

# ST45\_1 Uncertainty measures for economic accounts <sup>1</sup>

Nino Mushkudiani

Statistics Netherlands, The Hague, The Netherlands - n.mushkudiani@cbs.nl

Jeroen Pannekoek

Statistics Netherlands, The Hague, The Netherlands - j.pannekoek@cbs.nl

Li-Chun Zhang

S3RI/University of Southampton, Southampton, UK - L.Zhang@soton.ac.uk

Statistics Norway, Oslo, Norway - lcz@ssb.no

## Abstract

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.*

---

<sup>1</sup>Part of this work has been carried out as part of the ESSnet on Quality in Multisource Statistics, and has been funded by the European Commission.

# 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.

At the official statistics departments of National Accounts a large number of economic constraints are applied to produced consistent tables. National accounts departments collect data from large number of sources, included data from other departments. These sources could be surveys, administrative sources, information of branch organization, or even web-sites. When we want to give the estimate of a "variance" of a figure in a final publication of the national accounts table, we try to trace this specific figure all the way back to the source where it came from. For example consider micro-figures obtained from a survey. For these figures we have an idea about how big the survey error is. When these figures are aggregated or merged in the following steps we still can have an estimate of the error margins. In the final steps of making consistent tables national accounts specialists apply benchmarking and reconciliation on data. These include adjustments of the figures so that the certain constraints hold true. In this paper are looking for the measures that capture the "error" or "variance" that is introduced by adjusting data according to the constraints.

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  $A_0 = [\dots]$  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  $(\tilde{Y}_1, \tilde{Y}_2, \tilde{Z})$  that do satisfy the constraint, so that we have

$$A = [\tilde{Y}_1\tilde{Y}_2 = \tilde{Z}]. \quad (3)$$

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  $(\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_p, \tilde{Z})$ , denoted by  $f(\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_p, \tilde{Z})$ . 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  $\hat{X} = (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_p, \hat{Z})$  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,  $\Sigma_{\tilde{X}}$  say of  $\tilde{X} = (\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_p, \tilde{Z})$ . 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 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  $\tilde{X}$ . We shall consider two different approaches for defining a scalar measure of uncertainty: covariance approach and deviation approach. Respectively by  $\tau$  and  $\Delta$  we define scalar measures of the covariance and the deviation approach. These scalar measures would be helpful in at least two respects:

1. Let  $\dot{X} = (\dot{Y}_1, \dot{Y}_2, \dots, \dot{Y}_p, \dot{Z})$  be an alternative adjustment to  $\tilde{X}$ , 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  $\tilde{X}$ . 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  $\Delta$ .

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  $\Delta$ .

Our intuitive approach in search of such a measure was the following: For the multiplicative account when  $p = 2$  and  $Z = z$  we can construct a one-one transformation:

$$\begin{pmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \end{pmatrix} \leftrightarrow \begin{pmatrix} z \\ D \end{pmatrix} = \begin{pmatrix} \tilde{Y}_1 \tilde{Y}_2 \\ D(\tilde{Y}_1, \tilde{Y}_2) \end{pmatrix}$$

By sufficiency principle,  $(z, D)$  summarizes all the stochastic variation in  $(\tilde{Y}_1, \tilde{Y}_2)$ . Given  $z$ , the variation of  $A = [\tilde{Y}_1 \tilde{Y}_2 = z]$  is equivalent to that of  $D$ . Invariance measure of the variance of  $D$  is feasible provided symmetric  $D$ , i.e.  $D(\tilde{Y}_1, \tilde{Y}_2) = D(\tilde{Y}_2, \tilde{Y}_1)$ . For instance for  $D(\tilde{Y}_1, \tilde{Y}_2) = \tilde{Y}_1/\tilde{Y}_2 = \tilde{Y}_1^2/z$  is not symmetric, since  $D(\tilde{Y}_2, \tilde{Y}_1) = \tilde{Y}_2/\tilde{Y}_1 = z/\tilde{Y}_1^2$ . Neither is  $D(\tilde{Y}_1, \tilde{Y}_2) = D(\tilde{Y}_1)$  symmetric since  $D(\tilde{Y}_2, \tilde{Y}_1) = D(\tilde{Y}_2) = z/\tilde{Y}_1$ . Put a symmetric  $D$ :

$$D(\tilde{Y}_1, \tilde{Y}_2) = \tilde{Y}_1 + \tilde{Y}_2 = \tilde{Y}_2 + \tilde{Y}_1 = D(\tilde{Y}_2, \tilde{Y}_1).$$

And if  $z = E(\tilde{Y}_1)E(\tilde{Y}_2)$ , we have

$$V(D) = V(\tilde{Y}_1) + V(\tilde{Y}_2) + 2Cov(\tilde{Y}_1, \tilde{Y}_2) = V(\tilde{Y}_1) + V(\tilde{Y}_2)$$

In order to define measures such as expectation and variance for  $A$  we need to have a more formal definition for it.

The first, *covariance*, approach starts with the covariance matrix  $\Sigma$  of  $\tilde{X}$  as a multivariate measure of the expected deviation of  $\tilde{X}$  from its expected value. This multivariate measure is then reduced to a scalar summary. The second *deviation* approach first reduces the deviation of  $\tilde{X}$  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:

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

In other words, the appearance of the  $\tilde{Y}$ 's and  $\tilde{Z}$  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,  $[\text{Value} \cdot (1/\text{Quantum}) = \text{Price}]$  is possibly just as acceptable as  $[\text{Quantum} \cdot \text{Price} = \text{Value}]$ , whereas  $[\text{Quantum} \cdot (1/\text{Value}) = (1/\text{Price})]$  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:

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

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  $\tilde{Y}$ '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  $(\tilde{Y}_i, \tilde{Z})$  satisfy the constraint (2). For notational convenience we will sometimes combine  $\tilde{Y}$  and  $\tilde{Z}$  in one vector  $\tilde{X}$ :

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

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

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

Scalar measures for the covariance approach are defined as follows:

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

where  $\tau_1(A) = V(\sum_{k=1}^p \tilde{Y}_k + \tilde{Z}) = V(2\tilde{Z})$ , and  $V(\tilde{Z}) = \mathbf{1}^T \Sigma_{\tilde{Y}} \mathbf{1}$ , and  $\text{Trace}(\Sigma_{\tilde{X}}) = \sum_{k=1}^p V(\tilde{Y}_k) + V(\tilde{Z})$ .  $\tau_2$  is also called the total variation, see Seber (1984).

Observe that  $\tau_2$  is strongly permutation invariant, whilst  $\tau_1$  is only weakly permutation invariant. To see this consider again the two permutations of  $A$

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

Define  $\tilde{X}_1 = (\tilde{Y}_1, \dots, \tilde{Y}_p, \tilde{Z})^T$  and  $\tilde{X}_2 = ((-\tilde{Z}), \tilde{Y}_2, \dots, \tilde{Y}_p, (-\tilde{Y}_1))^T$ , with covariance matrices  $\Sigma_{\tilde{X}_1}$  and  $\Sigma_{\tilde{X}_2}$ . Then, since  $\mathbf{1}^T \Sigma_{\tilde{X}_1} \mathbf{1} = 4V(\tilde{Z})$  and  $\mathbf{1}^T \Sigma_{\tilde{X}_2} \mathbf{1} = 4V(-\tilde{Y}_1)$  we would have different values for  $\tau_1$ . This does not hold for  $\tau_2$ , which remains the same for all permutations.

Due to the constraint (2),  $\tau_1$  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  $\tau_1$  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  $\tilde{Y} = (\tilde{Y}_1, \dots, \tilde{Y}_{p^*})^T$ , 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

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

Then  $\Sigma_{\tilde{X}}$  is the variance-covariance matrix of  $\tilde{Y}$  and we will have that  $\tau_1 = \mathbf{1}^T \Sigma_{\tilde{X}} \mathbf{1} \propto \mathbf{1}^T \Sigma_{\tilde{Y}} \mathbf{1} = V(z) = 0$ . 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

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

and

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

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  $V(D) = \mathbf{1}^T \Sigma_{\tilde{Y}} \mathbf{1}$ . Recall the total variance formulae, then for  $D$  and  $\tilde{Z}$  we can write that:

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

and using the definition of covariance

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

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

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

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

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

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

### 2.2.1 Extension to multiple additive constraints

On many occasions, e.g. supply and use tables in national accounts, we will have to consider multiple additive constraints. These constraints may be connected because they have variables in common, which calls for a simultaneous approach. In such cases we do not consider a single equation such as (2) but multiple such equations. The  $Y$ -variables on the left-hand side of these equations are different, possibly overlapping, subsets of the

complete  $p$ -set of  $Y$ -variables. The  $Z$ -variables are considered to be linear functions of the  $Y$ -variables and appear only on the right-hand side of the equations. The  $Z$ -values need not be all different because there may be different subsets of  $Y$ -variables that add-up to the same  $Z$ -value. These multiple additive constraints can formally be denoted by the system of equations:

$$\mathbf{H}Y = Z, \quad (10)$$

with  $\mathbf{H}$  a  $q \times p$  matrix with elements equal to 0 and 1,  $q$  the number of constraints and  $Z$  a  $q$ -vector with  $Z$ -variables.

The vector  $\tilde{X}$  is now generalized as

$$\tilde{X} = (\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_p, \tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_q) = (\tilde{Y}, \mathbf{H}\tilde{Y}) = (\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_p, \tilde{X}_{p+1}, \tilde{X}_{p+2}, \dots, \tilde{X}_{p+q}).$$

The variance-covariance matrix  $\Sigma_{\tilde{X}}$  of this extended  $\tilde{X}$  has the following formal structure

$$\Sigma_{\tilde{X}} = \begin{pmatrix} \Sigma_{\tilde{Y}} & \Sigma_{\tilde{Y}\tilde{Z}} \\ \Sigma_{\tilde{Z}\tilde{Y}} & \Sigma_{\tilde{Z}} \end{pmatrix} = \begin{pmatrix} \Sigma_{\tilde{Y}} & \Sigma_{\tilde{Y}}\mathbf{H}^T \\ \mathbf{H}\Sigma_{\tilde{Y}} & \mathbf{H}\Sigma_{\tilde{Y}}\mathbf{H}^T \end{pmatrix}$$

Scalar measures for the covariance approach can be defined analogous to the single constraint case as

$$\begin{aligned} \tau_1(A) &= \mathbf{1}^T \Sigma_{\tilde{X}} \mathbf{1} = V\left(\sum_{k=1}^p \tilde{Y}_k + \sum_{l=1}^q \tilde{Z}_l\right) = \mathbf{1}_p^T \Sigma_{\tilde{Y}} \mathbf{1}_p + 2\mathbf{1}_q^T \mathbf{H} \Sigma_{\tilde{Y}} \mathbf{1}_p + \mathbf{1}_q^T \mathbf{H} \Sigma_{\tilde{Y}} \mathbf{H}^T \mathbf{1}_q \quad (11) \\ \tau_2(A) &= \text{Trace}(\Sigma_{\tilde{X}}) = \sum_{k=1}^p V(\tilde{Y}_k) + \sum_{l=1}^q V(\tilde{Z}_l) = \text{Trace}(\Sigma_{\tilde{Y}}) + \text{Trace}(\Sigma_{\tilde{Z}}) \end{aligned}$$

where  $\mathbf{1}_p$  and  $\mathbf{1}_q$  are a  $p$ -vector and a  $q$ -vector with ones, respectively. The single constraint special case of  $\tau_1$  is obtained by setting  $\mathbf{H}$  equal to the row-vector  $\mathbf{1}_p^T$  and noting that in this case  $\mathbf{1}_q$  consists of just a single element equal to 1.

As in the single constraint case, the measure  $\tau_1$  is not strongly permutation invariant because the covariance terms in  $\Sigma_{\tilde{X}}$  are affected by the change of sign that occurs if a variable is moved from the left hand site to the right hand site of a constraint. The measure  $\tau_2$  is again strongly permutation invariant because it depends only on the variances of the variables involved and these will not change if the variables are permuted.

The measure  $\tau_1$  can be zero if some values are fixed. For instance, if the variables in  $Z$  are all fixed we have  $V(Z) = V(\mathbf{H}\tilde{Y}) = \text{COV}(\tilde{Y}, Z) = 0$ . Now, if  $\mathbf{1}_p^T$  is contained in the row-space of  $\mathbf{H}$ ,  $\mathbf{1}_p^T \tilde{Y}$  will be equal to a linear combination of  $Z$ -values and hence has zero variance, thus  $V(\mathbf{1}_p^T \Sigma_{\tilde{Y}} \mathbf{1}_p) = 0$  and, since  $V(Z) = 0$  and  $\text{COV}(\tilde{Y}, Z) = 0$ ,  $\tau_1 = 0$ .



## 2.3 Deviation approach

Let us again consider the additive account at first. Let  $M_X = (M_1, \dots, M_p, M_{p+1} = M_Z)^T$  be the expectation of  $\tilde{X}$ , such that  $\tilde{X} - M_X$  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  $\delta = \delta(\tilde{X} - M_X)$ . Different choices are possible. Two of them are

$$\delta_1 = \sum_{k=1}^{p+1} w_k |\tilde{X}_k - M_k| \quad \text{and} \quad \delta_2 = \sum_{k=1}^{p+1} w_k (\tilde{X}_k - M_k)^2 \quad (12)$$

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

$$\Delta(A) = E(\delta). \quad (13)$$

Depending on the choice of  $w_k$ 's, the deviation approach (13) can easily be made strongly permutation invariant. For  $\delta_1$  and  $\delta_2$  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  $\Delta_2$  is closely related to  $\tau_2$  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  $\Delta_2$  to be a generalization of the covariance approach based on  $\tau_2$ .

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  $\tilde{Y}_1, \dots, \tilde{Y}_p$ , or as a classification of the amount  $z$  into  $p$ -categories. The differences  $\tilde{Y}_k - M_k$  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 ( $\hat{Y}_k$ ) to the other ( $\tilde{Y}_k$ ) (see e.g. Hitchcock 1941). The measure  $\delta_1$  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  $\delta_1$  is the amount of  $z$  that is classified differently by  $\tilde{Y}_k$  than by  $M_k$ . For  $\delta_2$  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

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

For  $m_k = E(\log(\tilde{Y}_k))$  for  $k = 1, 2$  and  $\tilde{Z}$ , we have that  $\mu(B) = [m_1 + m_2 = m_{\tilde{Z}}] = [e^{m_1} e^{m_2} = e^{m_{\tilde{Z}}}]$ . It seems natural to use the proportional deviation here, i.e.  $\tilde{Y}_1/e^{m_1}$ ,  $\tilde{Y}_2/e^{m_2}$  and  $\tilde{Z}/e^{m_{\tilde{Z}}}$ . 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

$$\delta_1 = \sum_{k=1}^3 w_k |\log(\tilde{X}_k) - m_k| \quad \text{and} \quad \delta_2 = \sum_{k=1}^3 w_k (\log(\tilde{X}_k) - m_k)^2 \quad (14)$$

and the corresponding  $\Delta(A)$  by (13). Again, this amounts to measuring the uncertainty of a multiplicative account  $A$  in terms of the corresponding  $B$  on the log scale.

## 2.4 Expected norm as scalar uncertainty measure

The covariance and deviation approaches above have two main limitations. Firstly, they summarise either the variances of the adjusted estimators or their deviations from the respective expectations. This is justified provided the adjusted accounts are unbiased, i.e. if the expected accounts are the true accounts, where all the estimates are replaced by the true values of their respective estimates. Otherwise, the bias would be missed, which is important to uncertainty evaluation.

Secondly, in both approaches the generic accounts (10) have fixed constraint matrix  $\mathbf{H}$ . While this is the case in many applications, including the supply-and-use tables for National Account, there does exist situations where the constraint matrix is random itself. For example, let  $y$  be the vector of domains of population counts in year  $t$ , and  $z$  the corresponding vector in year  $t + 1$ ; let  $\mathbf{H}$  be the gross-flow distribution matrix, whose elements are non-negative and sum to 1 in each column. One may then consider  $\mathbf{H}y = z$  as a population account with random constraint matrix  $\mathbf{H}$ .

We now introduce norm as a scalar measure, which overcomes these two limitations and unifies the covariance and deviation approaches above.

Consider an account (10), with fixed or random constraint matrix  $\mathbf{H}$ . Let  $X_{K \times 1}$  be a vector consisting of all the unknowns involved. Each account may be expressed as  $A(x) = [\mathbf{H}y = z; X = x]$ , where  $\mathbf{H}(x)y(x) = z(x)$ . Let the *account space* of  $X$  be given by

$$\Omega = \{x; \mathbf{H}(x)y(x) = z(x)\}$$

which contains all the  $x$ -vectors that satisfy the account. In particular, let  $x_0$  be the true  $x$ -vector, where  $x_0 \in \Omega$ , and  $A_0 = A(x_0) = [\mathbf{H}y = z; X = x_0]$ . Let  $\hat{X}$  be the vector of initial input estimates, where  $\hat{X} \notin \Omega$ , as long as  $\mathbf{H}(\hat{X})y(\hat{X}) \neq z(\hat{X})$ . Let  $\tilde{X}$  be the vector of adjusted estimates, where  $\tilde{X} \in \Omega$  and  $\mathbf{H}(\tilde{X})y(\tilde{X}) = z(\tilde{X})$ , with the corresponding adjusted account  $\tilde{A} = A(\tilde{X}) = [\mathbf{H}y = z; X = \tilde{X}]$ .

Thus, the account space  $\Omega$  is a subset of the vector space of  $X$ ; the true vector  $x_0$  is a point in  $\Omega$ ; the initial vector  $\hat{X}$  is generally a point outside of  $\Omega$ ; the adjusted vector  $\tilde{X}$  is a point in  $\Omega$ , which is typically obtained from  $\hat{X}$  via some minimum changes subjected to the constraints (10). The difference  $\hat{X} - x_0$  is the error of the initial vector of estimates, the difference  $\tilde{X} - \hat{X}$  is required in order to move from  $\hat{X}$  outside  $\Omega$  to  $\tilde{X}$  inside  $\Omega$ , and the

final difference  $\tilde{X} - x_0$  is the error of the adjusted vector of estimates.

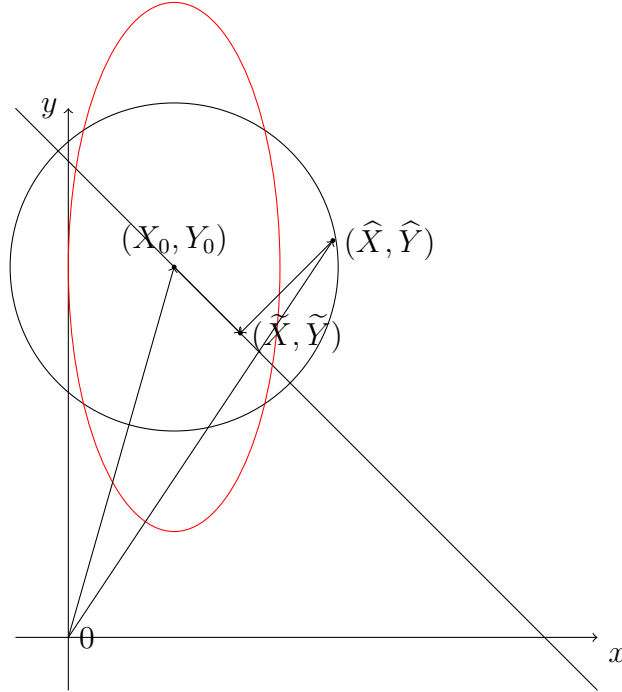
One may envisage the initial and final vectors of error, i.e.  $\hat{X} - x_0$  and  $\tilde{X} - x_0$ , as two vectors relocated to  $x_0$  as the origin, i.e. one may centre the vector space of  $X$  at  $x_0$ . The measure of the size (or length) of each error vector is then just a question of norm. The sizes of the errors are unknown for the observed  $\hat{X}$  and  $\tilde{X}$  because  $x_0$  is unknown. However, it is possible to measure the average size over the distribution of  $\hat{X}$  and  $\tilde{X}$ , respectively. For instance,  $E(\hat{X} - x_0)$  is the bias of  $\hat{X}$  and  $E(\hat{X} - x_0)^2$  its mean squared error (MSE). Such averaging measures are referred to as *statistical* uncertainty measures. In particular, for a random vector  $v_{K \times 1}$ , consider the weighted Euclidean norm (WEN) given by

$$\|v\|_{wen} = \sum_{i=1}^K w_i v_i^2 \quad (15)$$

where  $w_i \geq 0$ , and the corresponding statistical uncertainty measure given by

$$\delta_v = E(\|v\|_{wen}) \quad (16)$$

Notice that for  $v = \tilde{X} - x_0$ ,  $\delta_v$  is just the weighted MSE of  $\tilde{X}$ , which does not require  $\tilde{X}$  to be unbiased for  $x_0$ . Moreover, setting  $w_i \equiv 1$ , we recover  $\tau_2$  by the covariance approach, where  $\mathbf{H}$  is assumed to be fixed; whereas  $\|v\|_{wen}$  reduces to  $\delta_2$  by (12) in the simple case of one additive constraint. Finally, the WEN (15) is easily seen to be strongly permutation invariant, provided component-specific weights  $w_i$ . Thus, using the expected norm (16) as a scalar statistical uncertainty measure of accounting equations is a general approach that unifies the covariance and deviation approaches.



### 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  $\Sigma_Y$  is a diagonal matrix with the diagonal values  $\sigma_k^2 = V(Y_k)$ , for  $k = 1, \dots, p$ . Suppose also that  $\tilde{Z}$  is fixed,  $\tilde{Z} = z$  and the variances of the unadjusted variables  $\sigma_k^2$ 's are proportional to the corresponding expected values,  $\mu_k$ 's:

$$\sigma_k^2 = \sigma^2 \mu_k, \quad \text{for } k = 1, \dots, p. \quad (17)$$

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  $\tilde{Y}$  satisfies the following constraint:

$$A = [\tilde{Y}_1 + \dots + \tilde{Y}_p = z]. \quad (18)$$

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

$$\tilde{Y}_k = Y_k + (z - \sum_{j=1}^p Y_j) \nu_k \quad k = 1, \dots, p, \quad (19)$$

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

#### 3.2 Adjustment methods

It is not obvious what the best choice is for the adjustment weights in (19). We consider two different adjusted vectors  $\tilde{Y}_1$  and  $\tilde{Y}_2$  for  $\nu_{1k} = 1/p$  and  $\nu_{2k} = \mu_k / \sum_{j=1}^p \mu_j$ :

$$\tilde{Y}_{1k} = Y_k + \frac{1}{p}(z - \sum_{j=1}^p Y_j) \quad k = 1, \dots, p, \quad (20)$$

and

$$\tilde{Y}_{2k} = Y_k + \frac{\mu_k}{\sum_{j=1}^p \mu_j} (z - \sum_{j=1}^p Y_j) \quad k = 1, \dots, p. \quad (21)$$

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  $\tilde{Y}$  defined in (19), and for the special cases  $\tilde{Y}_1$  and  $\tilde{Y}_2$  we can derive  $\Delta_1(A|z)$  and  $\Delta_2(A|z)$  analytically. With the notation  $(\cdot|z)$  we want to indicate that  $\tilde{Z} = z$  is fixed. First we consider  $\Delta_1(A|z)$ . Following the deviation approach from (12) and (13) we have that:

$$\Delta_1(A|z) = E(\delta_{1|z}) = \sum_{k=1}^{p+1} E(w_k |\tilde{X}_k - M_k|) = \sum_{k=1}^p E(w_k |\tilde{Y}_k - M_k|), \quad (22)$$

since  $E(z) = z$ . Here  $M_k = E(\tilde{Y}_k)$ , for  $k = 1, \dots, p$  are defined in the Appendix. Note that the random variables  $\tilde{Y}_k - M_k$ ,  $k = 1, \dots, p$  are negatively correlated with each other and have normal distributions  $N(0, \tilde{\sigma}_k^2)$ , where  $\tilde{\sigma}_k^2$  is the variance of  $\tilde{Y}_k$  defined in (35) in the Appendix. Then  $|\tilde{Y}_k - M_k|$ , for  $k = 1, \dots, p$  have half-normal distributions with expectation equal to  $\tilde{\sigma}_k \sqrt{2/\pi}$  and we will have that:

$$\Delta_1(A|z) = \sum_{k=1}^p w_k \tilde{\sigma}_k \sqrt{2/\pi}. \quad (23)$$

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

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

Applying results from (23) and (36) in the Appendix we can derive the uncertainty measures  $\Delta_1(A_1|z)$  and  $\Delta_1(A_2|z)$  for the adjustments  $\tilde{Y}_{1k}$  and  $\tilde{Y}_{2k}$  defined in (20) and (21):

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

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:

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

For the second adjustment method the uncertainty measure  $\Delta_1$  is smaller and therefore we would choose the second adjustment.

For the uncertainty measure  $\Delta_2(A|z)$  by again applying (12) and (13) we will have that:

$$\Delta_2(A|z) = E(\delta_{2|z}) = \sum_{k=1}^p E(w_k(\tilde{Y}_k - M_k)^2) = \sum_{k=1}^p w_k V(\tilde{Y}_k) = \sum_{k=1}^p w_k \tilde{\sigma}_k^2. \quad (24)$$

As above, consider adjustments defined in (20) and (21) with  $\sigma_k^2 = \sigma^2 \mu_k$ . Next we apply results in (37) in the Appendix:

$$\begin{aligned} \Delta_2(A_1|z) - \Delta_2(A_2|z) &= \sum_{k=1}^p w_k \tilde{\sigma}_{1k}^2 - \sum_{k=1}^p w_k \tilde{\sigma}_{2k}^2 \\ &= \sum_{k=1}^p w_k (V(\tilde{Y}_{1k}) - V(\tilde{Y}_{2k})) = \sum_{k=1}^p w_k \sigma^2 \left( \frac{\sqrt{\sum_{j=1}^p \mu_j}}{p} - \frac{\mu_k}{\sqrt{\sum_{j=1}^p \mu_j}} \right)^2. \end{aligned}$$

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 (MC) 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  $\Sigma_Y$  is a diagonal matrix with the diagonal values  $\sigma_k^2 = V(Y_k)$  and consider the same constraint as above:

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

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  $\delta_{1i}$  and  $\delta_{2i}$  for each  $\tilde{Y}^{(i)}$ , for  $i = 1, \dots, N$ ;
8. Estimate  $\Delta_1$  and  $\Delta_2$  as the average of  $\delta_{1i}$  and  $\delta_{2i}$ , respectively over all  $i = 1, \dots, N$  under the deviation approach.

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  $(Y_1, \dots, Y_{10})$  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  $\sigma_Y^2$  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:  $A|z = [Y_1 + \dots + Y_{10} = z]$ , where  $z = \sum_{i=1}^{10} \mu_i$ .
3. Apply (19) to estimate  $\tilde{Y}$  from  $Y$ . As before consider  $\tilde{Y}_1$  and  $\tilde{Y}_2$  for  $\nu_{1k}^1 = 1/p$  and  $\nu_{2k}^2 = \mu_k / \sum \mu_k$ , respectively.
4. For  $N = 1000$  simulate  $N$  vectors  $Y$  from  $N_p(\mu, \Sigma_Y)$  distribution;
5. For each  $Y^{(i)}$ ,  $i = 1, \dots, N$ , derive  $\tilde{Y}_1^{(i)}$  and  $\tilde{Y}_2^{(i)}$ ;
6. Estimate  $\Sigma_{\tilde{Y}}$  as the sample variance. Following the covariance approach we will have that

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

Here  $\hat{V}(\tilde{Y}_k) = \frac{1}{N-1} \sum_{i=1}^N (\tilde{Y}_k^{(i)} - \mu_k)^2$  is unbiased sample variance. The simulated values of  $\tau_2$  for  $Y$ ,  $\tilde{Y}_1$  and  $\tilde{Y}_2$  are given in Table 2.

7. Calculate  $\delta_1$  and  $\delta_2$  for each  $Y$ ,  $\tilde{Y}_1$  and  $\tilde{Y}_2$ , take  $w_k$  to be  $1/2$  for  $\delta_1$  and  $1/10$  for  $\delta_2$ .
8. Estimate  $\Delta_1$  and  $\Delta_2$  as the average of  $\delta_1$  and  $\delta_2$ , see the values in 'Simulated' of Table 2.

Table 1: Uncertainty measures  $\Delta_1$ ,  $\Delta_2$  and  $\tau_2$  for  $Y$ ,  $\tilde{Y}_1$  and  $\tilde{Y}_2$ .

	$\Delta_1(Y)$	$\Delta_2(Y)$	$\tau_2(Y)$	$\Delta_1(\tilde{Y}_1)$	$\Delta_2(\tilde{Y}_1)$	$\tau_2(\tilde{Y}_1)$	$\Delta_1(\tilde{Y}_2)$	$\Delta_2(\tilde{Y}_2)$	$\tau_2(\tilde{Y}_2)$
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

From (23) and (24) we also calculated the exact values for  $\Delta_1$ ,  $\Delta_2$  and  $\tau_2$ , see 'Exact' in Table 2.

Observe that the values of  $\Delta_1$  are greater than the values of  $\Delta_2$ . For  $N = 1000$  the simulated values of the uncertainty measures are quite close to the exact ones. We have carried out more simulations for different values of  $N$  and observed that the uncertainty measures for this specific example are very stable and converge to the exact values. The values of  $\Delta_2$  in  $\tilde{Y}_2$  are overall smaller than in  $\tilde{Y}_1$  and  $Y$ . Both adjusted variables  $\tilde{Y}_1$  and  $\tilde{Y}_2$  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  $\tilde{Y}_2$  is smaller than of  $Y$ . This does not hold for  $\tilde{Y}_1$ . It seems better to apply a 'smart' weighted adjustment as for  $\tilde{Y}_2$ .

### 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  $\tilde{Y}_1$  or  $\tilde{Y}_2$  using the relative efficiency on the results in Table 1:

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

## 4 Empirical example

In this section we apply the MC simulation approach described in section 3.4 on a empirical data example. In Daalmans and Di Fonzo (2014) this data set is used for comparing different reconciliation techniques. It contains figures of EU Quarterly Sector Accounts (EUQSA), flow variables only. Data set consist of multiple time series. These time series are assumed to be independent from each other and should be adjusted by using benchmark values. We have high frequency series  $Y$  and low frequency series  $Z$ , we have 28 quarterly values and 7 yearly values for 61 independent series. All yearly values are the benchmarks for the quarterly values. Di Fonzo and Marini (2011) describe this data in more detail.

We will follow the steps of the simulation approach defined above:

1. For  $Y$  define or estimate its expectation and variance. For each data set we have  $Y_{ijk}$  unadjusted variable, where the index  $i$  ( $i = 1, \dots, p$ ) stands for quarters or months,  $j$  ( $j = 1, \dots, J$ ) for years and  $k$  ( $k = 1, \dots, K$ ) for time series. For our data set



Table 2: Uncertainty measures

	$\tau_2(\tilde{Y})$	$\Delta_1(\tilde{Y})$	$\Delta_2(\tilde{Y})$
Denton-Cholette additive	25930.7	260.8	25904.8
Denton-Cholette proportional	25906.1	260.0	25880.2
Chow-Lin	25589.1	255.2	25563.5
Fernandez	25711.4	258.1	25685.7

$p = 4$ . Since we have only the unadjusted time series and no additional information is available, we set the expected values to be equal to the unadjusted values and define the variance as in the previous section by  $\sigma_{ijk}^2 = \sigma^2 \times Y_{ijk}$ , for some fix  $\sigma^2$ .

- Identify constraint that should be fulfilled; For these data constraint in (2) will become as follows:

$$Y_{1jk} + \dots + Y_{ijk} + \dots + Y_{pjk} = z_{jk}, \text{ for } k = 1, \dots, K, \quad (25)$$

$z_{jk}$  are the annual benchmark values and are fixed.

- Define a method for estimating  $\tilde{Y}$  from  $Y$ , satisfying the account  $A$ ; we will apply three different adjustment methods: Denton-Cholette (see Dagum and Cholette 2006), Fernandez and Chow-Lin (see Chow and Lin (1971)).
- For large  $N$ , simulate  $N$  matrices  $Y_{ijk}^{(l)} \sim N_p(Y_{ijk}, \sigma^2 \times Y_{ijk})$ ,  $l = 1, \dots, N$ .
- For each  $Y_{ijk}^{(l)}$  estimate  $\tilde{Y}_{ijk}^{(l)}$  using the methods defined in step 3; Here we carry out the benchmarking by using the package "tempdisagg" of R program, see Sax and Steiner (2003).
- 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).

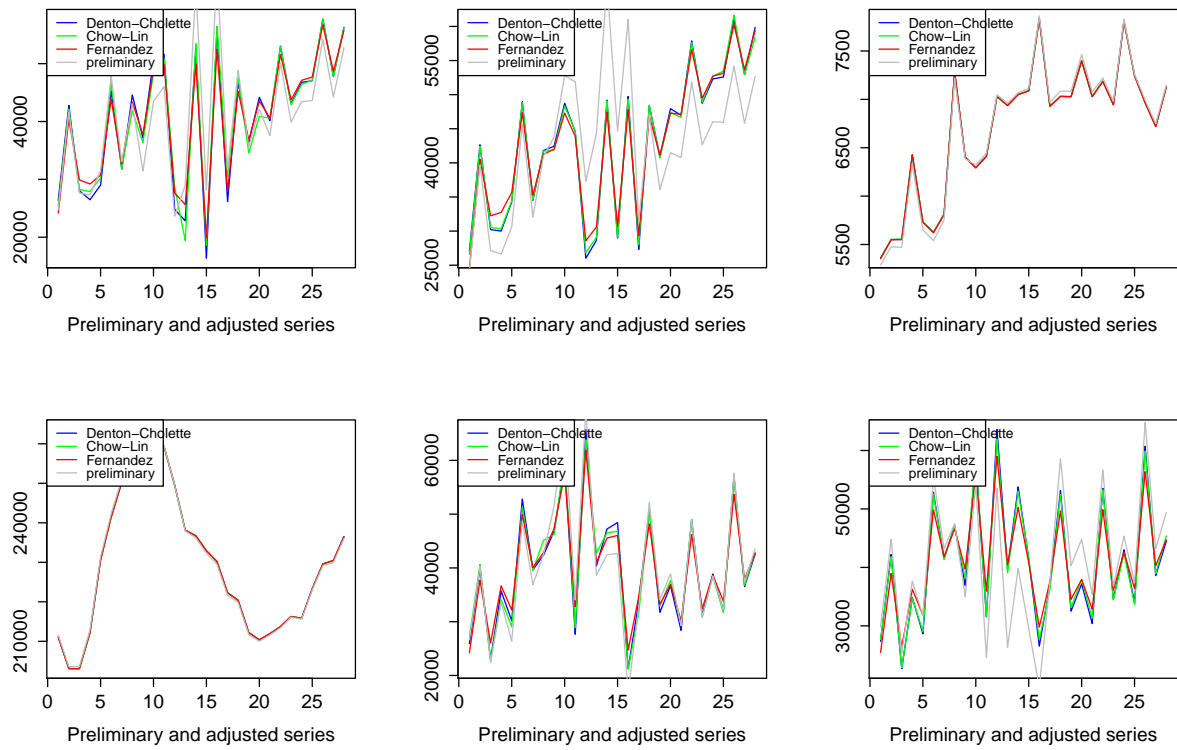
Figure 1 shows the preliminary and benchmarked figures for six different series. For two series the adjustment methods do not make any difference. For the other four series it is not obvious from these plots which method to choose. In Table 2 we give the average of the estimates uncertainty measures over all 61 series. Observe that for this data example Chow-Lin method has smallest figures for all measures. Even though for this case the difference is rather small.

## 5 Illustration: Multiplicative account

Consider the multiplicative account:

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

Figure 1: Preliminary and adjusted series



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:  $\widehat{I}_Q \cdot \widehat{I}_P \neq \widehat{I}_V$ .

## 5.1 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:  $E(\widehat{I}_Q, \widehat{I}_P, \widehat{I}_V) = (I_Q, I_P, I_V)$
3. Independent initial estimates:  $\widehat{X} = (\widehat{I}_Q, \widehat{I}_P, \widehat{I}_V)$ . Then the covariance matrix  $\Sigma_{\widehat{X}}$  is a diagonal matrix. Also on the log-scale for  $\widehat{x} = \log(\widehat{X}) := (\widehat{i}_Q, \widehat{i}_P, \widehat{i}_V)$  the covariance matrix  $\Sigma_{\widehat{x}}$  will be diagonal. Assume that:

$$\begin{aligned} V(\widehat{I}_V) &= K_1 V(\widehat{I}_P) & \text{and} & & V(\widehat{I}_Q) &= K_2 V(\widehat{I}_P); \\ V(\widehat{i}_V) &= k_1 V(\widehat{i}_P) & \text{and} & & V(\widehat{i}_Q) &= k_2 V(\widehat{i}_P) \end{aligned}$$

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

## 5.2 Adjustment methods

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

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

$$A^* = [\widetilde{I}_Q = \widehat{I}_V / \widehat{I}_P] = [\widetilde{I}_Q \cdot \widehat{I}_P = \widehat{I}_V] \quad (26)$$

and for the account on the log-scale:

$$B^* = [\widetilde{i}_Q = \widehat{i}_V - \widehat{i}_P] = [\widetilde{i}_Q + \widehat{i}_P = \widehat{i}_V] \quad (27)$$

- **Joint adjustment** yielding  $A = [\widetilde{I}_Q \cdot \widetilde{I}_P = \widetilde{I}_V]$ , starting from  $(\widehat{I}_Q, \widehat{I}_P, \widehat{I}_V)$  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:

$$(\widetilde{i}_V - \widehat{i}_V)^2 + (\widetilde{i}_Q - \widehat{i}_Q)^2 + (\widetilde{i}_P - \widehat{i}_P)^2 \quad \text{subjected to} \quad \widetilde{i}_V = \widetilde{i}_Q + \widetilde{i}_P \quad (28)$$

Using the method of Lagrange multipliers we will obtain that

$$\begin{cases} \tilde{i}_V = \hat{i}_V - \alpha = \frac{2}{3}\hat{i}_V + \frac{1}{3}\hat{i}_Q + \frac{1}{3}\hat{i}_P \\ \tilde{i}_Q = \hat{i}_Q + \alpha = \frac{1}{3}\hat{i}_V + \frac{2}{3}\hat{i}_Q - \frac{1}{3}\hat{i}_P \\ \tilde{i}_P = \hat{i}_P + \alpha = \frac{1}{3}\hat{i}_V - \frac{1}{3}\hat{i}_Q + \frac{2}{3}\hat{i}_P \end{cases} \Leftrightarrow \begin{cases} \tilde{I}_V = \hat{I}_V \exp(-\alpha) \\ \tilde{I}_Q = \hat{I}_Q \exp(\alpha) \\ \tilde{I}_P = \hat{I}_P \exp(\alpha). \end{cases} \quad (29)$$

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

### 5.3 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 (27), we will have  $\tilde{i}_V = \hat{i}_V$ , and  $\tilde{i}_P = \hat{i}_P$ , and  $\tilde{i}_Q = \hat{i}_V - \hat{i}_P$ . It follows that

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

Next, the joint adjustment (29) yields

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

Putting these together, we have

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

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  $\Delta_2$ , when  $w_i$ 's equal to 1.

### 5.4 Weakly permutation invariant uncertainty measure

Applying covariance approach using the weakly permutation invariant uncertainty measure  $\tau_1$  on the log-scale defined in (5), for the deflation adjustment (27) we will obtain that:

$$\tau_1(B^*) = 4V(\tilde{i}_V) = 4V(\hat{i}_V) = 4k_1V(\hat{i}_P) \quad (30)$$

On the other hand for the joint adjustment method estimates:

$$\tau_1(B) = 4V(\tilde{i}_V) = 4V\left(\frac{2}{3}\hat{i}_V + \frac{1}{3}\hat{i}_Q + \frac{1}{3}\hat{i}_P\right) = \frac{4}{9}(4k_1 + k_2 + 1)V(\hat{i}_P) \quad (31)$$

and

$$\tau_1(B) \leq \tau_1(B^*) \text{ if } k_2 \leq 5k_1 - 1 \quad (32)$$

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

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

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

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

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

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

## 6 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 than 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 extend 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

- Agafitei, M., Gras F., Kloek W., Reis F. and Văju S. (2015), Measuring Output Quality for Multisource Statistics in Official Statistics: Some Directions. *Statistical Journal of the IAOS* 31, pp. 203-211.
- Bikker, R., Daalmans, J., and Mushkudiani, N. (2013). Benchmarking large accounting frameworks: a generalized multivariate model. *J. Econ. Sys. Res.*, 25:4, 290-408.
- Daalmans, J., and Di Fonzo, T. (2014). Denton PFD and GRP benchmarking are friends. An empirical evaluation on Dutch Supply and Use Tables. Technical report.
- Di Fonzo, T., and M. Marini (2011). Simultaneous and two-step reconciliation of systems of time series: methodological and practical issues. *J. of the Royal statistical Society C*, 60, 143-164.
- 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.
- Sax, C., and Steiner, P. (2003). Temporal Disaggregation of Time Series, *The R Journal* Vol. 5/2, ISSN 2073-4859.
- 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.

# Appendix

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

$$\tilde{Y}_k = Y_k + (z - \sum_{j=1}^p Y_j) \nu_k \quad k = 1, \dots, p, \quad (33)$$

where  $\nu_k$ ,  $k = 1, \dots, p$  denote relative weights and sum up to 1. Note that in (33)  $\tilde{Y}_k$  is a sum of independent normal random variables and is also a normal random variable. Its expectation and variance can be derived as follows:

$$E(\tilde{Y}_k) = E(Y_k) + E(z - \sum_{j=1}^p Y_j) \nu_k = \mu_k + (z - \sum_{j=1}^p \mu_j) \nu_k =: M_k \quad (34)$$

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

$$V(\tilde{Y}_k) = (1 - \nu_k)^2 V(Y_k) + \nu_k^2 \sum_{j \neq k} V(Y_j) = \nu_k^2 \sum_{j=1}^p V(Y_j) + V(Y_k)(1 - 2\nu_k) =: \tilde{\sigma}_k^2. \quad (35)$$

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

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

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

then

$$V(\tilde{Y}_k) = \nu_k^2 \sigma^2 \sum_{j=1}^p \mu_j + \sigma^2 \mu_k - 2\nu_k \sigma^2 \mu_k. \quad (36)$$

Observe that it does not always hold that  $V(\tilde{Y}_k) < V(Y_k)$ . 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 (20) and (21). Observe that:

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

which does not always hold true and

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

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

$$\begin{aligned}
V(\tilde{Y}_{1k}) - V(\tilde{Y}_{2k}) &= \sigma^2 \left( \frac{1}{p^2} \sum_{j=1}^p \mu_j - \frac{2}{p} \mu_k - \left( \frac{\mu_k}{\sum_{j=1}^p \mu_j} \right)^2 \sum_{j=1}^p \mu_j + \frac{2\mu_k}{\sum_{j=1}^p \mu_j} \mu_k \right) \\
&= \sigma^2 \left( \frac{1}{p^2} \sum_{j=1}^p \mu_j - \frac{2}{p} \mu_k + \frac{\mu_k^2}{\sum_{j=1}^p \mu_j} \right) = \sigma^2 \left( \frac{\sqrt{\sum_{j=1}^p \mu_j}}{p} - \frac{\mu_k}{\sqrt{\sum_{j=1}^p \mu_j}} \right)^2 \quad (37)
\end{aligned}$$

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