adjustment method the uncertainty measure Δ_1 is smaller and therefore we would choose the second adjustment.

For the uncertainty measure $\Delta_2(A|z)$ by again applying (10) and (11) we will have that:

$$\Delta_2(A|z) = E(\delta_{2|z}) = \sum_{k=1}^{p} E(w_k(\tilde{Y}_k - M_k)^2) = \sum_{k=1}^{p} w_k V(\tilde{Y}_k) = \sum_{k=1}^{p} w_k \tilde{\sigma}_k^2. \tag{20}$$

As above, consider adjustments defined in (16) and (17) with $\sigma_k^2 = \sigma^2 \mu_k$. Next we apply results in (32) in the Appendix:

$$\Delta_2(A_1|z) - \Delta_2(A_2|z) = \sum_{k=1}^{p} w_k \tilde{\sigma}_{1k}^2 - \sum_{k=1}^{p} w_k \tilde{\sigma}_{2k}^2$$

$$= \sum_{k=1}^{p} w_k (V(\tilde{Y}_{1k}) - V(\tilde{Y}_{2k})) = \sum_{k=1}^{p} w_k \sigma^2 \left(\frac{\sqrt{\sum_{j=1}^{p} \mu_j}}{p} - \frac{\mu_k}{\sqrt{\sum_{j=1}^{p} \mu_j}} \right)^2 .$$

Hence $\Delta_2(A_1|z) \geq \Delta_2(A_2|z)$ and $\Delta_2(A_1|z)$ is always more uncertain than $\Delta_2(A_2|z)$.

Both our measures are more uncertain for adjustment $\tilde{Y}_1$. Hence we would obviously choose the second adjustment $\tilde{Y}_2$.

For more complex situations we might not have analytic results for these uncertainty measures, but we can estimate $\Delta(A|z)$ by Monte Carlo simulation.

3.4 Monte Carlo simulation

Suppose again that $\tilde{Z} = z$ is fixed and the unadjusted variables $Y = (Y_1, \dots, Y_p) \sim N_p(\mu, \Sigma_Y)$, where $\mu = (\mu_1, \dots, \mu_p)$ and Σ_Y is a diagonal matrix with the diagonal values $\sigma_k^2 = V(Y_k)$ and consider the same constraint as above:

$$A = [\tilde{Y}_1 + \cdots + \tilde{Y}_p = z]$$

In general the adjusted $\tilde{Y}$ cannot always be given in a closed expression. In such case we have to apply some method to estimate $\tilde{Y}$ indirectly. For example $\tilde{Y}$ could be defined by solving the optimization problem: Find $\tilde{Y}$ close to Y in some sense, so that the constraint A holds true. Here we are not concerned by the method for estimating $\tilde{Y}$. We want to use the measures defined above to compare different adjusted $\tilde{Y}$'s. For convenience we will call the process of obtaining $\tilde{Y}$, a scheme: $(Y, z) \rightarrow \tilde{Y}$; In general we can calculate $\tau_1(A|z)$ and $\tau_2(A|z)$, and $\Delta_1(A|z)$ and $\Delta_2(A|z)$ by the following simulation approach:

1. For Y define or estimate its expectation and variance;
2. Identify constraint that should be fulfilled, an account A;
3. Define a method for estimating $\tilde{Y}$ from Y, satisfying the account A;
4. For large N simulate N vectors $Y^{(i)} \sim N_p(\mu, \sigma_Y)$, $i = 1, \dots, N$;
5. For each $Y^{(i)}$ estimate $\tilde{Y}^{(i)}$ using the method defined in step 3;
6. Estimate $M_{\tilde{Y}}$ and $\Sigma_{\tilde{Y}}$ from $\tilde{Y}^{(i)}$, $i = 1, \dots, N$ and $\tau_1(A|z)$ and $\tau_2(A|z)$ under the covariance approach (4). Observe that for a fixed $\tilde{Z} = z$, $\tau_1(A|z) = V(z) = 0$;
7. Calculate δ_{1i} and δ_{2i} for each $\tilde{Y}^{(i)}$, for $i = 1, \dots, N$;
8. Estimate Δ_1 and Δ_2 as the average of δ_{1i} and δ_{2i}, respectively over all $i = 1, \dots, N$ under the deviation approach.

Table 3.1 Misclassification measures Δ_1, Δ_2 and τ_2 for $Y, \widetilde{Y}_1$ and $\widetilde{Y}_2$.

	$\Delta_1(Y)$	$\Delta_2(Y)$	$\tau_2(Y)$	$\Delta_1(\widetilde{Y}_1)$	$\Delta_2(\widetilde{Y}_1)$	$\tau_2(\widetilde{Y}_1)$	$\Delta_1(\widetilde{Y}_2)$	$\Delta_2(\widetilde{Y}_2)$	$\tau_2(\widetilde{Y}_2)$
Simulated	8.38	7.45	74.56	8.74	6.71	67.18	7.50	5.59	55.95
Exact	8.32	7.34	73.42	8.68	6.61	66.08	7.42	5.49	54.87

This simulation method can be generalized in case when we have more than one account and Z is not fixed.

For the illustration we consider the following data example:

1. Suppose we have a vector $(Y_1, \ldots, Y_{10})$ of $p = 10$ variables with mean

$\mu = (150, 160, 140, 20, 15, 7, 2000, 2100, 2200, 550)$

and $\sigma_i^2 = \sigma^2 * \mu_i$, where $\sigma = 0.1$. Then σ_Y^2 will be

$(1.5, 1.6, 1.4, 0.2, 0.15, 0.07, 20, 21, 22, 5.5)$.

2. Suppose we have one additive account: $A|z = [Y_1 + \cdots + Y_{10} = z]$, where $z = \sum_i^{10} \mu_i$.

3. Apply (15) to estimate $\widetilde{Y}$ from Y. As before consider $\widetilde{Y}_1$ and $\widetilde{Y}_2$ for $v_{1k}^1 = 1/p$ and $v_{2k}^2 = \mu_k / \sum \mu_k$, respectively.

4. For $N = 1000$ simulate N vectors Y from $N_p(\mu, \Sigma_{\widetilde{Y}})$ distribution;

5. For each $Y^{(i)}$, $i = 1, \ldots, N$, derive $\widetilde{Y}_1^{(i)}$ and $\widetilde{Y}_2^{(i)}$;

6. Estimate $\Sigma_{\widetilde{Y}}$ as the sample variance. Following the covariance approach we will have that

$$\tau_1 = V(z) = 0 \text{ and } \tau_2 = \text{Trace}(\hat{\Sigma}_{\widetilde{Y}}) = \sum_{k=1}^{p} \hat{V}(\widetilde{Y}_k);$$

Here $\hat{V}(\widetilde{Y}_k) = \frac{1}{N-1} \sum_{i=1}^{N} (\widetilde{Y}_k^{(i)} - \mu_k)^2$ is unbiased sample variance. The simulated values of τ_2 for $Y, \widetilde{Y}_1$ and $\widetilde{Y}_2$ are given in Table 3.1.

7. Calculate δ_1 and δ_2 for each $Y, \widetilde{Y}_1$ and $\widetilde{Y}_2$, take w_k to be $1/2$ for δ_1 and $1/10$ for δ_2.

8. Estimate Δ_1 and Δ_2 as the average of δ_1 and δ_2, see the values in 'Simulated' of Table 3.1.

From (19) and (20) we also calculated the exact values for Δ_1, Δ_2 and τ_2, see 'Exact' in Table 3.1.

Observe that the values of Δ_1 are greater than the values of Δ_2. For $N = 1000$ the simulated values of the misclassification measures are quite close to the exact ones. We have carried out more simulations for different values of N and observed that the misclassification measures for this specific example are very stable and converge to the exact values. The values of Δ_2 in $\widetilde{Y}_2$ are overall smaller than in $\widetilde{Y}_1$ and Y. Both adjusted variables $\widetilde{Y}_1$ and $\widetilde{Y}_2$ have the same expected values, since here $z = \sum_{i=1}^{10} E(Y_{ji}), j = 1, 2$. It is easy to show that the variance of $\widetilde{Y}_2$ is smaller than of Y. This does not hold for $\widetilde{Y}_1$. It seems better to apply a `smart' weighted adjustment as for $\widetilde{Y}_2$.

Relative efficiency

In section 1 we defined the relative efficiency as $\tau(A_1)/\tau(A_2)$, or $\Delta_1(A)/\Delta_2(A)$ measure that facilitates a choice of adjustment method. In the example above for $z = \sum \mu_k$, we can substantiate our choice of an adjustment between $\widetilde{Y}_1$ or $\widetilde{Y}_2$ using the relative efficiency on the results in Table 1:

$$\Delta_1(\widetilde{Y}_2)/\Delta_1(\widetilde{Y}_1) = \frac{7.42}{8.68} < 1 \text{ and } \Delta_2(\widetilde{Y}_2)/\Delta_2(\widetilde{Y}_2) = \frac{5.49}{6.61} < 1.$$

4 Illustration: Multiplicative account

Consider the multiplicative account:

(Quantum Index) · (Price Index) = Value Index $\quad \Leftrightarrow \quad I_Q \cdot I_P = I_V$.

The most common application of a multiplicative account is for index figures. *The index problem* arises because, in general, for the direct estimates the theoretical constraint does not hold true: $\hat{I}_Q \cdot \hat{I}_P \neq \hat{I}_V$.

4.1 Assumptions

Our aim here is to compare alternative adjustment methods for the index problem using a scalar measures defined above. To keep focus we assume the following.

1. The true account: $A_0 = [1 \cdot 1 = 1] \Leftrightarrow B_0 = [0 + 0 = 0]$ on the log-scale
2. Unbiased initial estimates: $E(\hat{I}_Q, \hat{I}_P, \hat{I}_V) = (I_Q, I_P, I_V)$
3. Independent initial estimates: $\hat{X} = (\hat{I}_Q, \hat{I}_P, \hat{I}_V)$. Then the covariance matrix $\Sigma_{\hat{X}}$ is a diagonal matrix. Also on the log-scale for $\hat{x} = \log(\hat{X}) := (\hat{i}_Q, \hat{i}_P, \hat{i}_V)$ the covariance matrix $\Sigma_{\hat{x}}$ will be diagonal. Assume that:

$$V(\hat{I}_V) = K_1 V(\hat{I}_P) \quad \text{and} \quad V(\hat{I}_Q) = K_2 V(\hat{I}_P);$$
$$V(\hat{i}_V) = k_1 V(\hat{i}_P) \quad \text{and} \quad V(\hat{i}_Q) = k_2 V(\hat{i}_P)$$

where $\hat{i}_P = \log \hat{I}_P$, and similarly for the other indices and their logs.

4.2 Adjustment methods

Consider two different adjustments of $\hat{I}_Q \cdot \hat{I}_P \neq \hat{I}_V$.

- **Deflation adjustment** is common for e.g. in National Accounts, by which the quantum index is derived from the other two, yielding

$$A^* = [\tilde{I}_Q = \hat{I}_V/\hat{I}_P] = [\tilde{I}_Q \cdot \hat{I}_P = \hat{I}_V] \tag{21}$$

and for the account on the log-scale:

$$B^* = [\tilde{i}_Q = \hat{i}_V - \hat{i}_P] = [\tilde{i}_Q + \hat{i}_P = \hat{i}_V] \tag{22}$$

– **Joint adjustment** yielding $A = [\tilde{I}_Q \cdot \tilde{I}_P = \tilde{I}_V]$, starting from $(\hat{I}_Q, \hat{I}_P, \hat{I}_V)$ and using a Lagrangian of choice. Treating proportional adjustment in a symmetric manner leads us to consider the following minimization problem on the log-scale:

$$(\tilde{\imath}_V - \hat{\imath}_V)^2 + (\tilde{\imath}_Q - \hat{\imath}_Q)^2 + (\tilde{\imath}_P - \hat{\imath}_P)^2 \quad \text{subjected to} \quad \tilde{\imath}_V = \tilde{\imath}_Q + \tilde{\imath}_P \tag{23}$$

Using the method of Lagrange multipliers we will obtain that

$$\begin{cases} \tilde{\imath}_V = \hat{\imath}_V - \alpha = \frac{2}{3}\hat{\imath}_V + \frac{1}{3}\hat{\imath}_Q + \frac{1}{3}\hat{\imath}_P \\ \tilde{\imath}_Q = \hat{\imath}_Q + \alpha = \frac{1}{3}\hat{\imath}_V + \frac{2}{3}\hat{\imath}_Q - \frac{1}{3}\hat{\imath}_P \\ \tilde{\imath}_P = \hat{\imath}_P + \alpha = \frac{1}{3}\hat{\imath}_V - \frac{1}{3}\hat{\imath}_Q + \frac{2}{3}\hat{\imath}_P \end{cases} \quad \Leftrightarrow \quad \begin{cases} \tilde{I}_V = \hat{I}_V \exp(-\alpha) \\ \tilde{I}_Q = \hat{I}_Q \exp(\alpha) \\ \tilde{I}_P = \hat{I}_P \exp(\alpha). \end{cases} \tag{24}$$

where $3\alpha = \hat{\imath}_V - (\hat{\imath}_Q + \hat{\imath}_P)$.

4.3 Strongly permutation invariant uncertainty measure

Consider the covariance approach. For this approach the scalar measure $\tau_2 = \text{Trace}(\Sigma_{\tilde{x}})$. By applying (6) on the deflation adjustment (22), we will have $\tilde{\imath}_V = \hat{\imath}_V$, and $\tilde{\imath}_P = \hat{\imath}_P$, and $\tilde{\imath}_Q = \hat{\imath}_V - \hat{\imath}_P$. It follows that

$$\tau_2(B^*) = V(\hat{\imath}_V) + V(\hat{\imath}_P) + V(\hat{\imath}_V - \hat{\imath}_P) = 2(k_1 + 1)V(\hat{\imath}_P).$$

Next, the joint adjustment (24) yields

$$\tau_2(B) = \frac{2}{3}(k_1 + k_2 + 1)V(\hat{\imath}_P).$$

Putting these together, we have

$$\tau_2(B) \le \tau_2(B^*) \quad \Leftrightarrow \quad k_2 \le 2(k_1 + 1).$$

Unless the variance of the initial quantum index estimator is at least double the sum of the other two, the deflation adjustment is more uncertain than the joint adjustment.

Deviation approach will lead to the same results for Δ_2, when w_i's equal to 1.

4.4 Weakly permutation invariant uncertainty measure

Applying covariance approach using the weakly permutation invariant uncertainty measure τ_1 on the log-scale defined in (5), for the deflation adjustment (22) we will obtain that:

$$\tau_1(B^*) = 4V(\tilde{\imath}_V) = 4V(\hat{\imath}_V) = 4k_1 V(\hat{\imath}_P) \tag{25}$$

On the other hand for the joint adjustment method estimates:

$$\tau_1(B) = 4V(\tilde{\imath}_V) = 4V(\frac{2}{3}\hat{\imath}_V + \frac{1}{3}\hat{\imath}_Q + \frac{1}{3}\hat{\imath}_P) = \frac{4}{9}(4k_1 + k_2 + 1)V(\hat{\imath}_P) \tag{26}$$

and

$$\tau_1(B) \le \tau_1(B^*) \ \text{ if } \ k_2 \le 5k_1 - 1 \tag{27}$$

Applying the decomposition approach (9) under the deflation adjustment, where $\widetilde{Y}_1 = \tilde{I}_Q = \hat{I}_V/\hat{I}_P$, $\widetilde{Y}_2 = \tilde{I}_P = \hat{I}_P$ and $\widetilde{Z} = \tilde{I}_V = \hat{I}_V$, we obtain

$$\begin{aligned}
\tau(A^*) &= E\left[V(\frac{\hat{I}_V}{\hat{I}_P}|\hat{I}_V)\right] + E\left[V(\hat{I}_P|\hat{I}_V)\right] + 2E(\hat{I}_V) - 2E\left[E(\frac{\hat{I}_V}{\hat{I}_P}|\hat{I}_V)E(\hat{I}_P|\hat{I}_V)\right] + V(\hat{I}_V) \\
&\approx E\left[V(\frac{\hat{I}_V}{I_P^2}(\hat{I}_P - I_P)|\hat{I}_V)\right] + V(\hat{I}_P) + 2 - 2E(\hat{I}_V) + V(\hat{I}_V) \\
&= E(\hat{I}_V^2)V(\hat{I}_P) + V(\hat{I}_P) + 2 - 2 + V(\hat{I}_V) \\
&= (V(\hat{I}_V) + 1^2)V(\hat{I}_P) + V(\hat{I}_P) + V(\hat{I}_V) \\
&\approx (2 + K_1)V(\hat{I}_P).
\end{aligned}$$

Here we applied the Taylor expansion for $f(\hat{I}_P) = \frac{\hat{I}_V}{\hat{I}_P}$ at I_P and that $\hat{I}_P$ and $\hat{I}_V$ are independent of each other, and the true I_P and I_V are both assumed to be equal to 1, and typically $V(\hat{I}_V)V(\hat{I}_P) \ll V(\hat{I}_P)$.

Whereas, for the joint adjustment, we have $\tilde{I}_V = \hat{I}_V^{\frac{2}{3}}\hat{I}_Q^{\frac{1}{3}}\hat{I}_P^{\frac{1}{3}}$, $\tilde{I}_Q = \hat{I}_V^{\frac{1}{3}}\hat{I}_Q^{\frac{2}{3}}\hat{I}_P^{-\frac{1}{3}}$, and $\tilde{I}_P = \hat{I}_V^{\frac{1}{3}}\hat{I}_Q^{-\frac{1}{3}}\hat{I}_P^{\frac{2}{3}}$. Let $D = \tilde{I}_Q + \tilde{I}_P$, so that $\tau(A)$ is $E[V(D|\tilde{I}_V)] + V(\tilde{I}_V)$ by the definition leading to (9). We have

$$D = \hat{I}_V^{\frac{1}{3}}\hat{I}_Q^{\frac{1}{6}}\hat{I}_P^{\frac{1}{6}}[(\hat{I}_Q/\hat{I}_P)^{\frac{1}{2}} + (\hat{I}_Q/\hat{I}_P)^{-\frac{1}{2}}] \approx 2\hat{I}_V^{\frac{1}{3}}\hat{I}_Q^{\frac{1}{6}}\hat{I}_P^{\frac{1}{6}} = 2\tilde{I}_V^{\frac{1}{2}}$$

where we use $1/x + x \approx 2$ for $x \approx 1$, i.e. D is approximately a constant given $\tilde{I}_V$. It follows that $E[V(D|\tilde{I}_V)] \approx 0$, so that

$$\tau(A) \approx V(\tilde{I}_V) \approx \frac{1}{9}(4K_1 + K_2 + 1)V(\hat{I}_P)$$

Again, unless the initial quantum index estimator has considerably larger variance than the other two, the deflation adjustment is more uncertain than the joint adjustment, since

$$\tau(A) \leq \tau(A^*) \quad \Leftrightarrow \quad K_2 \leq 17 + 5K_1$$

5 Discussion of an application under development

In this paper we propose a new approach for measuring account uncertainty. We defined uncertainty measures as a single scalar, rather then a joint covariance or mean squared error matrix. In some cases when the account consistency problem is very complex and it is difficult or impossible to estimate the joint covariance matrix directly, we have practically no tool to measure uncertainty of a final estimates. An example of such a situation is the reconciliation and benchmarking of supply and use (SU) tables of national accounts. Often an optimization model is applied to solve this problem (see Bikker, et.al. 2013). This approach only provides a set of new estimates but not the covariance matrix or any other uncertainty measure. A scalar uncertainty measure may be very helpful.

Uncertainty measures we defined here can only deal with a single constraint or a set of constraints that are independent. For benchmarking problem of one dimensional time series we can apply these measures. However for more complex systems such as SU tables we will have to extent our approach for example for multiple set of dependent constraints. From the simple simulation example we can observe that the scalar measures we proposed here directly capture the adjustment effect as well as the relative contribution of the various input estimates to the final estimated account. Yet another challenge is further research in understanding the properties of these measures.

References

Bikker, R., Daalmans, J., and Mushkudiani, N. (2013). Benchmarking large accounting frameworks: a generalized multivariate model. J. Econ. Sys. Res., 25:4, 290-408.

Hitchcock, F.L. (1941). The distribution of a product from several sources to numerous localities. J. Math. Phys., 20:224-230.

Peña, D., and Rodríguez, R. (2000). Descriptive measures of multivariate scatter and linear dependence. J. of multivariate analyses, 85, 361-374.

Seber, G.A.F. (1984). Multivariate observations. John Wiley & Sons: New York.

Stone, J.R.N., Champerowne, D.A., and Maede, J.E. (1942). The Precision of National Income Accounting Estimates. The Review of Economic Studies, 9, 111-125.

Agafiţei, M., Gras F., Kloek W., Reis F. and Vâju S. (2015), Measuring Output Quality for Multisource Statistics in Official Statistics: Some Directions. Statistical Journal of the IAOS 31, pp. 203-211.

Appendix

Here we derive the expectation and the variance of the adjusted variable:

$$\widetilde{Y}_k = Y_k + (z - \sum_{j=1}^{p} Y_j)\nu_k \quad k = 1, \dots, p, \tag{28}$$

where ν_k, $k = 1, \dots, p$ denote relative weights and sum up to 1. Note that in (28) $\widetilde{Y}_k$ is a sum of independent normal random variables and is also a normal random variable. Its expectation and variance can be derived as follows:

$$E(\widetilde{Y}_k) = E(Y_k) + E(z - \sum_{j=1}^{p} Y_j)\nu_k = \mu_k + (z - \sum_{j=1}^{p} \mu_j)\nu_k =: M_k \tag{29}$$

and if we rewrite $\widetilde{Y}_k$ as the sum of the independent random variables:

$$V(\widetilde{Y}_k) = (1 - \nu_k)^2 V(Y_k) + \nu_k^2 \sum_{j \neq k} V(Y_j) = \nu_k^2 \sum_{j=1}^{p} V(Y_j) + V(Y_k)(1 - 2\nu_k) =: \widetilde{\sigma}_k^2. \tag{30}$$

The additive constraint in our account $A : [\widetilde{Y}_1 + \dots + \widetilde{Y}_p = z]$ implies that $\widetilde{Y}_k$'s are negatively correlated with each other.

Suppose that the variances of the unadjusted variables are proportional to their expected values:

$$V(Y_k) = \sigma_k^2 = \sigma^2 \mu_k, \text{ for } k = 1, \dots, p,$$

then

$$V(\widetilde{Y}_k) = \nu_k^2 \sigma^2 \sum_{j=1}^{p} \mu_j + \sigma^2 \mu_k - 2\nu_k \sigma^2 \mu_k. \tag{31}$$

Observe that it does not always hold that $V(\widetilde{Y}_k) < V(Y_k)$. Only if

$$\sum_{j=1}^{p} \mu_j < \frac{2}{\nu_k} \mu_k$$

the adjusted variables will have smaller variance than the unadjusted variables.

We want to compare the variances for two different adjustment methods defined in (16) and(17). Observe that:

$$V(\widetilde{Y}_{1k}) < V(Y_k) \text{ if } \sum_{j=1}^{p} \mu_j < 2p\mu_k,$$

which does not always hold true and

$$V(\widetilde{Y}_{2k}) < V(Y_k) \text{ if } \frac{\mu_k}{\sum_{j=1}^{p} \mu_j} \sum_{j=1}^{p} \mu_j < 2\mu_k,$$

which always holds true for $\mu_k \geq 0$. Next we compare the variances of the adjustments:

$$V(\widetilde{Y}_{1k}) - V(\widetilde{Y}_{2k}) = \sigma^2 \left(\frac{1}{p^2} \sum_{j=1}^{p} \mu_j - \frac{2}{p} \mu_k - \left(\frac{\mu_k}{\sum_{j=1}^{p} \mu_j} \right)^2 \sum_{j=1}^{p} \mu_j + \frac{2\mu_k}{\sum_{j=1}^{p} \mu_j} \mu_k \right)$$

$$= \sigma^2 \left(\frac{1}{p^2} \sum_{j=1}^{p} \mu_j - \frac{2}{p} \mu_k + \frac{\mu_k^2}{\sum_{j=1}^{p} \mu_j} \right) = \sigma^2 \left(\frac{\sqrt{\sum_{j=1}^{p} \mu_j}}{p} - \frac{\mu_k}{\sqrt{\sum_{j=1}^{p} \mu_j}} \right)^2 \tag{32}$$

which is never negative and hence $V(\widetilde{Y}_{1k}) \geq V(\widetilde{Y}_{2k})$.

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress: Statistics Netherlands, Grafimedia
Design: Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order
verkoop@cbs.nl
Fax +31 45 570 62 68
ISSN 1572-0314

xxxxxxxxxxxxx