Taylor & Francis
Taylor & Francis Group

Uncertainty measures for economic accounts

ABSTRACT. The problem of adjusting economic or social accounts can be quite complex when large accounting equation systems are considered. This is especially true if they must fulfill predefined, known functional relationships. For such complex systems, evaluating the accuracy of the estimates after the adjustment is difficult since they are defined by unadjusted initial estimates, the accounting equations and the adjustment method. In this paper, we consider such systems as a single entity and develop scalar uncertainty measures that capture the adjustment effect as well as the relative contribution of the various input estimates to the final estimated account. The scalar measures are based on the first two moments of the joint distribution of the underlying true accounting system without requiring specification of the distribution in full. Scalar measures can help to effectively communicate to the users the relevant uncertainty of disseminated macro-economic accounts, and can assist the producer in choosing and improving adjustment method and input estimators. The proposed approach will be illustrated both analytically and by simulation. Applications to supply and use tables and to time series data will be presented.

Keywords: macro account; input estimates adjustment; scalar uncertainty measure; permutation invariance.

1. Introduction

Macro-economic indicators are generally more meaningful when viewed in relation to other variables. Often, such variables are presented together to emphasize and depict their relationship, and can even be the main feature of the presentation. This is especially the case for variables that form equations. For instance, the total labor cost for a sector of the economy can be given as the product of the number of full-time equivalent employees and the average wage. Together these variables represent the structure of the labor cost. In practice, however, it is often the case that each is estimated separately and independently by different sources and means. Decompositions of GDP (gross domestic product) are another prime example of a macro-economic account. It is based on the sum of consumption, investment, government spending, and net exports should be equal to to the sum of consumption, private sector savings and net taxes, see for e.g. Miller and Blair (2009). Similarly, the total value of the production of an economy can be decomposed as the sum of the production values of a number of categories of goods and services. Each data point in these accounts is an aggregated estimate, obtained via a meticulous process that combines data from diverse sources, including sample surveys, administrative data, business organizations, web-sites, etc. Here we consider economic and accounting estimates that are defined at any level of aggregation. The variables in these estimated accounts are estimates of subpopulation figures, e.g. totals or means, and will have different properties than micro-level accounts, consisting of measurements of e.g. individual firms or businesses. The initial estimates have inherent bias or uncertainty for instance due to sampling or non-response and generally

must be adjusted to satisfy the balancing of accounts, which is typically enabled via a set of constraints. The task of the national statistical office is to produce consistent figures, estimates that satisfy required macro-economic or other accounts and are close to the initial figures. Adjustment methods are applied to obtain these consistent estimates. Nice overviews on adjustment methods applied within the systems of national accounts can be found in for e.g. Buono et.al. (2018) or International Monetary Fund (2017).

In this paper, we define scalar measures of the uncertainty of the obtained consistent estimates of aggregates in systems of additive or multiplicative accounting equations. When a number of estimated aggregates must satisfy some accounting equations, we record estimates before the accounting adjustment and after the adjustment. These estimates have variances that change due to the adjustment. The revised variances depend on the initial variances, accounting constraints and adjustment methods. With the scalar measure we aim to capture the uncertainty of the estimates after the adjustment. In order to define the scalar measure, we assume a hypothetical fixed true account exists. The true account is defined by the hypothetical true values of the accounting system and the accounting constraints. In this accounting system, true values satisfy the accounting constraints. In practice, we have available only the initial estimates. These estimates do not satisfy the accounting constraints due to measurement, non-response, sampling and other errors. After an adjustment method is applied, one obtains the adjusted values that satisfy the accounting constraints. Within this framework, different realizations of initial estimates, different accounting constraints and/or different adjustment methods will lead to different adjusted values. We introduce a scalar measure for the uncertainty in the final estimates. This scalar measure depends on all three attributes of the framework: the uncertainty of the initial estimates, accounting constraints and the adjustment method. This measure could be helpful (i) to capture the effect of the uncertainty in the input data on the final estimates, and (ii) in making a choice between different adjustment methods. For the latter purpose, one may use the relative efficiency of two adjustment methods, which is the ratio of the corresponding scalar measures. Another advantage of our scalar measure is that it is easy to interpret as the expectation of a chosen norm of the error vector. A third advantage is that for large accounting system the scalar measures of its sub-systems can be easily combined to yield a scalar measure of the whole system. In contrast, one cannot directly combine variance-covariance matrices of sub-systems into that of the whole system, unless they can be obtained independently of each other.

Manski (2014) argues for the importance of communicating uncertainty in official statistics. Three types of uncertainty are distinguished, exemplified and discussed: *Transitory* statistical uncertainty arises because data collection takes time, e.g. the first few rounds of GDP revision due to incomplete data. *Permanent* statistical uncertainty arises from incompleteness or inadequacy of data collection that does not diminish with time. Examples are sampling and non-sampling errors such as survey non-response. *Conceptual* uncertainty arises from incomplete understanding of the information that official statistics provide about well-defined economic concepts or from lack of clarity in the concepts themselves. It concerns the interpretation of statistics rather than their magnitudes. Seasonal adjustment is used as an example.

As pointed out by Manski (2014), permanent statistical uncertainty traditionally has been a preoccupation of survey statisticians. Many studies have emerged concerning uncertainty assessment of data from administrative sources. See for example Oberski et al. (2017), Zhang (2012), Wallgren and Wallgren (2014). But rarely is the issue addressed, regarding the uncertainty of a set of macro-economic figures that relate to each other via constraints.

For macro-economic accounts considered in this paper, several approaches exist regarding the quantification of the uncertainty in the input estimates, how this knowledge may be incorporated into the adjustment method, and how to evaluate the resulting integrated outputs. A typical problem in this context is that limited information may be available about the data sources and the original figures. In Mushkudiani et al. (2018), uncertainty measures of the original figures are

defined for a practical example of integration of supply and use tables in national accounts. The uncertainty of the input estimates are summarised in the reliability weights for integration, which are based on the knowledge of National Accounts specialists. The choice of weights is relative for different figures, and all series are classified in discrete classes of reliability, ranging from "least reliable" to "most reliable". A similar ranking of classes of data quality is described by Rodrigues (2014). Moreover, Rodrigues (2014) uses a Bayesian approach to balancing, related to earlier work of Golan and Vogel (2000), in which the prior is a multivariate random vector of economic data items that may be unbalanced and the posterior is obtained by minimising a loss-function (cross-entropy) that measures the difference between prior and posterior distributions, subject to the balance constraints. It is shown that several conventional balancing methods can be obtained as special cases, including proportional fitting, weighted or generalised least squares with different kinds of weights.

Rodrigues (2016) stresses the input uncertainty for macroeconomic reconciliation processes, and proposes to approximate the probability distribution of the input items only based on estimates of (up to) the first two moments. When the items are constrained by an additive constraint, the paper provides approximations to the correlations between the data items. The probability distribution is derived using the maximum entropy principle (MEP) for two cases: (I) when only an estimate of the expected value is available and (II) when also an estimate of the variance is available. For case (I) it is shown that the MEP induces an exponential distribution with standard deviation equal to its expected value, and in case (II) it is a truncated normal distribution.

The normal distribution provides often a starting point in the literature for calculating the uncertainty of the final integrated figures. Magnus et al. (2000) define a Bayesian setting for the constrained optimization problem, including linearized ratio constraints, and provide closed-form estimates of the mean and variance of the posterior under the normality assumption. Under the same setting, Boonstra et al. (2011) emphasize the inclusion of inequality constraints within the linear data integration framework. The method is based on a normal approximation to the truncated multivariate normal distribution. The approximate moments method with Kullback-Leibler divergence is applied to obtain closed-form approximation to the mean and variance of the estimates, where the approximated variances are adopted as the uncertainty measures.

Golan (2018) proposes a general framework for the constrained optimization approach in the present context, when only partial, uncertain or noisy information is available. Also the constraints can be uncertain themselves. As more than a single inference exists that is logically consistent with the available information, an inferential method and a decision criterion are selected. The author applies the maximum entropy optimization procedure for solving these under-determined problems. In this framework, estimates of the uncertainty can be found numerically; see also Temursho (2018).

In this paper the final disseminated figures are considered to be the result of the specific realisation of the input data as well as the application of the constraints and the specific adjustment method. For the measurement of the effect of the application of constraints alone, several scalar summary measures of the discrepancy between the input estimates and the adjusted estimates have been proposed, see Jackson and Murray (2004) for a review and applications.

Often a covariance matrix of the adjusted figures is also considered. This matrix can be derived for some adjustments when a covariance matrix for the original figures is available. But when inequality constraints are added to the adjustment it becomes difficult to construct the covariance matrix of the adjusted figures for large systems, see e.g. Boonstra et al. (2011). Scalar measures that merely measure the closeness of adjusted figures to the original figures are easy to derive and are often used, see e.g. Temursho (2018), Mushkudiani et al. (2018).

Our perspective in this paper fits the tenet that "Statistical agencies could mitigate misinterpretation of official statistics if they were to measure uncertainty and report it in their news releases and technical publications" Manski (2014) p. 1, according to which statistical uncertainty should

have explicit and meaningful bearings on macro economic accounts. Rather than presenting numerous uncertainty measures for each of the final estimates, which could easily hamper the communication, we instead propose scalar measures for the set of disseminated macro-economic accounts, in which all the relevant estimates are considered together, all at once. To illustrate the working of these measures we present several scalar measures and their application. The trace is one such measure. It is often used in the literature and has the nice property of permutation invariance. The measures we consider can be applied to multiplicative accounts by taking the logarithm.

2.   Uncertainty of macro accounts

We consider the logical or theoretical constraints that can be expressed as multiplicative or additive [1] *accounting equations*, which can be given, respectively, as

$$y_1 y_2 \cdots y_p = z \tag{1}$$

and

$$y_1 + \ldots + y_i + \ldots + y_p = z. \tag{2}$$

Note that for the multiplicative constraint (1) we consider only the case $p = 2$ in the rest of this paper. The approach we develop allows $p > 2$ for the multiplicative constraint as well, but the case of $p = 2$, as in the examples above of the total labor cost as the product of the number of full-time equivalent employees and the average wage, is the most common with multiplicative accounting equations. Moreover, we assume that the components of a multiplicative accounting equation are all positive by definition.

Let the true values be denoted by $(y_{01}, ..., y_{0p}, z_0)$. Then, for a multiplicative accounting equation with $p = 2$, we denote the *true account* $\mathbb{A}_0$ by

$$\mathbb{A}_0 = [y_{01} y_{02} = z_0],$$

which satisfies (1) by definition. We denote an additive accounting equation similarly. The notation $\mathbb{A}_0 = [\cdots]$ is introduced to emphasize that the account is conceived as a *single* entity. Denote by $(\hat{y}_1, \hat{y}_2, \hat{z})$ the initial *input estimates* of (y_{01}, y_{02}, z_0), which do not satisfy the equation (1), i.e. $\hat{y}_1 \hat{y}_2 \neq \hat{z}$.

Suppose that, after an adjustment of choice, we obtain adjusted values defined by $(\widetilde{y}_1, \widetilde{y}_2, \widetilde{z})$ that do satisfy the constraint, so that we obtain

$$\mathbb{A} = [\widetilde{y}_1 \widetilde{y}_2 = \widetilde{z}]. \tag{3}$$

We shall refer to (3) as the *estimated account*. To develop a scalar measure of uncertainty, we shall consider the estimated account to be a *single* random outcome. Possible different realisations could be, for instance, $[100 \cdot 1.2 = 120]$ or $[80 \cdot 1.5 = 120]$ or $[90 \cdot 1.2 = 108]$.

To measure the variance of an estimated account we need a notion of its expectation. 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 $f(\widetilde{y}_1, \widetilde{y}_2, \cdots, \widetilde{y}_p, \widetilde{z})$. The expected additive account is obtained by direct substitution of the expectations of the component variables. The expected multiplicative account is less obvious because $E(\widetilde{y}_1)E(\widetilde{y}_2) \neq E(\widetilde{z})$. As a general

---

[1] The possibility of a combination of multiplicative and additive accounts, i.e. when the y-variables in (2) can be seen as the r.h.s. (z-variables) of multiplicative accounts (1), as suggested by a reviewer, is an extension that is not treated in this paper.

option, we propose to log-transform a multiplicative account to an additive account, and define the expected account on the log-scale. An alternative approach on the original scale, will be detailed in Section 4.2.

To motivate the need of a scalar uncertainty measure of an estimated account, we notice that an obvious measure of the uncertainty in the input vector $\hat{\mathbf{x}} = (\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_p, \hat{z})$ is the $(p+1) \times (p+1)$ variance-covariance matrix, denoted by $\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}$. Similarly, let $\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}}$ be the variance-covariance matrix of the adjusted estimates $\widetilde{\mathbf{x}} = (\widetilde{y}_1, \widetilde{y}_2, \cdots, \widetilde{y}_p, \widetilde{z})$, which includes the adjustment effect. This approach to measuring the uncertainty in accounts is investigated by, among others, Stone et al. (1942). A drawback of this approach is its multidimensional nature, which makes it difficult to interpret for the user, just as it hampers the comparison of different estimates of the same account for the producer.

In contrast, we propose to develop a *scalar* measure of uncertainty. This scalar measure depends on three elements: the vector of input estimates, the accounting equation and the adjustment method used to obtained the estimated account. The scalar measure would be helpful in at least two respects:

I. Let $\dot{\mathbf{x}} = (\dot{y}_1, \dot{y}_2, \cdots, \dot{y}_p, \dot{z})$ be an alternative adjustment to $\widetilde{\mathbf{x}}$, both having the same input estimates $\hat{\mathbf{x}}$. Let $\dot{\mathbb{A}}$ and $\mathbb{A}$ be the estimated accounts, respectively for $\dot{\mathbf{x}}$ and $\widetilde{\mathbf{x}}$. Let $u(\mathbb{A})$ and $u(\dot{\mathbb{A}})$ be the respective scalar uncertainty measures. The relative efficiency (RE), i.e. $u(\mathbb{A})/u(\dot{\mathbb{A}})$, would facilitate the choice between the two.

II. Write $u(\mathbb{A}; F_{\hat{\mathbf{x}}})$ to signify its dependence on $F_{\hat{\mathbf{x}}}$, the distribution of $\hat{\mathbf{x}}$. Any change in the input uncertainty would result in a different $F'_{\hat{\mathbf{x}}}$, and $u(\mathbb{A}; F'_{\hat{\mathbf{x}}})$ accordingly. The RE $u(\mathbb{A}; F'_{\hat{\mathbf{x}}})/u(\mathbb{A}; F_{\hat{\mathbf{x}}})$ could then help us to identify and to assess the changes, or improvements, in the input data, which are most effective in terms of the uncertainty of the final estimated account.

We shall investigate two different approaches to define a scalar measure of uncertainty: the covariance approach and the deviation approach. The basic ideas of the two approaches are as follows. For the *covariance* approach, one starts with the covariance matrix $\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}}$ as a multivariate measure of the expected deviation of $\widetilde{\mathbf{x}}$ from its expected value. This multivariate measure is then reduced to a scalar summary. 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 e.g. Seber (1984), which will be shown to have many desirable properties. For the *deviation* approach one first reduces the deviation of $\widetilde{\mathbf{x}}$ from its expectation to a scalar summary measure. Scalar summary measures based on L_1 and L_2 norms will be considered. The uncertainty measure is then defined as the expectation of this scalar measure. Thus, the *covariance* approach considers a scalar summary of the expected values of the multivariate deviations, whereas the *deviation* approach considers the expected value of a scalar summary of the multivariate deviations.

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. An issue is therefore to clarify in which sense a chosen measure summarizes as much as possible the relevant uncertainty. As we explain in Section 4.4, a unified interpretation can be given in terms of expected norm. Notice that using norm as a measure of the differences between vectors is a natural idea, also in the context of macro accounts; see e.g. de Mesnard (1990, 2004). A key difference is the statistical nature of the measure we propose in this paper. Let us explain in terms of the familiar biproportional techniques for matrix balancing in input-output analysis (e.g. Lahr and de Mesnard (2004)).

The problem is the estimation of a matrix based on the structure of a prior (matrix) and the margins (of target matrix). Since the true matrix, denoted by $\mathbf{Y}^*$ is not identifiable based on available information, a constrained optimisation approach is needed. In particular, let $\widehat{\mathbf{Y}}$ be the

initial matrix, by RAS, following Stone et al. (1942), one obtains

$$\widetilde{\mathbf{Y}} = \mathbf{R}\widehat{\mathbf{Y}}\mathbf{S}$$

where $\mathbf{R}$ and $\mathbf{S}$ are diagonal matrices of, respectively, the cumulated row and column multi-plicative adjustments. Lahr and de Mesnard (2004) refer to the difference between $\widetilde{\mathbf{Y}}$ and $\mathbf{Y}^*$ as the *gap* between the projection and the target. The uncertainty about the prior matrix and the target margins is referred to as *reliability*, and its incorporation in the biproportional techniques is discussed, as is noted in Section 1 above.

In the notation of this paper, denote by $\widehat{\mathbf{x}}$ the vector of all relevant input variables, including the margins. Some of the variables may have zero-variances, which are fully reliable and will not be adjusted. We denote the target by $\mathbf{x}^*$. Each target margin corresponds to an accounting equation, and the margin is fixed if it is fully reliable. By any chosen matrix balancing technique, which may or may not explicitly incorporate the reliability of $\widehat{\mathbf{x}}$, we obtain the adjusted $\widetilde{\mathbf{x}}$, with at least some components different to those of $\widehat{\mathbf{x}}$. Under the assumption $E(\widetilde{\mathbf{x}}) = \mathbf{x}^*$, each proposed scalar measure of accounting uncertainty is simply the *expected* gap between $\widetilde{\mathbf{x}}$ and $\mathbf{x}^*$ according to a chosen norm $\|\widetilde{\mathbf{x}} - \mathbf{x}^*\|$, as explained in Section 4.4. A statistical measure (i.e., the expected gap) will be necessary because one cannot estimate the gap between $\widetilde{\mathbf{x}}$ and $\mathbf{x}^*$ directly, insofar as $\mathbf{x}^*$ is unknown and is estimated by $\widetilde{\mathbf{x}}$. Of course, whether or not $E(\widetilde{\mathbf{x}}) = \mathbf{x}^*$, a scalar measure is always the expected gap between $\widetilde{\mathbf{x}}$ and $E(\widetilde{\mathbf{x}})$. To include the bias $E(\widetilde{\mathbf{x}}) - \mathbf{x}^*$ is generally difficult in reality, which is a matter beyond the scope of this paper.

3.   Example: An inconsistency in input estimates in supply and use tables

To explain the practical use of a scalar uncertainty measure we include, as a motivating example, a greatly simplified set of supply and use tables, see Bikker et al. (2013). With this example we illustrate the need to have an uncertainty measure to compare reconciliation methods and the convenience of having a scalar measure. In Table 1 the supply table is given at the top and the use table at the bottom. The rows of the supply table are related to the supply of products and services, and the columns to the producing sectors. The first two rows of the use table show the demand for products and services, and the first two columns show the customer sectors. The economy depicted in this example is a closed one, there is no trading with foreign countries. The tables do not contain import, taxes on products, subsidies, trade and transport margins. The grand totals of both tables can be derived directly from the other totals. There are only two sectors, industry and services, and two goods groups, industrial products and services, see Bikker et al. (2013) for more details.

In general the supply and use tables are subject to multiple constraints including the non-negativity of all variables. Constraints that should be satisfied for this tables are:

- total supply equals total use for industrial products and services:

$$y_1 + y_2 + (-y_5) + (-y_6) + (-y_7) = 0$$

$$y_3 + y_4 + (-y_8) + (-y_9) + (-y_{10}) = 0$$

- the sums of the entries of the tables must also equal the row and column totals:

$$y_1 + y_3 + (-y_5) + (-y_8) + (-y_{11}) + (-y_{13}) = 0$$

$$y_2 + y_4 + (-y_6) + (-y_9) + (-y_{12}) + (-y_{14}) = 0$$

Table 1. Supply and use table

		Industry	Services	Total	
Industrial products		700 (y_1)	300 (y_2)	1000	
Services		100 (y_3)	400 (y_4)	500	
Total		800	700		

	Industry		Services		Consumption		Total
Industrial products	50	(y_5)	190	(y_6)	860	(y_7)	1100
Services	170	(y_8)	100	(y_9)	180	(y_{10})	450
Wages	450	(y_{11})	350	(y_{12})			800
Operating surplus	130	(y_{13})	60	(y_{14})			190
Total	800		700		1040		

- all entries must be positive:

$$y_i > 0, \ i = 1, ..., 14.$$

Two constraints are not satisfied in the starting situation: total supply is unequal to total use for industrial products and services. To obtain adjusted values that satisfy these multiple constraints, we here use an optimization method that finds adjusted values that satisfy all constraints by minimizing the weighted sum of squared adjustments (see e.g. Stone et al. (1942)). This difference (the objective function of the optimization problem) can be defined in different ways, giving rise to different adjustment methods. The weighted least squares criterion is a common choice for the objective function. The weights determine the relative size of the adjustments: values with large weights are adjusted less than those with smaller weights. A common choice of weights is the inverse of the variance of the initial estimates. This means that initial values with a small variance get a large weight and are therefore less adjusted than initial values with large variances. Note that for the weights only the relative sizes of the variances are relevant, so we only need to know the variance up to a constant factor. Moreover, in many applications the relative variances cannot all be estimated and expert judgments are made about the relative accuracy. In the absence of knowledge about the relative accuracy, a weighting scheme that is sometimes used is weighting by the inverse of the initial values. With these weights, larger values are adjusted more than the smaller ones which means that we are minimizing a relative error rather than an absolute one.

The resulting adjusted values are determined by the choice of weights and to investigate the impact of this choice, we applied three common weighting schemes: equal, inverse variance and inverse value, see Section 5.4 for details. The results are in Table 2. The variances of $\widehat{y}$ were chosen arbitrarily for the purpose of this illustration. In addition, we have chosen $V(\widehat{y}_5 + \widehat{y}_8 + \widehat{y}_{11} + \widehat{y}_{13}) = 3400$ and $V(\widehat{y}_6 + \widehat{y}_9 + \widehat{y}_{12} + \widehat{y}_{14}) = 3000$. We assume independent input estimators otherwise. All the three sets of adjusted figures satisfy the constraints above.

In practice one would like to have criteria which can help to choose among the different solutions and determine the most appropriate adjustment method. For all three adjustment methods we can calculate the corresponding variance-covariance matrices. In the previous sections we propose to use a scalar measure of uncertainty. This measure is based on the covariance of the adjusted figures. As mentioned above we can use this scalar measure to compare different adjustment methods. The relative efficiency (RE) will facilitate the choice between the different methods.

Table 2. Adjusted values $\widetilde{y}^{(a)}$ with the equal weight, $\widetilde{y}^{(b)}$ with the weights equal to the inverse of the variance of $\widehat{y}$, $\widetilde{y}^{(c)}$ with the weights equal to the inverse of $\widehat{y}$.

	$\widehat{y}$	Variance	$\widetilde{y}^{(a)}$	$\widetilde{y}^{(b)}$	$\widetilde{y}^{(c)}$
y_1	700	100	718.5709	703.3753	725.7539
y_2	300	1000	318.5709	322.6111	315.5935
y_3	100	1000	88.5709	90.6055	93.6004
y_4	400	100	388.5709	397.9464	380.4763
y_5	50	500	31.4291	33.1235	48.1604
y_6	190	1000	171.4291	167.3889	180.1241
y_7	860	1000	834.2873	825.4771	813.0654
y_8	170	1000	181.4291	179.3945	180.8793
y_9	100	1000	111.4291	120.5363	104.8809
y_{10}	180	1000	184.2873	188.6246	188.3182
y_{11}	450	700	457.1418	450.5389	458.0027
y_{12}	350	700	357.1418	358.3382	350.9089
y_{13}	130	1200	137.1418	130.9239	132.3119
y_{14}	60	1200	67.1418	74.2941	60.1558

4. Scalar approaches to accounting uncertainty

In this section we develop two different scalar approaches, and present a unified interpretation. However, we would first like to clarify a necessary property, permutation invariance, that any admissible measure should possess, see e.g. Watrous (2018).

4.1. *Permutation invariance*

Under the additive constraint (2), any of the following *permutation rearrangements* of the estimated account 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 placement of a component variable in an additive account should not matter, because we can always rearrange their positions in the account without changing the underlying constraint. Thus, a scalar uncertainty measure is said to be *strongly permutation invariant* if it remains the same for any permutation rearrangement of the accounting equation $\mathbb{A}$. In practical applications it could be convenient to rearrange the account. If for e.g. we know that a certain initial entry of the account should be treated as constant, meaning in our setting that it has variance equal to zero, we would place that entry on the right side of the accounting equation.

Similarly, permutation rearrangements can be applied to a multiplicative accounting equation under the same constraint. But the interpretation equivalence may be less intuitive. For instance, [Value $\cdot$ (1/Quantum) = Price] is possibly just as acceptable as [Quantum $\cdot$ Price = Value], whereas [Quantum $\cdot$ (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 equivalency of permutation rearrangements is more readily acceptable:

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

The transformation on the log scale makes it also easier to deal with the accounts uncertainties, see Section 4.2.2.

Secondly, a scalar uncertainty measure is said to be *weakly permutation invariant* if it remains the same for any permutation rearrangement of the left-hand side of an account $\mathbb{A}$, 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 4.2 below.

4.2. *Covariance approach*

4.2.1. *Single additive account*

We start with the additive account (2). Consider the variance-covariance matrix $\mathbf{\Sigma}_{\widetilde{\mathbf{x}}}$ of the adjusted vector

$$\widetilde{\mathbf{x}} = (\widetilde{y}_1, \widetilde{y}_2, \cdots, \widetilde{y}_p, \widetilde{z})^T,$$

we have that

$$\mathbf{\Sigma}_{\widetilde{\mathbf{x}}} = \begin{pmatrix} \mathbf{\Sigma}_{\widetilde{\mathbf{y}}} & \mathbf{\Sigma}_{\widetilde{\mathbf{y}}\widetilde{z}} \\ \mathbf{\Sigma}_{\widetilde{z}\widetilde{\mathbf{y}}} & \mathbf{\Sigma}_{\widetilde{z}} \end{pmatrix}$$

Here $\mathbf{\Sigma}_{\widetilde{\mathbf{y}}}$ is the variance-covariance matrix of $\widetilde{\mathbf{y}} = (\widetilde{y}_1, \widetilde{y}_2, \cdots, \widetilde{y}_p)^T$, and $\mathbf{\Sigma}_{\widetilde{z}}$ is the variance of $\widetilde{z}$ as it is just one variable and

$$\mathbf{\Sigma}_{\widetilde{\mathbf{y}}\widetilde{z}} = (cov(\widetilde{y}_1, \widetilde{z}), cov(\widetilde{y}_2, \widetilde{z}), ..., cov(\widetilde{y}_p, \widetilde{z}))^T$$

and

$$\mathbf{\Sigma}_{\widetilde{z}\widetilde{\mathbf{y}}} = (cov(\widetilde{z}, \widetilde{y}_1), cov(\widetilde{z}, \widetilde{y}_2), ..., cov(\widetilde{z}, \widetilde{y}_p))$$

are $(p, 1)$ and $(1, p)$ dimensional matrices. As $\widetilde{\mathbf{y}}$'s are the adjusted vectors and satisfy the account in (2) we have that

$$\sum_{k=1}^{p} \widetilde{y}_k = \widetilde{z} \tag{4}$$

and we will also have that:

$$\sum_{k=1}^{p} cov(\widetilde{z}, \widetilde{y}_k) = \sum_{k=1}^{p} cov(\widetilde{y}_k, \widetilde{z}) = V(\widetilde{z})$$

and for $k = 1, ..., p$

$$\sum_{i=1}^{p} cov(\widetilde{y}_k, \widetilde{y}_i) = cov(\widetilde{y}_k, \widetilde{z}).$$

Now, to define a simple scalar summary of the variance-covariance matrix $\mathbf{\Sigma}_{\widetilde{\mathbf{x}}}$ we consider the following quantities: the sum of all elements (variances and covariances) of $\mathbf{\Sigma}_{\widetilde{\mathbf{x}}}$ and the sum of the diagonal elements (variances) only. The first scalar, defined by $\tau_1(\mathbb{A})$ will be the sum of variances of $\widetilde{y}$'s and $\widetilde{z}$ and will include all covariances. The second scalar, $\tau_2(\mathbb{A})$ is actually the

trace of the matrix $\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}}$:

$$\tau_1(\mathbb{A}) = \mathbf{1}^T \boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}} \mathbf{1} = \sum_{k,i=1}^{p} cov(\widetilde{y}_k, \widetilde{y}_i) + 2\sum_{k=1}^{p} cov(\widetilde{y}_k, \widetilde{z}) + V(\widetilde{z}) = 4V(\widetilde{z}) \tag{5}$$

and

$$\tau_2(\mathbb{A}) = \text{Trace}(\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}}) = \sum_{k=1}^{p} V(\widetilde{y}_k) + V(\widetilde{z}). \tag{6}$$

The second measure τ_2 is also called the total variation (Seber, 1984) and, in the context of accounts with constraints, it will generally be preferable to the first measure τ_1 for several reasons.

First, due to the constraint (4), τ_1 actually contains 'less' of the information in $\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}}$, since $\tau_2 = \text{Trace}(\boldsymbol{\Sigma}_{\widetilde{\mathbf{y}}}) + \tau_1/4$. We lose the variances of $\widetilde{y}$'s.

Next, observe that τ_2 is strongly permutation invariant, whilst τ_1 is only weakly permutation invariant. To see this consider the two permutations of the estimated account:

$$[\widetilde{y}_1 + \widetilde{y}_2 + \cdots + \widetilde{y}_p = \widetilde{z}] \quad \text{and} \quad [(-\widetilde{z}) + \widetilde{y}_2 + \cdots + \widetilde{y}_p = (-\widetilde{y}_1)].$$

Define $\widetilde{\mathbf{x}}_1 = (\widetilde{y}_1, ..., \widetilde{y}_p, \widetilde{z})^T$ and $\widetilde{\mathbf{x}}_2 = ((-\widetilde{z}), \widetilde{y}_2..., \widetilde{y}_p, (-\widetilde{y}_1))^T$, with covariance matrices $\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}_1}$ and $\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}_2}$. Then we will have that $\tau_1^{\widetilde{\mathbf{x}}_1}(\mathbb{A}) = 4V(\widetilde{z})$ and $\tau_1^{\widetilde{\mathbf{x}}_2}(\mathbb{A}) = 4V(-\widetilde{y}_1)$ and we would have different values for τ_1. This is not the case for τ_2, which remains the same for all permutations. Moreover, in many practical situations, some of the inputs may be treated as 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, and the variables on the left hand side, denoted by $\widetilde{\mathbf{y}} = (\widetilde{y}_1, ..., \widetilde{y}_{p^*})^T$, sum up to a constant z, where p^* is the number of unfixed components. For $\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}}$ we will then have $\tau_1 = \mathbf{1}^T \boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}} \mathbf{1} \propto \mathbf{1}^T \boldsymbol{\Sigma}_{\widetilde{\mathbf{y}}} \mathbf{1} = V(z) = 0$. In contrast, $\tau_2 = \text{Trace}(\boldsymbol{\Sigma}_{\widetilde{\mathbf{y}}}) + V(z) = \text{Trace}(\boldsymbol{\Sigma}_{\widetilde{\mathbf{y}}})$ would remain a meaningful measure of the accounting uncertainty. From now on we consider only τ_2.

4.2.2. *Multiplicative account*

Consider now the multiplicative account with $p = 2$. As noted previously, a general option is to measure the accounting uncertainty on the log scale, which is additive. Let the log-transformed account be $\mathbb{B} = [\log(\widetilde{y}_1) + \log(\widetilde{y}_2) = \log(\widetilde{z})]$. We can define

$$\tau_2(\mathbb{B}) = V(\log(\widetilde{y}_1)) + V(\log(\widetilde{y}_2)) + V(\log(\widetilde{z})). \tag{7}$$

In Section 2 we also mentioned an alternative approach on the original scale, which is weakly permutation invariant. This approach can be described as follows. Let $(\widetilde{z}, D)$ be a one-one transformation of $(\widetilde{y}_1, \widetilde{y}_2)$, so that D summarizes all the *conditional* variation in $(\widetilde{y}_1, \widetilde{y}_2)$ given $\widetilde{z}$. Weakly permutation invariance is achieved provided $V(D)$ is invariant for $\widetilde{\mathbf{y}}$-permutations, i.e. D is symmetric in $(\widetilde{y}_1, \widetilde{y}_2)$. A simple admissible choice is $D(\widetilde{y}_1, \widetilde{y}_2) = \widetilde{y}_1 + \widetilde{y}_2$, for which the expectation and variance are well defined. In particular, we obtain $V(D) = \mathbf{1}^T \boldsymbol{\Sigma}_{\widetilde{\mathbf{y}}} \mathbf{1}$. By the total

variance argument we can write that

$$V(D) = E_{\widetilde{z}}[V(D|\widetilde{z})] + V_{\widetilde{z}}[E(D|\widetilde{z})]. \tag{8}$$

The conditional covariance between $\widetilde{y}_1$ and $\widetilde{y}_2$ is given by

$$Cov(\widetilde{y}_1, \widetilde{y}_2|\widetilde{z}) = E(\widetilde{y}_1\widetilde{y}_2|\widetilde{z}) - E(\widetilde{y}_1|\widetilde{z})E(\widetilde{y}_2|\widetilde{z}) = \widetilde{z} - E(\widetilde{y}_1|\widetilde{z})E(\widetilde{y}_2|\widetilde{z}),$$

which is needed for the first term on the right-hand side of (8). One may consider the second term on the right-hand side of (8) to be a summary of the variance of $\widetilde{z}$, in cases it is not treated as fixed. We therefore define a scalar measure, denoted by $\tau(\mathbb{A})$, to be

$$\begin{aligned}
\tau(\mathbb{A}) &\stackrel{def}{=} E[V(D|\widetilde{z})] + V(\widetilde{z}) \\
&= E[V(\widetilde{y}_1|\widetilde{z})] + E[V(\widetilde{y}_2|\widetilde{z})] + 2E(\widetilde{z}) - 2E[E(\widetilde{y}_1|\widetilde{z})E(\widetilde{y}_2|\widetilde{z})] + V(\widetilde{z}).
\end{aligned} \tag{9}$$

The measure (9) involves in one way or another the first two moments of $\widetilde{y}_1$, $\widetilde{y}_2$ and $\widetilde{z}$ on the original scale. It is weakly permutation invariant given any symmetric D. Since this measure is based on the decomposition of the total variance, we will also refer to this particular form of covariance approach for multiplicative accounts as the decomposition approach.

4.2.3. *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 will have multiple equations like (2). The y-variables on the left-hand side of these equations are different and 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:

$$\mathbf{H}\mathbf{y} = \mathbf{z}, \tag{10}$$

where $\mathbf{H}$ is a $q \times p$ matrix with constant elements such as 0 and 1, and q is the number of constraints and the size of $\mathbf{z}$-vector. The vector $\widetilde{\mathbf{x}}$ is now given as

$$\widetilde{\mathbf{x}} = (\widetilde{y}_1, \widetilde{y}_2, \cdots, \widetilde{y}_p, \widetilde{z}_1, \widetilde{z}_2, \cdots, \widetilde{z}_q) = (\widetilde{\mathbf{y}}, \mathbf{H}\widetilde{\mathbf{y}}) = (\widetilde{x}_1, \widetilde{x}_2, \cdots, \widetilde{x}_p, \widetilde{x}_{p+1}\widetilde{x}_{p+2}, \cdots, \widetilde{x}_{p+q}).$$

The variance-covariance matrix $\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}}$ has the following structure

$$\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}} = \begin{pmatrix} \boldsymbol{\Sigma}_{\widetilde{\mathbf{y}}} & \boldsymbol{\Sigma}_{\widetilde{\mathbf{y}}\widetilde{\mathbf{z}}} \\ \boldsymbol{\Sigma}_{\widetilde{\mathbf{z}}\widetilde{\mathbf{y}}} & \boldsymbol{\Sigma}_{\widetilde{\mathbf{z}}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\widetilde{\mathbf{y}}} & \boldsymbol{\Sigma}_{\widetilde{\mathbf{y}}}\mathbf{H}^T \\ \mathbf{H}\boldsymbol{\Sigma}_{\widetilde{\mathbf{y}}} & \mathbf{H}\boldsymbol{\Sigma}_{\widetilde{\mathbf{y}}}\mathbf{H}^T \end{pmatrix}.$$

The scalar measure τ_2 can be defined as in the single-constraint case:

$$\tau_2(\mathbb{A}) = \text{Trace}(\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}}) = \sum_{k=1}^{p} V(\widetilde{y}_k) + \sum_{l=1}^{q} V(\widetilde{z}_l) = \text{Trace}(\boldsymbol{\Sigma}_{\widetilde{\mathbf{y}}}) + \text{Trace}(\boldsymbol{\Sigma}_{\widetilde{\mathbf{z}}}).$$

Apart from the supply and use table example cited above, multiple additive constraints also arise when accounts are hierarchical, with a higher level datum being itself the sum of lower-level data. Such a hierarchical or nested account can be rewritten as two equations, one is the higher level account with the lower-level variables substituted and the other consists of the lower-level variables and their sum (the higher-level datum).

4.3. *Deviation approach*

4.3.1. *Additive account*

For the additive account (2), let the vector $E(\mathbf{x}) = (M_1, ..., M_p, M_{p+1})^T$ be the expectation of the true $\mathbf{x} = (y_1, y_2, ..., y_p, z)$, such that $\widetilde{\mathbf{x}} - E(\mathbf{x})$ contains the deviation of all the adjusted components from those of their expected values. We can adopt a suitable scalar summary of the component-wise deviations, generically denoted by $\delta = \delta(\widetilde{\mathbf{x}} - M_{\mathbf{x}})$. Different choices are possible. Two simple choices could be

$$\delta_1 = \sum_{k=1}^{p+1} w_k |\widetilde{x}_k - M_k| \qquad \text{and} \qquad \delta_2 = \sum_{k=1}^{p+1} w_k (\widetilde{x}_k - M_k)^2, \tag{11}$$

where $\widetilde{x}_k$ is the k-th element of $\widetilde{\mathbf{x}}$ and $w_k \geq 0$, to be referred to as the *averaging* weights. As a scalar uncertainty measure we use

$$\Delta_i(\mathbb{A}) = E(\delta_i) \qquad (i = 1, 2). \tag{12}$$

The deviation approach (12) 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 (6). 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,[2] we can view the account as a decomposition of this total z into its components $\widetilde{y}_1, \ldots, \widetilde{y}_p$, or as a classification of the amount z into p-categories. The differences $\widetilde{y}_k - M_k$ can be interpreted as the classification errors that sum to zero over k, i.e. 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 move 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. An EMD-related choice is $w_k = 1/2$, in which case δ_1 is the amount of z that is classified differently by $\widetilde{y}_k$ than by M_k. For δ_2 a related choice is $w_k = 1/p$, in which case it becomes the mean squared classification error.

4.3.2. *Multiplicative account*

For the multiplicative account with $p = 2$, the expected account on the log scale equals

$$\mu(\mathbb{B}) = [E(\log(\widetilde{y}_1)) + E(\log(\widetilde{y}_2)) = E(\log(\widetilde{z}))].$$

Let $m_k = E(\log(\widetilde{y}_k))$, for $k = 1, 2$, and $m_{\widetilde{z}} = E(\log(\widetilde{z}))$. We have $\mu(\mathbb{B}) = [m_1 + m_2 = m_{\widetilde{z}}]$, which implies $e^{\mu(\mathbb{B})} = [e^{m_1} e^{m_2} = e^{m_{\widetilde{z}}}]$. 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

---

[2] In some practical applications it could occur that some figures can not be adjusted, for e.g. if figures are already published.

of unity, a transformation to the log scale is employed. This yields then

$$\delta_1 = \sum_{k=1}^{3} w_k |\log(\widetilde{x}_k) - m_k| \qquad \text{and} \qquad \delta_2 = \sum_{k=1}^{3} w_k (\log(\widetilde{x}_k) - m_k)^2, \qquad (13)$$

and the corresponding $\Delta(\mathbb{A})$ by (12). Again, this amounts to measuring the uncertainty of a multiplicative account $\mathbb{A}$ in terms of the corresponding $\mathbb{B}$ on the log scale. In section 5.2 we illustrate this approach for the index problem.

4.4. *General framework for a scalar uncertainty measure*

In both the covariance and deviation approaches above 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 the National Accounts, there do exist situations where the constraint matrix is random itself. For example, let $\mathbf{y}$ be the vector of domain some population counts in year t, and $\mathbf{z}$ 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 consider $\mathbf{Hy} = \mathbf{z}$ as a population account with random constraint matrix $\mathbf{H}$. The matrix $\mathbf{H}$ has the same dimension as in (10), however it has stochastic entries. We now present expected norm as a scalar measure of the generic *accounting system* (10), with fixed or random constraint matrix $\mathbf{H}$, which provides a unified interpretation of the covariance and deviation approaches above.

For a vector $\mathbf{x}^*$ that we want to adjust so that (10) is fulfilled, the accounting system may be expressed by $\mathbb{A}(\mathbf{x}^*) = [\mathbf{Hy} = \mathbf{z}; \mathbf{x} = \mathbf{x}^*]$. Define the *account space* Ω for all the $\mathbf{x}$-vectors that satisfy the account as follows:

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

In particular, let $\mathbf{x}_0$ be the true $\mathbf{x}$-vector, where $\mathbf{x}_0 \in \Omega$, and $\mathbb{A}_0 = \mathbb{A}(\mathbf{x}_0) = [\mathbf{Hy} = \mathbf{z}; \mathbf{x} = \mathbf{x}_0]$. Let $\widehat{\mathbf{x}}$ be the vector of initial input estimates, where $\widehat{\mathbf{x}} \notin \Omega$, if $\mathbf{H}(\widehat{\mathbf{x}})\mathbf{y}(\widehat{\mathbf{x}}) \neq \mathbf{z}(\widehat{\mathbf{x}})$. Let $\widetilde{\mathbf{x}}$ be the vector of adjusted estimates, where $\widetilde{\mathbf{x}} \in \Omega$ and $\mathbf{H}(\widetilde{\mathbf{x}})\mathbf{y}(\widetilde{\mathbf{x}}) = \mathbf{z}(\widetilde{\mathbf{x}})$, with the corresponding estimated system $\widetilde{\mathbb{A}} = \mathbb{A}(\widetilde{\mathbf{x}}) = [\mathbf{Hy} = \mathbf{z}; \mathbf{x} = \widetilde{\mathbf{x}}]$.

Thus, the account space Ω is a subset of the vector space of $\mathbf{x}$; the true vector $\mathbf{x}_0$ is a point in Ω; the initial vector $\widehat{\mathbf{x}}$ is generally a point outside of Ω; the adjusted vector $\widetilde{\mathbf{x}}$ is a point in Ω, which is typically obtained from $\widehat{\mathbf{x}}$ via some minimum changes subjected to the constraints (10). The difference $\widehat{\mathbf{x}} - \mathbf{x}_0$ is the error of the initial vector of estimates, the change $\widetilde{\mathbf{x}} - \widehat{\mathbf{x}}$ is needed in order to move from $\widehat{\mathbf{x}}$ outside Ω to $\widetilde{\mathbf{x}}$ inside Ω, and the final difference $\widetilde{\mathbf{x}} - \mathbf{x}_0$ is the error of the adjusted vector of estimates. An illustration of $[x_1 + x_2 = x_3]$ with fixed x_3 is given in Figure 1. The account space Ω is the line intersecting both axes in this case, on which (x_{01}, x_{02}) marks the point of the true account. The input estimator $\widehat{\mathbf{x}} = (\widehat{x}_1, \widehat{x}_2)$ is outside the account space Ω, insofar as the corresponding point is not on the line. Orthogonally projecting $\widehat{\mathbf{x}}$ onto the line (Ω) yields the adjusted point $(\widetilde{x}_1, \widetilde{x}_2)$. The change $\widetilde{\mathbf{x}} - \widehat{\mathbf{x}}$ is marked by the line connecting these two points.

Figure 1.   Illustration of $[x_1 + x_2 = x_3]$ with fixed x_3

One may envisage the initial and final vectors of error, i.e. $\widehat{\mathbf{x}} - \mathbf{x}_0$ and $\widetilde{\mathbf{x}} - \mathbf{x}_0$, as two vectors relocated to $\mathbf{x}_0$ as the origin, i.e. one may centre the vector space of $\mathbf{x}$ at $\mathbf{x}_0$. 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 $\widehat{\mathbf{x}}$ and $\widetilde{\mathbf{x}}$ because $\mathbf{x}_0$ is unknown. However, it is possible to

measure the average size over the distribution of $\widehat{\mathbf{x}}$ and $\widetilde{\mathbf{x}}$, respectively, if these are (assumed to) be known or can be estimated (see sections 5.1 and 5.3). Such averaging measures are referred to as *statistical* uncertainty measures. Figure 1 provides a geometric interpretation of the expected norm. All the possible points of $(\widehat{x}_1, \widehat{x}_2)$ can be placed on the infinite many concentric circles around (x_{01}, x_{02}). The average radius of these circles, i.e. $E[(\widehat{x}_1 - x_{01})^2 + (\widehat{x}_2 - x_{02})^2]$, which is the expected Euclidean norm of $\widehat{\mathbf{x}} - \mathbf{x}_0$, is marked by the circle in Figure 1. Meanwhile, all the adjusted points $(\widetilde{x}_1, \widetilde{x}_2)$ are on the line passing through (x_{01}, x_{02}), so that the expected Euclidean norm $E[(\widetilde{x}_1 - x_{01})^2 + (\widetilde{x}_2 - x_{02})^2]$ can be indicated by two points on the line with equal distance to (x_{01}, x_{02}).

Weights can be introduced in the norm. For instance, the weighted Euclidean norm (WEN) of $v = \widetilde{\mathbf{x}} - \mathbf{x}_0$ is given by

$$\|v\|_{wen} = \sum_{i=1}^{K} a_i v_i^2, \tag{14}$$

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

$$\Delta_v = E\big(\|v\|_{wen}\big).$$

The WEN (14) is strongly permutation invariant, provided component-specific weights a_i. Provided unbiased $\widetilde{\mathbf{x}}$, setting $a_i \equiv 1$, we recover τ_2 by the covariance approach; setting a_i to be the averaging weights w_i, we recover Δ_2 by the deviation approach, or Δ_1 if we replace the L_2 norm in (14) by the L_1 norm. Thus, using the expected norm as a scalar statistical uncertainty measure of an accounting system unifies the covariance and deviation approaches.

5. Illustrations and applications

In this section we illustrate the theory developed above for both additive and multiplicative accounts under simplified yet typical scenarios, for which closed expressions of the uncertainty measures can be obtained. For more complex situations, we outline a simulation-based computation approach and compare it to the analytic results obtained in the example of an additive account. Finally, we present an application to time series data subjected to additive constraints.

5.1. *Additive account with independent normal input estimators*

Suppose we have the input estimates $\widehat{\mathbf{y}} = (\widehat{y}_1, \ldots, \widehat{y}_p) \sim N_p(\mu, \Sigma_{\mathbf{y}})$, where $\mu = (\mu_1, \ldots, \mu_p)$ and $\Sigma_{\mathbf{y}}$ is a diagonal matrix with the diagonal values $\sigma_k^2 = V(y_k) = \sigma^2 \mu_k$, for $k = 1, \ldots, p$. The scenario of independent normal input estimators represents a standard practical situation, where one appeals to the Central Limit Theorem for the distribution of the estimators derived from independent large sample surveys. Suppose the additive accounting equation of the form (2), where $\widetilde{z} = z$ is treated as fixed. Consider a common benchmarking method, which yields the adjusted estimates

$$\widetilde{y}_k = \widehat{y}_k + (z - \sum_{j=1}^{p} \widehat{y}_j)\nu_k, \tag{15}$$

where ν_k are the adjustment weights that sum up to 1, for $k = 1, \ldots, p$, by which the total difference is simply apportioned to each component estimator.

In (A1) and (A2) from Appendix A we derive the expectation and variance for $\widetilde{y}_k$. It is easy to see that since y_k have the normal distributions, then the random variables $\widetilde{y}_k - M_k, k = 1, \ldots, p,$

also have normal distributions $N(0, \widetilde{\sigma}_k^2)$ and due to the additive constraint $\widehat{y}_1 + \cdots + \widehat{y}_p = z$, the $\widetilde{y}_k$'s are negatively correlated with each other.

Suppose one would like to compare two choices: $\nu_{1k} = 1/p$ and $\nu_{2k} = \mu_k / \sum_{j=1}^p \mu_j$, which yield, respectively, the following adjusted estimates

$$\widetilde{y}_{1k} = \widehat{y}_k + \frac{1}{p}(z - \sum_{j=1}^p \widehat{y}_j) \tag{16}$$

and

$$\widetilde{y}_{2k} = \widehat{y}_k + \frac{\mu_k}{\sum_{j=1}^p \mu_j}(z - \sum_{j=1}^p \widehat{y}_j). \tag{17}$$

Denote the reconciled estimated account by $\mathbb{A}_1$ based on (16) and by $\mathbb{A}_2$ based on (17).

We want to compare uncertainty measures defined above for these accounts. It is obvious from (5) that $\tau_1(\mathbb{A}_1) = 0$, since z is fixed. For the measure τ_2 given by (6) from (A4) follows that:

$$\tau_2(\mathbb{A}_1) - \tau_2(\mathbb{A}_2) = \sum_{j=1}^p V(\widetilde{y}_{1k}) - \sum_{j=1}^p V(\widetilde{y}_{2k}) > 0.$$

This implies that according to τ_2, the adjustment method (16) is always more uncertain than (17) as long as the μ_k's are not all equal to each other.

Similarly for Δ_2 given by (12) derived from δ_2 in (11), for arbitrary positive $w_k \geq 0$

$$\Delta_2(\mathbb{A}_1|z) - \Delta_2(\mathbb{A}_2|z) = E(\sum_{k=1}^p w_k(\widetilde{y}_{1k} - M_k)^2) - E(\sum_{k=1}^p w_k(\widetilde{y}_{2k} - M_k)^2) \tag{18}$$

$$= \sum_{k=1}^p w_k(V(\widetilde{y}_{1k} - M_k) - V(\widetilde{y}_{2k} - M_k)) \tag{19}$$

$$= \sum_{k=1}^p w_k \sigma^2 \left(\frac{\sqrt{\sum_{j=1}^p \mu_j}}{p} - \frac{\mu_k}{\sqrt{\sum_{j=1}^p \mu_j}} \right)^2 \geq 0. \tag{20}$$

where the notation $(\cdot|z)$ emphasizes that $\widetilde{z} = z$ is fixed. Again we conclude that the adjustment method (16) is always more uncertain than (17).

Moreover, $|\widetilde{y}_k - M_k|$ has the half-normal distribution with expectation equal to $\widetilde{\sigma}_k \sqrt{2/\pi}$, we obtain the measure Δ_1 given by (12) derived from δ_1 in (11):

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

We have that

$$\Delta_1(\mathbb{A}_1|z) - \Delta_1(\mathbb{A}_2|z) = \sigma \sqrt{2/\pi} \sum_{k=1}^p w_k \left(\sqrt{\frac{1}{p^2} \sum_{j=1}^p \mu_j + \mu_k(1 - \frac{2}{p})} - \sqrt{\mu_k - \frac{\mu_k^2}{\sum_{j=1}^p \mu_j}} \right).$$

Again, the adjustment method (16) is more uncertain than (17) because

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

– see (A4) in Appendix A, where the inequality holds as long as the μ_k's are not all equal to each other.

5.2.   *A multiplicative account: The index problem*

Here we illustrate the covariance and decomposition approaches to the multiplicative account:

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

The account above is related to the so-called *index problem*, where in general the directly calculated indices do not satisfy the constraint, i.e. $\widehat{I}_Q \cdot \widehat{I}_P \neq \widehat{I}_V$. To focus on the comparison between alternative adjustment methods for the index problem, we make the following simplifying assumptions.

- The true account is $\mathbb{A}_0 = [1 \cdot 1 = 1]$, or $\mathbb{B}_0 = [0 + 0 = 0]$ on the log-scale.
- The initial estimates are unbiased: $E(\widehat{\mathbf{x}}) = \mathbf{x}_0$, where $\widehat{\mathbf{x}} = (\widehat{I}_Q, \widehat{I}_P, \widehat{I}_V)$ and $\mathbf{x}_0 = (I_Q, I_P, I_V)$.
- Let $\widehat{\iota}_P = \log \widehat{I}_P, \widehat{\iota}_Q = \log \widehat{\iota}_Q$ and $\widehat{\iota}_V = \log \widehat{I}_V$. The initial estimates are independent and

$$V(\widehat{I}_V) = c_1 V(\widehat{I}_P) \qquad \text{and} \qquad V(\widehat{I}_Q) = c_2 V(\widehat{I}_P);$$
$$V(\widehat{\iota}_V) = \gamma_1 V(\widehat{\iota}_P) \qquad \text{and} \qquad V(\widehat{\iota}_Q) = \gamma_2 V(\widehat{\iota}_P).$$

Consider two different adjustments of $\widehat{I}_Q \cdot \widehat{I}_P \neq \widehat{I}_V$. First, the *deflation adjustment* is common e.g. in the National Accounts, by which the quantum index is derived from the other two, yielding $\mathbb{A}^* = [\widetilde{I}_Q = \widehat{I}_V / \widehat{I}_P] = [\widetilde{I}_Q \cdot \widehat{I}_P = \widehat{I}_V]$, and on the log-scale:

$$\mathbb{B}^* = [\widetilde{\iota}_Q = \widehat{\iota}_V - \widehat{\iota}_P] = [\widetilde{\iota}_Q + \widehat{\iota}_P = \widehat{\iota}_V]. \tag{21}$$

Second, treating proportional adjustment in a symmetric manner leads us to consider *joint adjustment* by the following minimization problem on the log-scale:

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

Denote the corresponding estimated account by $\mathbb{A} = [\widetilde{I}_Q \cdot \widetilde{I}_P = \widetilde{I}_V]$ and $\mathbb{B} = [\widetilde{\iota}_Q + \widetilde{\iota}_P = \widetilde{\iota}_V]$ on the log-scale. Let $3\alpha = \widehat{\iota}_V - (\widehat{\iota}_Q + \widehat{\iota}_P)$. Using the method of Lagrange multipliers we will obtain the second estimate

$$\begin{cases} \widetilde{\iota}_V = \widehat{\iota}_V - \alpha = \frac{2}{3}\widehat{\iota}_V + \frac{1}{3}\widehat{\iota}_Q + \frac{1}{3}\widehat{\iota}_P \\ \widetilde{\iota}_Q = \widehat{\iota}_Q + \alpha = \frac{1}{3}\widehat{\iota}_V + \frac{2}{3}\widehat{\iota}_Q - \frac{1}{3}\widehat{\iota}_P \\ \widetilde{\iota}_P = \widehat{\iota}_P + \alpha = \frac{1}{3}\widehat{\iota}_V - \frac{1}{3}\widehat{\iota}_Q + \frac{2}{3}\widehat{\iota}_P \end{cases} \quad \Leftrightarrow \quad \begin{cases} \widetilde{I}_V = \widehat{I}_V \exp(-\alpha) \\ \widetilde{I}_Q = \widehat{I}_Q \exp(\alpha) \\ \widetilde{I}_P = \widehat{I}_P \exp(\alpha). \end{cases} \tag{22}$$

Consider first the scalar measure τ_2 on the log-scale given by (7). For the deflation adjustment (21), we have $\widetilde{\iota}_V = \widehat{\iota}_V, \widetilde{\iota}_P = \widehat{\iota}_P$ and $\widetilde{\iota}_Q = \widehat{\iota}_V - \widehat{\iota}_P$. It follows that

$$\tau_2(\mathbb{B}^*) = V(\widehat{\iota}_V) + V(\widehat{\iota}_P) + V(\widehat{\iota}_V - \widehat{\iota}_P) = 2(\gamma_1 + 1)V(\widehat{\iota}_P).$$

The joint adjustment (22) yields

$$\tau_2(\mathbb{B}) = \frac{2}{3}(\gamma_1 + \gamma_2 + 1)V(\widehat{\iota}_P).$$

Putting these together, we have

$$\tau_2(\mathbb{B}) \leq \tau_2(\mathbb{B}^*) \quad \Leftrightarrow \quad \gamma_2 \leq 2(\gamma_1 + 1).$$

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.

Consider next the decomposition approach (9) which is weakly permutation invariant. Under the deflation adjustment, where $\widetilde{y}_1 = \widetilde{I}_Q = \widehat{I}_V / \widehat{I}_P$, $\widetilde{y}_2 = \widetilde{I}_P = \widehat{I}_P$ and $\widetilde{z} = \widetilde{I}_V = \widehat{I}_V$, we obtain

$$\tau(\mathbb{A}^*) \approx (2 + c_1) V(\widehat{I}_P)$$

– see Appendix B. Whereas, for the joint adjustment, we have $\widetilde{I}_V = \widehat{I}_V^{\frac{2}{3}} \widehat{I}_Q^{\frac{1}{3}} \widehat{I}_P^{\frac{1}{3}}$, $\widetilde{I}_Q = \widehat{I}_V^{\frac{1}{3}} \widehat{I}_Q^{\frac{2}{3}} \widehat{I}_P^{-\frac{1}{3}}$ and $\widetilde{I}_P = \widehat{I}_V^{\frac{1}{3}} \widehat{I}_Q^{-\frac{1}{3}} \widehat{I}_P^{\frac{2}{3}}$. Let $D = \widetilde{I}_Q + \widetilde{I}_P$. Using $1/x + x \approx 2$ for $x \approx 1$, we have

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

i.e. D is approximately a constant given $\widetilde{I}_V$. It follows that $E[V(D|\widetilde{I}_V)] \approx 0$, so that

$$\tau(\mathbb{A}) \approx V(\widetilde{I}_V) \approx \frac{1}{9}(4c_1 + c_2 + 1) V(\widehat{I}_P).$$

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

$$\tau(\mathbb{A}) \leq \tau(\mathbb{A}^*) \quad \Leftrightarrow \quad c_2 \leq 17 + 5c_1.$$

5.3. *Simulation-based computation*

5.3.1. *General method*

Generally speaking, the adjusted estimates $\widetilde{\mathbf{x}}$ for a general accounting system $\mathbb{A}$ by (10) may not be available in a closed expression. A simulation-based approach to the computation of the uncertainty measures can be generically given as follows.

1. Specify and estimate the joint distribution of the input estimators, denoted by $f(\widehat{\mathbf{x}})$;
2. For a large J, simulate J input vectors $\mathbf{x}^{(j)} \sim f(\widehat{\mathbf{x}})$ independently for $j = 1, \ldots, J$;
3. For each simulated $\mathbf{x}^{(j)}$, obtain the corresponding adjusted vector $\widetilde{\mathbf{x}}^{(j)}$, in the same way as the actual adjusted vector $\widetilde{\mathbf{x}}$ based on the realized input vector $\widehat{\mathbf{x}}$;
4. Estimate $E(\widetilde{\mathbf{x}})$ and $\Sigma_{\widetilde{\mathbf{x}}}$, respectively, by the resample mean and variance-covariance matrix over the J simulated vectors $\widetilde{\mathbf{x}}^{(1)}, \ldots, \widetilde{\mathbf{x}}^{(J)}$, and the estimated measure τ_2 by the covariance approach;
5. Calculate the deviation summaries $\delta_1^{(j)}$ and $\delta_2^{(j)}$ corresponding to each $(\widetilde{\mathbf{x}}^{(j)}, E(\widetilde{\mathbf{x}}))$; estimate Δ_1 and Δ_2 by the deviation approach, respectively, as the resample mean of $\delta_1^{(j)}$ and $\delta_2^{(j)}$ over the J simulated deviation summaries.

The computation approach can easily be modified for the multiplicative account (1), whether the uncertainty measure is defined on the log scale or by the decomposition approach (9).

5.3.2. *Illustration using an additive account with normal inputs*

To illustrate the simulation-based computation approach, and to compare it with analytical results, we consider a single additive account as in Section 5.1 with the following specifications. Let $\widehat{\mathbf{y}} \sim N(\mu, \Sigma_{\widehat{\mathbf{y}}})$, with $p = 10$, and

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

and $\Sigma_{\mathbf{y}}$ a diagonal matrix with diagonal elements $0.1^2 \mu_k$, for $k = 1, \ldots, 10$. This completes Step 1, which specifies the distribution $f(\widehat{\mathbf{y}})$, now that z is treated as fixed. Let Step 2 be completed

with $J = 1000$. For Step 3, we calculate the two adjusted vectors by (16) and (17), respectively, for each simulated vector $\mathbf{y}^{(j)}$, $j = 1, ..., 1000$. Finally, complete Step 4 to obtain the estimated measure τ_2 and Step 5 to obtain the estimated measures Δ_1 and Δ_2.

Table 3. Uncertainty measures Δ_1, Δ_2 and τ_2 for $\widehat{\mathbf{y}}$, $\widetilde{\mathbf{y}}_1$ by (16) and $\widetilde{\mathbf{y}}_2$ by (17).

	$\Delta_1(\widehat{\mathbf{y}})$	$\Delta_2(\widehat{\mathbf{y}})$	$\tau_2(\widehat{\mathbf{y}})$	$\Delta_1(\widetilde{\mathbf{y}}_1)$	$\Delta_2(\widetilde{\mathbf{y}}_1)$	$\tau_2(\widetilde{\mathbf{y}}_1)$	$\Delta_1(\widetilde{\mathbf{y}}_2)$	$\Delta_2(\widetilde{\mathbf{y}}_2)$	$\tau_2(\widetilde{\mathbf{y}}_2)$
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

In Table 3, the simulation-based measures are compared to the analytic results described in Section 5.1. In addition, we included the scalar measures of the input estimator $\widehat{\mathbf{y}}$, to give a better appreciation of the adjustment effect. We observe the following. (I) Given the vector of interest, which can be the initial $\widehat{\mathbf{y}}$, $\widehat{\mathbf{y}}_1$ by (16) or $\widehat{\mathbf{y}}_2$ by (17), the absolute values of τ_2, Δ_1 and Δ_2 differ from each other, which serves as a reminder that the uncertainty measures are to be interpreted relatively between the different vectors, rather than in absolute terms. (II) With $J = 1000$ the simulation-based values of the uncertainty measures are already quite close to the exact ones. We have carried out simulations for larger values of J and observed that the uncertainty measures converge to the corresponding exact values. More importantly, the simulation-based results agree with the exact results even better when the measures are used relatively. For instance, $\Delta_1(\widetilde{\mathbf{y}}_2)/\Delta_1(\widetilde{\mathbf{y}}_1) = 7.42/8.68 = 0.855$ based on the exact values, and it is $7.50/8.74 = 0.858$ based on the simulation-based results. Similarly for the other comparisons. (III) The numerical results confirm the analytic conclusion in Section 5.1, namely, the adjustment method (16) is more uncertain than (17). Indeed, the adjusted vector by (16) is more uncertain than the input vector $\widehat{\mathbf{y}}$, whereas the adjusted vector by (17) is less uncertain than the input vector $\widehat{\mathbf{y}}$; see Appendix A for the analytic details in this respect. Finally, (IV) in Section 1 we defined the relative efficiency as $\tau(\mathbb{A}_1)/\tau(\mathbb{A}_2)$, or $\Delta(\mathbb{A}_1)/\Delta(\mathbb{A}_2)$, which facilitates a choice of adjustment method. Based on the exact results in Table 3, we obtain

$$\frac{\tau_2(\mathbb{A}_2)}{\tau_2(\mathbb{A}_1)} = \frac{54.87}{66.08} < 1, \quad \frac{\Delta_1(\mathbb{A}_2)}{\Delta_1(\mathbb{A}_1)} = \frac{7.42}{8.68} < 1, \quad \frac{\Delta_2(\mathbb{A}_2)}{\Delta_2(\mathbb{A}_1)} = \frac{5.49}{6.61} < 1.$$

The conclusion is the same as with the simulation-based results.

5.4. *Application to the supply and use tables*

The supply and use tables introduced in Section 3 are subject to multiple constraints including the non-negativity of all variables. For the multiple constraint case of the supply and use tables where an optimization approach to adjustment is used, no analytical results are available and we applied the simulation approach to assess the uncertainty measures, with the following details:

- The initial estimates are the values $\widehat{y}_i$, $i = 1, \ldots, 14$ from Table 2. The adjustment objective amounts to satisfying the accounting system (10) where $\mathbf{Hy} = \mathbf{z}$ is given by

$$y_1 + y_2 + (-y_5) + (-y_6) + (-y_7) = 0$$
$$y_3 + y_4 + (-y_8) + (-y_9) + (-y_{10}) = 0$$
$$y_1 + y_3 + (-y_5) + (-y_8) + (-y_{11}) + (-y_{13}) = 0$$
$$y_2 + y_4 + (-y_6) + (-y_9) + (-y_{12}) + (-y_{14}) = 0$$

and all adjusted values $\widetilde{y}_i$ should be positive.

- We assume independent truncated (at zero) normal distributions for $\widehat{y}_i, i = 1 \ldots 14$, with the expectation set at the corresponding observed value, and the variance as in Table 2.
- We consider three weighted least squares schemes: equal, inverse variance and inverse value.
- We set $J = 1000$ as the number of simulations.
- We obtain the benchmark adjustments by the R-package `rspa` (van de Loo 2015). Note that, while the initial estimates, are taken to be independent, due to the application of the constraints to these initial estimates, the adjusted estimates have non-zero covariances.
- We calculate the scalar measures τ_2, Δ_1 and Δ_2 for each adjustment method.

Table 4. Uncertainty measures for different weighting methods for the supply and use tables.

Weighting method	τ_2	Δ_1	Δ_2
Equal	8026.1	156.0	772.2
Inverse variance	7535.7	152.7	786.8
Inverse value	10239.1	175.5	1006.5

Tabel 4 shows that the adjustment method using the inverse of the variance as weights is overall the best performing method, whereas using weights that are not based on the relative accuracy (here the inverse value) can have a considerable adverse impact on the uncertainty measures. Despite the variances have been set arbitrarily in this case, which are not always bigger for the variables with larger initial estimates (Table 2), the illustration demonstrates that generally the weights should be chosen carefully in practice.

5.5. *An application to time series data*

In this section we apply the simulation approach described in Section 5.3 to the benchmarking of time series data Daalmans and Di Fonzo (2014). The data set consists of multiple time series for the EU Quarterly Sector Accounts (EUQSA), flow variables only. The time series are assumed to be independent of each other and should be adjusted to agree with benchmark values. Denote the quarterly series by $\widehat{\mathbf{y}}$ and the yearly series by $\mathbf{z}$, where we have 28 quarterly values and 7 yearly values for 61 independent series. All yearly values are treated as benchmarks for the quarterly values. Di Fonzo and Marini (2011) describe this data in more detail. We apply the simulation approach with the following details:

- Let $\widehat{y}_{ilk}$ be the input quarterly figure, where $i = 1, \ldots, 4$ for quarters, $l = 1, ..., 7$ for years and $k = 1, \ldots, 61$ for time series. Let z_{lk} be the corresponding yearly figures, which are treated as fixed. The benchmarking objective amounts to the accounting system (10) given by

$$\mathbb{A} = [\widetilde{y}_{1lk} + \ldots + \widetilde{y}_{4lk} = z_{lk}; l = 1, ..., 7; k = 1, ..., 61], \text{ with } \widetilde{y}_{ilk} \geq 0.$$

- We assume independent truncated normal distributions of $\widehat{y}_{ilk}$, with the expectation set at the corresponding observed value, and variance equal to $\sigma^2 \widehat{y}_{ilk}$, for some fixed σ^2.
- We consider three different adjustment methods, denoted by Denton-Cholette, Fernandez and Chow-Lin; see e.g.(Denton 1971), (Dagum and Cholette 2006) and (Chow and Lin 1971), for details.
- We set $J = 1000$ as the number of simulations.
- We obtain the benchmark adjustments by the R-package `tempdisagg` (Sax and Steiner 2003).
- We calculate the scalar measures τ_2, Δ_1 and Δ_2 of the account $\mathbb{A}$ for each combination of (l, k) and present their average values in Table 5.

Table 5.   Averaged uncertainty measures for different adjustments methods for the time series data

Adjustment Method	τ_2	Δ_1	Δ_2
Denton-Cholette	25906.1	260.0	25880.2
Chow-Lin	25589.1	255.2	25563.5
Fernandez	25711.4	258.1	25685.7

Figure 2.   Six selected input (preliminary) and adjusted time series

Figure 2 shows 6 selected input (or preliminary) and benchmarked time series. For two series the adjustments make little difference. For the other four series it is not obvious from these plots which method is the least uncertain. However, based on the results (Table 5) by the proposed approach, one may conclude that for these data the Chow-Lin method is the least uncertain according all the three scalar measures, but the difference is rather small in this case. For instance, the ratio between the smallest and the largest average Δ_2 is $25563.5/25904.8 = 0.987$.

6.   Final remarks

In this paper we propose a new approach for measuring the uncertainty of an estimated system of accounting equations. We consider an accounting system as a single entity, and define scalar uncertainty measures that capture both the uncertainty of the inputs estimates and the effects of adjustments: In Section 3 the same optimization method is applied on aggregated supply and use tables for three different weighting schemes, reflecting different uncertainties in the initial estimates; In Section 5.5 we compared different benchmarking methods for time series data. Compared to the use of a variance-covariance matrix of the adjusted estimators involved in an accounting system, our approach has three important practical advantages:

  (i) The scalar measures of different adjustment methods can be directly compared to each other to facilitate a univocal choice among them, as has been demonstrated for a number of different situations in Section 5.
 (ii) A scalar measure is easy to interpret as the expectation of a chosen norm of the error vector.
(iii) A large accounting system (like the Systems of National Accounts) is often divided into multiple (say, K) *satellite* systems. The scalar measures of the K satellite systems can be easily combined to yield a scalar measure of the whole system. For instance, in the case of an L_2 norm like (14), the scalar measures of satellite systems are additive, and the resulting sum is the scalar measure of the accounting system which consists of these satellite systems. In contrast, one cannot directly combine K variance-covariance matrices into that of the whole system, unless the satellite systems can be obtained independently of each other. Even then, combining the K matrices into a single block-diagonal matrix does not make it simpler to interpret, or easier for one to choose among alternative adjustment methods.

To define and compute the uncertainty measures we need to specify three components: the initial estimates, including their joint distribution; the system of accounting equations; and the adjustment method. In simple situations the distribution of the initial estimators can be specified directly. However, generally this is not an easy task. Thus, in order to apply the proposed approach to a large accounting system, the most difficult task may still be the explicit specification of the joint distribution of the input vector and the constraints involved. For this one needs to

work with the subject experts, and there may be other technical issues that need to be resolved. It will require effort that is beyond the scope of this paper. But we have indicated a road map forward.

A.   Analytic results for Section 5.1

Let $\widetilde{y}_k = y_k + (z - \sum_{j=1}^p y_j)\nu_k$, for $k = 1, \ldots, p$. Its expectation and variance are given by

$$E(\widetilde{y}_k) = E(y_k) + E(z - \sum_{j=1}^p y_j)\nu_k = \mu_k + (z - \sum_{j=1}^p \mu_j)\nu_k =: M_k \tag{A1}$$

$$V(\widetilde{y}_k) = (1 - \nu_k)^2 V(y_k) + \nu_k^2 \sum_{j\neq k} V(y_j) = \nu_k^2 \sum_{j=1}^p V(y_j) + V(y_k)(1 - 2\nu_k) =: \widetilde{\sigma}_k^2. \tag{A2}$$

The additive constraint $\widetilde{y}_1 + \cdots + \widetilde{y}_p = z$ implies that the $\widetilde{y}_k$'s are negatively correlated with each other. Provided $V(y_k) = \sigma_k^2 = \sigma^2 \mu_k$, for $k = 1, \ldots, p$, we have

$$V(\widetilde{y}_k) = \nu_k^2 \sigma^2 \sum_{j=1}^p \mu_j + \sigma^2 \mu_k - 2\nu_k \sigma^2 \mu_k. \tag{A3}$$

Notice that $V(\widetilde{y}_k) < V(y_k)$ only if $\sum_{j=1}^p \mu_j < 2\mu_k/\nu_k$. This is not the case given (16), which would require $\sum_{j=1}^p \mu_j < 2p\mu_k$; but it is always the case given (17), since $\frac{\mu_k}{\sum_{j=1}^p \mu_j} \sum_{j=1}^p \mu_j < 2\mu_k$. Moreover,

$$V(\widetilde{y}_{1k}) - V(\widetilde{y}_{2k}) = \sigma^2 \Big(\frac{1}{p^2} \sum_{j=1}^p \mu_j - \frac{2}{p}\mu_k - \Big(\frac{\mu_k}{\sum_{j=1}^p \mu_j}\Big)^2 \sum_{j=1}^p \mu_j + \frac{2\mu_k}{\sum_{j=1}^p \mu_j}\mu_k\Big) \tag{A4}$$

$$= \sigma^2 \Big(\frac{1}{p^2} \sum_{j=1}^p \mu_j - \frac{2}{p}\mu_k + \frac{\mu_k^2}{\sum_{j=1}^p \mu_j}\Big) = \sigma^2 \Big(\frac{\sqrt{\sum_{j=1}^p \mu_j}}{p} - \frac{\mu_k}{\sqrt{\sum_{j=1}^p \mu_j}}\Big)^2 \geq 0. \tag{A5}$$

B. $\tau(\mathbb{A}^*)$ by the deflation adjustment

Apply the Taylor expansion for $f(\widehat{I}_P) = \widehat{I}_V/\widehat{I}_P$ at I_P, where $\widehat{I}_P$ and $\widehat{I}_V$ are independent of each other, and the true I_P and I_V are both assumed to be equal to 1, we have

$$\tau(\mathbb{A}^*) = E\Big[V(\frac{\widehat{I}_V}{\widehat{I}_P}|\widehat{I}_V)\Big] + E\Big[V(\widehat{I}_P|\widehat{I}_V)\Big] + 2E(\widehat{I}_V) - 2E\Big[E(\frac{\widehat{I}_V}{\widehat{I}_P}|\widehat{I}_V)E(\widehat{I}_P|\widehat{I}_V)\Big] + V(\widehat{I}_V)$$

$$\approx E\Big[V(\frac{\widehat{I}_V}{I_P^2}(\widehat{I}_P - I_P)|\widehat{I}_V)\Big] + V(\widehat{I}_P) + 2 - 2E(\widehat{I}_V) + V(\widehat{I}_V)$$

$$= E(\widehat{I}_V^2)V(\widehat{I}_P) + V(\widehat{I}_P) + 2 - 2 + V(\widehat{I}_V)$$

$$= (V(\widehat{I}_V) + 1^2)V(\widehat{I}_P) + V(\widehat{I}_P) + V(\widehat{I}_V)$$

$$\approx (2 + c_1)V(\widehat{I}_P).$$

Notice that in the above we have used the fact that typically $V(\widehat{I}_V)V(\widehat{I}_P) \ll V(\widehat{I}_P)$.

References

Bikker, R. Daalmans, J. and N. Mushkudiani. 2013. "Benchmarking large accounting frameworks: a generalized multivariate model." *Economic Systems Research.*

Boonstra, H.J., de Blois, C.J. and G.-J.M. Linders. 2011. "Macro-integration with inequality constraints: an application to the integration of transport and trade statistics." *Statistica Neerlandica.* 65(4):407–431.

Buono, D., Elliott, D., Mazzi, G. L., Bikker, R., Frölich, M., Gatto, R., Guardalbascio, B., Hauf, S., Infante, E., Moauro, F., Oltmanns, E., Palate, J., Safr, K., Stoltze, P. T. and F. Di Iorio. 2018. "ESS guidelines on temporal disaggregation, benchmarking and reconciliation." *Eurostat.*

Burda, M. and C. Wyplosz. 2017. "Macroeconomics. " 7th Edition. *Oxford University Press, Oxford, United Kingdom.*

Chow, G. C. and A.-L. Lin. 1971. "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series. " *Review of Economics and Statistics* 53, 372-375.

Daalmans, J., Di Fonzo, T., Mushkudiani, N. and R. Bikker. 2018. "GRP temporal benchmarking: drawbacks and alternative solutions." *Survey Methodology.*

Daalmans, J. and T. Di Fonzo. 2014. "Denton PFD and GRP benchmarking are friends. An empirical evaluation on Dutch Supply and Use Tables." *Technical report.*

Dagum, E.B. and P. A. Cholette. 2006. "Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series." *Springer-Verlag, Lecture Notes in Statistics, New York,* 186.

Denton, F.T. 1971. "Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach based Quadratic Minimization." *Journal of the American Statistical Association,* 66, 99-102.

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.

Golan, A. 2018. "Foundations of info-metrics: modeling, inference and imperfect information." *Oxford University Press, New York.*

Golan, A. and S.J. Vogel. 2000. "Estimation of non-stationary social accounting matrix coefficients with supply-side information." *Economic Systems Research.* 12, 447-471.

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

International Monetary Fund. 2017. " Quarterly National Accounts Manual, 2017 Edition." https://unstats.un.org/unsd/nationalaccount/docs/QNAManual2017.pdf.

Jackson, R. and A. Murray. 2004. "Alternative Input-Output Matrix Updating Formulations." *Economic Systems Research.* 16(2), 135-148.

Lahr, M. and L. de Mesnard. 2004. "Biproportional Techniques in Input-Output Analysis: Table Updating and Structural Analysis." *Economic Systems Research.* 16, 115-134.

Loo, M. van der. 2015. "rspa: Adapt Numerical Records to Fit (in)Equality Restrictions." *R-package version 0.1.8.* https://cran.r-project.org/web/packages/rspa.

Magnus, J.R., van Tongeren, J.W. and A.F. de Vos. "National accounts estimation using indicator ratios." *Review of Income and Wealth.* 46, 329–350.

Manski, C.F. 2014. "Communicating Uncertainty in Official Economic Statistics". *Working Paper 20098.* http://www.nber.org/papers/w20098".

de Mesnard, L. 1990. "Biproportional Method for Analyzing Interindustry Dynamics: the Case of France." *Economic Systems Research.* 2, 271-293.

de Mesnard, L. 2004. "On the idea of ex ante and ex post normalization of biproportional methods," *The Annals of Regional Science, Springer;Western Regional Science Association.* 38(4), 741-749.

Miller, R. E. and P. D. Blair. 2009. "InputOutput Analysis Foundations and Extensions." *Cambridge University Press.*

Mushkudiani, N., Daalmans, J., and R. Bikker. 2018. "Solving large-data consistency problems at Statistics Netherlands using macro-integration techniques". *Statistica Neerlandica.* 72, 553-573.

Oberski, D. L., Kirchner, A., Eckman, S. and F. Kreuter. (2017). "Evaluating the Quality of Survey and Administrative Data with Generalized Multitrait- Multimethod Models." *Journal of the American Statistical Association*

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

Rodrigues, J.D.F. 2014. "A Bayesian Approach to the Balancing of Statistical Economic Data." *Entropy.* 16, 1243-1271.

Rodrigues, J.D.F. 2016. "Maximum-Entropy Prior Uncertainty and Correlation of Statistical Economic Data." *Journal of Business & Economic Statistics.* 34:3, 357-367.

Sax, C., and P. Steiner. 2003. "Temporal Disaggregation of Time Series." *The R Journal.* 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 J.E. Maede. 1942. "The Precision of National Income Accounting Estimates." *The Review of Economic Studies.* 9, 111-125.

Temursho, U. 2018. "Entropy-based benchmarking methods". *Statistica Neerlandica.* 72, 421-446.

Wallgren, A. and B. Wallgren. 2014. "Register-Based Statistics Statistical Methods for Administrative Data." (Second ed.). *Chichester: John Wiley & Sons.*

Watrous, J. 2018. "The Theory of Quantum Information." *University of Waterloo, Cambridge University Press.*

Zhang, L.-C. 2012. "Topics of Statistical Theory for Register-Based Statistics and Data Integration." *Statistica Neerlandica.* 66, 41-63.