Reconciling the System of U.S. Accounts and Distribution of the Aggregate Statistical Discrepancy

(First Draft. Please do not circulate)

Baoline Chen

Bureau of Economic Analysis

1441 L Street, NW

Washington, DC 20230

Email: baoline.chen@bea.gov

December 20, 2010

Keyword: Data Reliability, Accounts Reconciliation, GLS Estimation

Abstract

This paper describes and illustrates a generalized least squares (GLS) method that systematically incorporates all available information on relative reliability of initial data in reconciling a disaggregated system of accounts. The GLS method is applied to reconciling the 2002 U.S. Input-Output accounts, GDP-by-industry accounts, and final expenditures from National Income and Product Accounts. The results show that using estimated relative reliabilities of initial data to remove inconsistencies in different sets of accounts produces a statistically meaningful balanced system of accounts. The estimated distribution of the statistical discrepancies by industry and by expenditure category traces the aggregate statistical discrepancy to its sources. The study demonstrates the empirical feasibility and computational efficiency of using a relative reliability-based GLS method to reconcile a large system of national accounts.

1. Introduction

Systems of national accounts are constructed using data from a variety of sources, and, thus, typically contain various types of measurement errors. Initial estimates of national accounts items rarely satisfy all accounting identities and restrictions of the system. The usual balancing procedure is to use accounting identities from different parts of the system to reduce measurement errors as much as possible and to record the residual between the major aggregates. For example, the Bureau of Economic Analysis (BEA) publishes estimates of U.S. Gross Domestic Product (GDP) and Gross Domestic Income (GDI). Although the two estimates are conceptually equivalent, the actual estimates are inconsistent. The residual between estimated GDP and estimated GDI is the aggregate statistical discrepancy. Currently there are no estimates of statistical discrepancy by industry and by expenditure category. Lack of such information hinders a good understanding of the sources of aggregate inconsistency and makes it difficult to identify improvements in source data and estimation methods needed to minimize the statistical discrepancy.

Traditional balancing procedures are purely numerical and often conducted manually. Automatic balancing using numerical procedures, such as the Iterative Proportional Fitting (Raking) method (Deming and Stephan, 1940), are sometimes done in the final stage. Stone et al. (1942) proposed a generalized least squares (GLS) method for reconciling national accounting matrices according to data reliabilities which are determined by measurement errors. The GLS reconciliation method has two empirical advantages. First, it has a firm Bayesian foundation and allows information on relative reliabilities of initial data to be used efficiently in the reconciliation process. Using this method, reconciliation is achieved by trading off relative degrees of uncertainty of data items in the system in order to adjust initial estimates to satisfy accounting constraints. Second, it provides flexibility to the balancing process by allowing, for example, reconciliation to be conducted hierarchically (Dagum and Cholette, 2006) and additional constraints to be easily imposed (Barker et al., 1984).

However, it was not practical to implement the GLS method (Stone et al., 1942) in large systems of accounts due to its large computational requirement. Thus, the RAS method (Bacharach, 1965) became very popular for balancing input-output matrices. The RAS method, however, does not allow varying degrees of uncertainty in initial estimates and constraints, provides dubious economic interpretation of the balancing results, and requires a large number of iterations for convergence. Byron (1978) introduced a more efficient alternative based on the conjugate gradient algorithm and, thus, made it empirically feasible to implement the GLS method in large accounting systems. The GLS reconciliation method has since been further developed (Stone, 1982; van der Ploeg, 1982a, b; Bartholdy, 1991; Weale, 1992), and its feasibility has been demonstrated by van der Ploeg (1982a, 1988) and Barker et al. (1984).

Despite these developments, the GLS reconciliation method has not been widely adopted by national accounting systems since its first inception by Byron (1978) for two reasons. The first obstacle, which has been overcome, was the large required computer memory for reconciling large systems of accounts. The second obstacle was, and still is, insufficient objective information on reliability of initial data. In previous applications, reliability of initial data was set subjectively (van der Ploeg, 1982a; Barker et al, 1984; Beaulieu and Bartelsman, 2004). Reconciliation based on subjectively set reliabilities may lead to incorrectly reconciled accounts and an inaccurately estimated distribution of the aggregate statistical discrepancy. The difficulty in obtaining sufficient information on measurement errors in initial data largely explains why the GLS reconciliation method has not been implemented widely by national accounting systems more than six decades after it was first introduced. The few countries (e.g. Australia, Canada, and UK) that publish reconciled annual estimates of GDP derive it as an average of GDP estimated via production, expenditures, and income approaches.

In a recent study (Chen, 2006) a generalize least squares (GLS) method was applied to reconciling, using the 1997 data, the U.S. industry accounts with GDP estimated via expenditure approach according to the estimated reliabilities of initial estimates in the industry accounts. The study made a first attempt to systematically collect all available objective information on the reliability of initial data and to use that information to reconcile the accounts. In that study, the GLS method produced a balanced system of industry accounts and estimated distribution of statistical discrepancy by industry. However, in that study, initial estimates of final expenditures were considered final and were not adjusted. Consequently, the aggregate statistical discrepancy was distributed entirely to the income side of the industry accounts. Although there were institutional justifications for not adjusting initial estimates of final expenditures, not allowing them to be adjusted implied zero measurement errors in initial estimates of final expenditures. This is an assumption not supported by empirical statistics.

The objective of this study is to extend the GLS method to reconcile the system of U.S. national accounts by allowing all initial data items to be adjusted according to their estimated reliabilities and to estimate distribution of the aggregate statistical discrepancy by industry and by final expenditure category. The GLS method is applied to reconciling the 2002 benchmark U.S. Input-Output (IO) accounts, the GDP-by-industry accounts, and final expenditures from the national income and product accounts at the level of detail of 65 industries (See Appendix) 69 commodity groups, 3 value-added (VA) components, 13 final expenditure categories including exports and imports. Before reconciliation, initial estimates in the IO accounts were not balanced; initial estimates of VA from the IO and the GDP-by-industry accounts were not consistent; and initial estimates of VA from neither set of industry accounts were consistent with the expenditure-based GDP.

The GLS method produced a balanced system of U.S. national accounts. The results show that using estimated relative reliabilities of initial data to reconcile different sets of accounts produces statistically meaningful balanced estimates. The distribution of the statistical discrepancy by industry and by expenditure category properly reflects the sources of the aggregate discrepancy. Moreover, this study demonstrates the empirical feasibility and computational efficiency of implementing the GLD method in a large system of national accounts according estimated relative reliabilities of initial data.

The plan for the paper is as follows. Section 2 discusses the major problems in the 2002 U.S. data. Section 3 describes the GLS method and a sensitivity analysis of the method. Section 4 discusses the reliability of initial data. Section 5 reports and discusses the balanced results. Section 6 concludes the paper.

2. Major Problems in the 2002 Data

U.S. industry accounts measure GDP by industry using production and income data. In the IO accounts, GDP produced by an industry is the VA of that industry, measured as the residual between industry gross output and intermediate inputs. VA of all industries from production must sum to GDP measured via the expenditure approach. In the GDP-by-industry accounts, VA of an industry measures the total income of that industry, and VA estimates of all industries sum to GDI. In theory, GDP estimated from production and income data should be equivalent to expenditure-based GDP. However, actual estimates are not. The presence of inconsistencies in GDP estimated via production, income and expenditure approaches are due to various sources of errors in initial data. There are four major sources of errors identified in initial estimates in the 2002 IO and GDP-by-industry accounts.

The first major source of errors was inconsistencies caused by differences in definitions and classifications of variables, data collection and estimation methods used by different statistical agencies. For the 2002 benchmark IO accounts, initial data on gross output and intermediate inputs were compiled mostly from the 2002 Economic Census and Census related surveys. The GDP-by-industry accounts contain estimates of VA by industry using data on compensation, taxes and subsidies, and gross operating surplus (GOS). The primary source data on compensation, taxes and subsidies were from the Bureau of Labor Statistics (BLS) and BEA. A major portion of data on GOS was from Statistics of Income (SOI) of the IRS. Data from the Federal Reserve Board, other government and regulatory agencies, and private trade companies were also used in both sets of the accounts.

The second major source was sampling and non-sampling errors in the source data. The Census Bureau and SOI provided information on sampling errors for their published estimates in terms of coefficients of variation (CV), computed as the ratio of the standard deviation to the mean. In addition, source data in both sets of the accounts also suffered from non-sampling errors such as double counting, misreporting, omission, misallocation, misspecification, and simple mistakes (ARS, 2005). Non-sampling errors were either due to inconsistencies in definition and classification of variables between statistical agencies and national accounts or the result of problems in data collection and estimation methods used for compiling initial data.

The third major source was errors in various adjustments made by the national accounts to correct non-sampling errors in source data. Adjustments, however, introduced additional uncertainty in initial estimates, because some adjustments were based on studies conducted some years ago and some were estimated judgmentally.

The forth major source was the official residual errors, i.e., the aggregate statistical discrepancy. Recorded as a separate item in the GDP-by-industry accounts, the aggregate statistical discrepancy was a major inconsistency to be removed.

3. Generalized Least Squares Method for Reconciling National Accounts

Following Byron (1996), this section describes a GLS method for reconciling a set of national accounts. Subsection 3.1 describes the GLS method in general. The three inputs to the method are initial estimates of variables to be adjusted, covariances which measure their accuracy, and variables to be held at their initial values and not to be adjusted. Subsection 3.2 discusses available sample information, in the application with the U.S. data in Section 5, for setting the measurement-accuracy covariances. Subsection 3.3 develops a sensitivity analysis of the method with respect to the measurement-accuracy covariances.

3.1 The General GLS Problem

Let α denote the nx1 vector of true, nonstochastic, and unknown values of variables in a linear system of national accounts. The system and α are said to be reconciled when they satisfy the linear accounting system

$$H\alpha = \beta, \tag{1}$$

which imposes m (< n) independent linear constraints on the n variables in α, for a given mxn matrix H and a given mx1 vector β. Independence of the constraints means that H has full row rank m, the elements of H are either 0 or ±1, and in the overall accounting there is usually one more constraint not included in (1) so as to keep H from being singular, so that the overall number of variables is n+1. See Byron (1996, p. 134).

Let α^0 denote an initial, unreconciled, estimate of α, produced by a statistical agency, so that $e_0 = H\alpha^0-\beta \neq 0$. Following Byron (1996), suppose that α^0 is considered a stochastic and unbiased estimate of true α, with positive definite covariance matrix Σ. The GLS method computes an adjusted and reconciled estimate denoted by α^*, which is as close as possible to α^0 in the sense that the weighted sum of squared adjustments, $(\alpha^*-\alpha^0)'\Sigma^{-1}(\alpha^*-\alpha^0)$, is minimal. Given α^0, β, H, and Σ, the GLS problem minimizes

$$S(\alpha^*) = (\alpha^* - \alpha^0)'\Sigma^{-1}(\alpha^* - \alpha^0) \tag{2}$$

with respect to α^*, subject to $H\alpha^* = \beta$. If indeed H has full row rank and Ω is positive definite, then, the problem has the unique solution

$$\alpha^* = \alpha^0 - \Sigma H'(H\Omega H')^{-1}(H\alpha^0 - \beta). \tag{3}$$

If true α is reconciled and initial α^0 is unbiased, then, revised α^* is also an unbiased estimate of true α (Byron, 1996). The idea of weighting by Ω^{-1} in objective function (2) is to induce small adjustment of presumably accurate initial estimates with small variances in an adjusted and reconciled accounting system and vice versa for inaccurate initial estimates. Optimal adjustment rule (3) has the Bayesian interpretation of being drawn from a posterior distribution of α (van der Ploeg, 1982).

Because (3) is based on α^0 being unbiased, one would like to statistically test this assumption. Supposing α^0 is normally distributed, Byron (1996) proposed testing this assumption with the quadratic form $g = (\alpha^*-\alpha^0)'H'(H\Omega^{-1}H')^{-1}H(\alpha^*-\alpha^0)$, distributed chi-squared with m degrees of freedom. As usual, a large and significant value of g rejects the null hypothesis that α^0 is unbiased. Normality is a reasonable assumption for initial estimates in national accounts (Bryon, 1996). In the application in Section 5, g = 159.9 and m = 134 imply that g has a p value of .937, which does not reject unbiasedness of α^0 at the 5% significance level.

3.2 Setting Ω for U.S. National Accounts Data

In the application in Section 5 with U.S. data, there are 134 independent accounting constraints (rows of H) and 10062 variables (columns of H and elements of α). The variables to be adjusted in α are gross output, intermediate inputs, VA, and final expenditures, including exports and imports. Since all variables are included in α to be adjusted, β is zero. The GLS method is applied to the 2002 benchmark input-output, GDP-by-industry, and final expenditures data, with 134 independent accounting constraints in equation (1), reflecting 65 industry and 69 commodity constraints. As usual, there is one more overall constraint equating total industry VA with total expenditure-based GDP, which is excluded from (1) in order to keep H nonsingular, so that α excludes one variable of the overall accounting.

In the application in Section 5, Σ is restricted to be diagonal, with positive diagonal elements. Following Dagum and Cholett (2006), the covariance matrix can be decomposed as

$$\Sigma = \Omega^{\lambda}\Phi\Omega^{\lambda}, \tag{4}$$

where Ω is nxn diagonal matrix with positive diagonal elements, Φ is nxn identity matrix, and λ = ½ or 1. If λ = 1, Σ is diagonal matrix with positive diagonal elements, $\sigma_{ii}^2 > 0$, set to be estimated variances of elements in α^0; if λ = ½, Σ is diagonal matrix with positive diagonal elements, $\sigma_{ii} > 0$, set to be estimated standard errors of elements in α^0.

Ideally, survey data would provide enough information to estimate all elements in Σ. However, U.S. survey data underlying the data used in Section 5 provided only information to estimate variances in Σ. In particular, only CV for IO data from the Census Bureau and CV for VA data from SOI were available and were converted to estimated variances by multiplying them by sample means and squaring. Accordingly, in the application, data were considered inaccurate according to their estimated variances. Restricting Σ to be a diagonal matrix of estimated variances or standard errors is standard practice in reconciliation using data from surveys (Dagum and Cholette, 2006). The sensitivity analysis in the next subsection suggests that specifying Σ to be diagonal is not overly restrictive.

3.3. Sensitivity Analysis for the GLS Method

The following sensitivity analysis provides an upper bound on the size of the adjustment or revision vector, $r = \alpha^{*}-\alpha^{0}$, in terms of the size of the initial-nonreconciliation-error vector, $e = H\alpha^{0}-\beta$, for diagonal Σ and any H. The following mathematical notions needed in the analysis are standard in numerical mathematics and are reviewed in, for example, Golub and Van Loan (1989, ch. 2, pp. 49-85).

For $p \geq 1$, let $||x||_p$ denote the p-norm of nx1 vector $x = (x_1, \ldots, x_n)'$, defined by $||x||_p = (|x_1|^p + \ldots + |x_n|^p)^{1/p}$; and let $||A||_p$ denote the p-norm of mxn matrix A induced by the vector p-norm and defined by $||A||_p$ = maximum of $||Ax||_p$ with respect to x, subject to $||x|| = 1$. For p = 1, $||A||_1$ = maximum of $\sum_{i=1}^{m} |a_{ij}|$ over j = 1, ..., n; for p = 2, $||A||_2$ = square root of the maximum (positive and real) eigenvalue of A'A; for $p = \infty$, $||A||_\infty$ = maximum of $\sum_{j=1}^{n} |a_{ij}|$ over i = 1, ..., m. Because the elements of H are either 0 or ±1, $||H||_1$ = maximum number of nonzero elements in any column of H and $||H||_\infty$ = maximum number of nonzero elements in any row of H. If A and B are compatible with product AB, then, for $p \geq 1$, (i) $||AB||_p \leq ||A||_p \cdot ||B||_p$. The p-norm condition number of a matrix A, denoted by $k_p(A)$ and defined by $k_p(A) = ||A||_p \cdot ||A^{-1}||_p$, satisfies $k_p(A) \geq 1$. Also, (ii) $||A||_2 \leq \sqrt{||A||_1 ||A||_\infty}$.

The 2-norms, inequality (i), and optimal adjustment rule (3) imply that (iii) $||r||_2 \leq k_2(\Sigma) \cdot ||H'||_2 \cdot ||e_0||_2$. For r and e_0 viewed as sample vectors, $s_r^2 = \sum_{i=1}^{n} r_i^2/n = ||r||_2^2/n$ and $s_e^2 = \sum_{i=1}^{n} e_{0i}^2/n = ||e_0||_2^2/n$ are sample second moments. If, as in the application, Σ is diagonal, with positive diagonal elements, and (after permutation, if necessary, without loss of generality) σ_{11}^{λ} and σ_{nn}^{λ} are the largest and smallest diagonal elements, then, $k_2(\Sigma) = \sigma_{11}^{\lambda}/\sigma_{nn}^{\lambda}$. After combining inequalities (ii) and (iii), we obtain the desired inequality

$$s_r/s_e \leq \sqrt{\frac{m}{n} \cdot ||H||_1 \cdot ||H||_\infty} \cdot k_2(\Sigma), \qquad (5)$$

which implies the following about the sensitivity of r to Σ.

First, inequality (5) gives a worst case analysis: for the given norms of H and condition number of Σ, the relative size of the adjustment, s_r/s_e, at most equals the right-side of (5), but could actually be much smaller. Second, optimal r given by (3) is invariant to the size of Σ in the sense of Σ being multiplied by a positive constant. Third, consequently, any diagonal or nondiagonal Σ can be scaled down without changing the GLS problem, so that its implied normal-distribution confidence region of a given probability fits into the normal-distribution confidence region of the same probability implied by the diagonal Σ being considered. In other words, in the GLS problem, any degree of absolute uncertainty of a Σ can be fit into the absolute uncertainty of the diagonal Σ being considered. Finally, in the application in Section 5, $m = 134$, $n = 9165$, $||H||_1 = 2$, and $||H||_\infty = 93$ imply $s_r/s_e \leq 1.64\cdot(\omega_{11}/\omega_{nn})$. In sum, the key implication of inequality (4) of the sensitivity analysis is that the adjustments are expected to be smaller if variables to be adjusted have equal uncertainties and larger if they have very differing degrees of relative uncertainties.

4. Reliability of Initial Data

This section discusses how reliability of initial estimates was determined. Because initial data come from various sources, an initial estimate consists of two parts, source data and adjustments, where source data are from official data collection agencies and adjustments aim to correct non-sampling errors in source data.

In the 2002 data, initial source and adjustment data items were identified by a reliability indicator, denoted, in a decreasing order of reliability, by $\theta = (1, 2, 3, 4, 5)$ (Table B in the appendix defines the 5 categories of reliability indicator). Distributions of initial data items in the 5 reliability categories, as shown in Table 1, vary among the variables. Initial source data items fall mostly into reliability categories 1 and 2, and adjustment data items fall largely into reliability categories 3, 4 and 5.

Reliability of source data is usually measured by estimated variances based on published estimates and CVs, when available, from statistical agencies. Data from the Economic Census have no assigned sampling errors, and, thus, have zero assigned CVs. Administrative data, such as salaries, wages, taxes, and subsidies are provided by regulatory agencies and, thus, are treated like the data from the Economic Census. Initial data from Economic Census and administrative data fall in category 1 of reliability.

Estimating reliability of adjustment data, however, is less straightforward, because there is inadequate information about the uncertainty in these data items. Stone et al. (1942) addressed this issue and suggested that, in the absence of standard errors, margins of error may be set judgmentally by experienced analysts. In this application, subjective margins of errors were systematically assigned based on information reflected from reliability indictors.

Let α^0_{i,A_θ} denote an initial estimate of the ith element in α^0 of category θ that has no CV, and let c denote the average CV of detailed initial data items from surveys conducted by data collection agencies. The subjective CV of α^0_{i,A_θ} is, then, computed as

$$\mathrm{CV}(\alpha^0_{i,A_\theta}) = c\cdot(\theta - 1). \tag{6}$$

For data in the IO accounts, c is the average CV of detailed data items from Economic Census related surveys, and in the GDP-by-industry accounts, c is the average CV of SOI estimates. The subjective standard error of an initial data item is, thus, the product of the subjective CV and the estimate of the data item. Based on the reliability indicator values assigned by experienced analysts, equation (6) is used to differentiate varying degrees of uncertainty in initial adjustment items. Here, the computed subjective CVs are 0, c, 2c, 3c and 4c. However, because adjustment data fall into reliability category 2 to 4, the computed subjective CVs for adjustments are effectively c, 2c, 3c and 4c.

Thus, reliability of an initial estimate in α^0, in the sense of its diagonal element of Σ, is measured by the sum of variances of the source data and the adjustments,

$$\sigma^{2\lambda}_{ii} = \sigma^{2\lambda}_{ii}(\alpha^0_{i,S} + \alpha^0_{i,A}) = (\mathrm{cv}(\alpha^0_{i,S})\cdot\alpha^0_{i,S})^{2\lambda} + \textstyle\sum_{\theta=1}^{5}(c\cdot(\theta-1)\cdot\alpha^0_{i,A_\theta})^{2\lambda}, \tag{6}$$

where $\alpha^0_{i,S}$ and $\alpha^0_{i,A}$ refer to the source data and adjustment components of the i^{th} item in α^0, and $\alpha^0_{i,A} = \sum_1^5 \alpha^0_{i,A_\theta}$. Table 1-b contains simple statistics of the estimated CVs by variable. The average CV for gross output in the IO accounts is the lowest, and that of GOS and final expenditures is much larger. Standard deviation for intermediate inputs and final expenditures is much larger than that for gross output and GOS, reflecting some large outliers in those components.

In Section 5, we contrast the uncertainty measures based on estimated variances and standard errors of initial estimates with a so-called neutral variant, defined as the square of an initial estimate, $\omega_{ii} = (\alpha_i^0)^2$ (Barker et al, 1984). The neutral variant, motivated by the idea that large initial estimates imply large discrepancies in the accounts, has been used previously to reconcile accounts (Beaulieu and Bartelsman, 2004). Reconciling different accounts using neutral variants is equivalent to assuming that all initial estimates have the same degree of uncertainty.

5. A Balanced System of Accounts

This section discusses the application using the 2002 U.S. initial estimates from the IO accounts, GDP-by-industry accounts, and initial estimates of final expenditures from NIPA. Accounts were reconciled according to three weighting schemes: relative reliabilities measured by estimated variances and standard errors of initial estimates, and neutral variants of initial estimates. Subsection 5.1 compares the three sets of balanced estimates at the aggregate level. 5.2 compares balanced estimates at the industry and commodity levels. Subsection 5.3 Discusses adjustments of the variables and the estimated distribution of statistical discrepancy by industry and by expenditure category. Subsection 5.4 discusses the implications of balanced results.

5.1 Balanced Estimates at the Aggregate Level

Tables 2-a and 2-b contain initial and three sets of balanced estimates at the aggregate industry and commodity levels. The first row of Table 2-a shows initial aggregate industry estimates of gross output, intermediate inputs and VA, and initial gap, in level and in percentages, between total industry gross output and total inputs measured as the sum of intermediate inputs and VA. Correspondingly, the first row of Table 2-b shows initial aggregate commodity estimates and initial gap between total commodity gross output and total commodity inputs measured as the sum of intermediate and final uses of commodities. Because initial estimates of final expenditures from NIPA were quite close to initial estimates of final uses in the IO accounts, the initial aggregate commodity gap of .57% is much smaller than the initial aggregate industry gap of 1.52%. Rows 2-4 in both tables show the three sets of balanced estimates at the aggregate level with the percentage adjustments of the variables in the parentheses.

Table 2-a and 2-b are here

We observe three features from balanced estimates at the aggregate. First, reconciliation based on relative reliabilities of initial estimates allows larger adjustments to less reliable components in the system; whereas reconciliation based on neutral variants tend to distinguish adjustments according to their relative sizes. For example, initial estimates of gross output were considered most reliable because they were compiled mostly from Economic Census data, and initial VA estimates, on the other hand, were considered much less reliable, because the GOS component of VA was compiled using a combination of SOI data which had larger CVs and various adjustments which were significant portions of the total GOS estimate. Thus, balanced estimates based on relative reliabilities reflect smaller percentage adjustments in gross output and much larger adjustments in VA. However, balanced estimates based on neutral variants reflect similar absolute percentage adjustments in gross output and VA, and the percentage adjustment in total intermediate inputs is the smallest because it is smallest component of all variables.

Second, compared with reconciliation based on estimated standard errors, reconciliation based on estimated variances generates, in this data set, smaller total adjustments in gross output, intermediate inputs and final uses, but higher adjustment in VA.

Third, allowing initial estimates of final uses to be adjusted, aggregate statistical discrepancy is distributed between final expenditures and VA estimates. Table 2-c shows that using relative reliabilities to reconcile the accounts, the aggregate statistical discrepancy is mostly distributed to VA in the GDP-by-industry accounts. Furthermore, if accounts were reconciled according to estimated variances, statistical discrepancy, in absolute value, distributed to VA is more than 11 times of that distributed to final expenditures. This relative distribution ratio drops to 3.3 if accounts were reconciles using estimated standard errors, and drops to 1.3 if accounts were reconciled using neutral variants.

Table 2-c is here

5.2 Balanced Estimates at Industry and Commodity Levels

Table 3-a and 3-b contain initial and balanced estimates at the 65-industry and 69-commodity levels. In both tables, Panel A shows, respectively, initial estimates of the variables and percentage gaps between initial estimates of gross output and total inputs by industry and by commodity. Panels B, C and D show, respectively, balanced estimates based on neutral variants of initial estimates and relative reliabilities of initial estimates derived from estimated variances and standard errors. As seen at the aggregate level, initial gaps by commodity are much smaller in general than initial gaps by industry, because initial estimates of final uses from NIPA were quite close to those of final uses in the IO accounts.

Table 3-a is here

Initial estimates of each variable are compared with balanced estimates based on the three weighting schemes. Sums of balanced estimates of gross output and intermediate inputs of all industries shown in Table 3-a match those of balanced estimates of gross output and intermediate inputs shown in Table 3-b; sum of balanced estimates of VA shown in Table 3-a matches that of balanced estimates of final uses shown in Table 3-b, indicating that the IO accounts are balanced and that the IO accounts, GDP-by-industry accounts, and final expenditures from NIPA are reconciled.

Table 3-b is here

We observe two significant features of the balanced estimates at the industry and commodity levels from Table 3-a and 3-b. First, if an industry or a commodity has a large initial gap, adjustments will be large for some variables (e.g. industry 334 and 532R; commodity 212 and 335). Second, the sizes of adjustments can be disproportional in gross output, intermediate inputs, and VA or final uses from reconciliation based on relative reliabilities; whereas adjustments tend to be more proportional among the variables for reconciliation based on neutral variants (e.g. industry 335 and 532R; commodity 335 and 482).

5.3 Adjustments in Variables and Distribution of Statistical Discrepancy by Industry and by Expenditure Category

The second feature of balanced estimates at industry and commodity levels are clearly depicted in Figure 1 and Table 4 on percentage adjustments in the variables. In accounts reconciled by relative reliabilities, the absolute means and standard deviations of percentage adjustments are much smaller for gross output and intermediate inputs than for VA and final uses. This is because initial estimates of gross output and intermediate inputs, compiled mostly from surveys from the Economic Census, and a smaller portion of initial estimates were in reliability categories of 3-5, are more reliable; whereas initial estimates of GOS in VA, compiled using SOI estimates which had larger sampling errors, and a larger portion of initial estimates were in reliability categories of 3-5, are less reliable. Similarly, initial estimates of final uses, compiled using data from various sources with most of initial estimates in reliability category 3, were also less reliable than those of gross output and intermediate inputs. By contrast, in accounts reconciled by neutral variants, relative sizes of initial estimates affect sizes of adjustments. Despite large varying degrees of relative reliabilities, differences are much smaller in absolute means and standard deviations of percentage adjustments from reconciliation based on neutral variants.

Table 4 is here

By allowing initial estimates of final expenditure to be adjusted, the aggregate statistical discrepancy is allocated between VA and final expenditures. Tables 5-a and 5-b show, respectively, estimated statistical discrepancy by industry and by final expenditure category. Panel A of Table 5-a shows initial gaps, in levels and percentages, between estimates of VA from IO and VA from the GDP-by-industry accounts. Statistical discrepancies based on neutral variants, relative reliability measured by relative variances and relative standard errors are respectively shown in panels B, C and D in Table 5-a. Two notable features echo those observed from Tables 3-a and 3-b. First, sizes of initial gaps in VA estimates between the two accounts affect sizes of statistical discrepancies distributed by industry (e.g. industry 324 and 532R). Second, if accounts are reconciled according to relative reliabilities, relative variances or relative standard errors of initial estimates determine statistical discrepancies by industry (e.g. 311F and 42); whereas if reconciliation is based on neutral variants, relative sizes of industry VA of total GDP determine the distributions (e.g. industry 313T and 531). Furthermore, statistical discrepancy distributed to each industry tends to be proportional to the estimated variances if reconciliation is based on relative variances; whereas statistical discrepancy tends to be proportional to the estimated standard errors if reconciliation is based on relative standard errors.

Table 5-a is here

Correspondingly, panels A, B and C of Table 5-b show, by expenditure category, distributed statistical discrepancy; statistical discrepancy as a percentage of final expenditure; and statistical discrepancy as a percentage of aggregate GDP, from reconciliation based on the three weighting schemes. The fourth column in panel A displays final expenditure of each category as a percentage of aggregate GDP, and the fourth column of panel C contains CVs of initial estimates by final expenditure category as a reference of their reliabilities. The distribution results in panel A show that if reconciliation is based on neutral variants, statistical discrepancy is distributed largely according to the sizes of final expenditure categories (e.g. F010, F030). By contrast, the results in panels B and C show that statistical discrepancy is distributed largely according to the relative reliabilities of initial estimates, if reconciliation is based on relative variances or relative standard errors (e.g. F030, F06C).

5.4 Significance of Balanced Estimates

The greatest value of the reliability-based GLS method is that it produces a statistically meaningful balanced system of accounts, and, thus, enhances the credibility of national accounts estimates. Balanced results from the reliability-based GLS method should improve the credibility of further data derived from the balanced estimates. For example, estimates of intermediate inputs data are used by BEA to compile annual industry KLEMS data and by the Federal Reserve Board to produce industry productivity indexes. Balanced estimates of intermediate inputs represent more accurate allocations of intermediate inputs across industries, and, hence, should help improve the accuracy of industry KLEMS data and productivity data compiled from them.

Degrees of impact on data derived from balanced estimates may vary. For example, BLS produces industry output and employment projections using estimates of gross output from annual IO accounts at BEA. Because initial estimates of gross output are considered reliable, using balanced estimates of gross output should not significantly affect industry output projections. However, because initial VA estimates are less reliable, the impact should be more significant on statistics compiled using balanced estimates of VA.

Another significant value of the GLS-based reconciliation method is that it helps identify the improvements needed in source data collection and in estimation methods used to compile initial estimates. Furthermore, it should enable a time series of distribution of the aggregate statistical discrepancy to be constructed. Such a time series could track the distribution of the aggregate statistical discrepancy over time and provide information on the improvements in the reliability of source data by industry and by expenditure category.

6. Conclusion

In this paper, a GLS method is applied to reconcile a disaggregated system of accounts using the 2002 U.S. data by allowing all initial data items to be adjusted. The results demonstrate that using estimated relative reliabilities of initial data to remove inconsistencies in different sets of accounts produces statistically meaningful balanced estimates. The distribution of the statistical discrepancies by industry and by final expenditure category properly traces the aggregate statistical discrepancy back to its sources. Moreover, the study also demonstrates that using a GLS method to reconcile a large system of national accounts according to estimated relative reliabilities of initial data is empirically feasible and computationally efficient.

In this study, the GLS method is applied to reconciling the system of U.S. national accounts of a benchmark year. Because Economic Census is conducted only for benchmark years, objective information on the reliability of source data is more available for benchmark years than for off-benchmark years. A future extension of this study is to combine time series analysis methods with the GLS method to balance the system of U.S. national accounts for multiple years.

Reference

ASR Analytics. (2005), "Study and Evaluation of Bureau of Economic Analysis National Income and Product Income Measures," unpublished report, Potomac, MD.

Bacharach, M. (1965), "Estimating Nonnegative Matrices from Marginal Data," ***International Economic Review***, 6, 294-310.

Barker, T., van der Ploeg, F. and Weale, M. (1984), "A Balanced System of National Accounts for the United Kingdom," ***Review of Income and Wealth***, 30(4), 461-485.

Bartholdy, K. (1991), "A Generalization of the Friedlander Algorithm for Balancing of National Accounts Matrices," ***Computer Science in Economics and Management,*** 4, 165-174.

Beaulieu, J. J., and Bartelsman, E. J. (2004), "Integrating Expenditure and Income Data: What to Do with the Statistical Discrepancy?" ***Unpublished paper***, Board of Governors of the Federal Reserve System.

Byron, R. P., (1978), "The Estimation of Large Social Account Matrices," ***Journal of Royal Statistics***, Series A, 141(3), 359-367.

Chen, Baoline, (2006), "A Balanced System of Industry Accounts for the US and Structural Distribution of Statistical Discrepancy," working paper no., Bureau of Economic Analysis, Washington, DC.

Dagum, E. B., and Cholette, A. P. (2006), *Benchmark, Temporal Distribution, and Reconciliation Methods for Time Series, Lecture Notes in Statistics*, Vol. 186, Springer publisher, Berlin, Germany.

Deming, W. E., and Stephan, D. F. (1940), "On the Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known," *Annuals of Mathematical Statistics*, 11, 427-444.
General Algebraic Modeling System (GAMS), GAMS Development Corporation, Washington, DC.

Golub, G., and Van Loan, C. (1989), *Matrix Computations*, 2nd edition, John Hopkins University Press: Baltimore, MD.

van der Ploeg, F. (1982a), "Reliability and the Adjustment of Sequences of Large Systems and Tables of National Accounting Matrices," *Journal of Royal Statistical Society*, Series A 145(2), 169-194.

van der Ploeg, F. (1982b), "Generalized Least Square Methods for Balancing Large Systems and Tables of National Accounts," *Review of Public Data Use.*

van der Ploeg, F. (1988), "Balancing Large Systems of National Accounts," *Computer Science in Economics and Management*, 1, 31-39.

Stone, R., Meade, J. E., and Champernowne, D. G. (1942), "The Precision of National Income Estimates," *Review of Economic Studies*, 9(2), 111-125.

Stone, R. (1982), "Balancing the National Accounts," *Demand Equilibrium and Trade: Essays in Honor of Ivor Pearce*, Macmillan, London.

Weale, M. (1992), "Estimation of Data Measured with Error and Subject to Linear Restrictions," *Journal of Applied Econometrics*, 7(2), 167-174.

Table 1-a: Distribution of Quality Index by Variable

Quality Index	Gross output	Inter. Inputs	Final Uses	Compen-sation	Taxes
θ=1	65.78	25.15	52.45	15.90	0.02
θ=2	9.72	30.39	2.78	81.17	0.00
θ=3	7.23	22.51	44.50	2.87	99.98
θ=4	14.01	10.46	0.20	0.03	0.00
θ=5	3.25	11.49	0.07	0.02	0.00

Table 1-b: Simple Statistics of CV by Variable

	Gross Output	Intermediate Inputs	GOS	Final Uses
Statistics	CV(x_{ij})	CV(z_{ij})	CV(VA_{i3})	CV(y_{dj})
Mean	0.016	0.070	0.091	0.199
Max	0.562	13.029	0.608	1.772
Min	0.000	0.000	0.002	0.000
Stdv	0.041	0.240	0.092	0.492

Table 2-a: Balanced Estimates of Industry Aggregates and Percentage Adjustments in Initial Estimates (Million dollars)

Model	Total Industry Gross Output	Total Industry Intermediate Inputs	Total Industry Value-Added	Total Industry Gap	Total Industry Gap (%)
Initial Estimates	19180034.33	8398244.5	10490589.9	291199.93	1.52
M1 (w=1/NV)	19003230.18 (-0.92)	8410057.06 (0.14)	10593173.11 (0.98)	0	0
M2(w=1/var)	19141485.85 (-0.20)	8484435.61 (1.03)	10657050.24 (1.59)	0	0
M3 (w=1/se)	19120561.11 (-0.31)	8491880.02 (1.11)	10628681.09 (1.32)	0	0

Table 2-b: Balanced Estimates of Commodity Aggregates and Percentage Adjustments in Initial Estimates (Million dollars)

Model	Total Commodity Gross Output	Commodity Intermediate Inputs	Total Final Expenditures	Total Commodity Gap	Total Commodity Gap (%)
Initial Estimates	19180034.33	8398244.5	10671888.8	109901.03	0.57
M1 (w=1/NV)	19003230.18 (-0.92)	8410057.06 (0.14)	10593173.11 (0.98)	0	0
M2(w=1/var)	19141485.85 (-0.20)	8484435.61 (1.03)	10657050.24 (1.59)	0	0
M3 (w=1/se)	19120561.11 (-0.31)	8491880.02 (1.11)	10628681.09 (-0.40)	0	0

Table 2-c: Distribution of Aggregate Statistical Discrepancy between VA and Final Uses (Million dollars and %)

Model	Total Industry Value-Added	Total Final Expenditures	Aggregate Statistical Discrepancy	Change in Total Value-Added	Chang in Total Final Uses	%Change in Total Value-Added	%Change in Total Final Uses
Initial Estimates	10490589.9	10671888.8	181298.9	-	-	-	-
M1 (w=1/NV)	10593173.11	10593173.11	0	102583.21	-78715.69	0.98	-0.74
M2(w=1/var)	10657050.24	10657050.24	0	166460.34	-14838.56	1.59	-0.14
M3 (w=1/se)	10628681.09	10628681.09	0	138091.19	-43207.71	1.32	-0.40

Table 3-a: Initial and Balanced Estimates for 65 Industries
(Million dollars)

	A				B			C			D		
	Initial Estimates				Balanced Estimates (w=1/NV)			Balanced Estimates (w=1/var)			Balanced Estimates (w=1/se)		
Ind ID	Initial Industry gap(%)	Industry gross output	Industry intermediate inputs	Industry VA	Industry Gross Output	Industry Inter. inputs	Industry VA	Industry Gross Output	Industry Inter. inputs	Industry VA	Industry Gross Output	Industry Inter. inputs	Industry VA
111C	0.00	220400	150330	70077	221493	150592	70900	219946	153402	66544	219449	150280	69169
113F	-16.52	50148	33187	25247	52885	30494	22391	53659	31988	21671	52873	29404	23469
211	-5.31	104225	49866	59896	103870	45930	57941	104005	37063	66943	103880	41747	62133
212	-6.14	47896	20060	30778	46412	18245	28167	47932	18759	29173	47723	17739	29984
213	17.24	32399	12992	13822	29858	14342	15517	32009	13489	18520	31491	14403	17087
22	-2.70	318367	126492	200470	330718	129725	200993	332487	128996	203491	326769	126944	199825
23	-6.49	967588	491684	538685	988033	471282	516751	996486	479113	517373	989194	456919	532275
311F	12.61	565284	343425	150572	536578	370535	166043	564969	349437	215533	561144	367669	193475
313T	8.98	75425	48283	20371	71618	50189	21429	75402	48638	26764	74976	50592	24383
315A	-8.14	47189	30536	20492	48599	28966	19632	47321	30105	17216	47506	28525	18982
321	-0.67	89459	59688	30369	89682	59698	29984	89465	59532	29933	89420	59123	30297
322	9.84	151697	89446	47326	146623	95161	51462	151623	91525	60098	150811	94655	56156
323	1.26	99015	49672	48099	100067	50133	49933	98932	49926	49005	98820	50020	48800
324	-14.64	211698	195028	47658	223204	180391	42813	211761	189326	22434	213338	185941	27397
325	7.66	448055	277373	136356	432953	288501	144452	447845	287041	160804	445895	294792	151104
326	8.13	170885	98581	58418	163271	101838	61433	170798	102116	68682	169890	105133	64757
327	6.19	93321	47811	39735	90603	49371	41232	93313	48018	45295	93051	49216	43836
331	1.79	137500	92043	43000	136819	93140	43679	137499	92055	45444	137421	92388	45033
332	3.44	242873	132864	101663	238713	134459	104254	242852	133548	109304	242429	136215	106214
333	5.92	241914	136499	91105	234939	140363	94577	241776	137398	104378	240559	141308	99250
334	24.41	354449	198436	69495	313778	226451	87326	353274	268384	84890	344721	275151	69569
335	-13.78	99414	56404	56706	103623	51382	52241	99421	55951	43471	99749	52842	46907
3361	2.46	461501	335307	114824	455669	338648	117021	461487	336186	125301	461057	339217	121840
3364	0.32	163772	99950	63303	163378	99351	64028	163768	98582	65186	163652	98576	65076
337	11.61	74318	38508	27179	70616	41281	29335	74267	44005	30262	73997	45314	28684
339	-0.59	122735	60555	62906	122455	59647	62808	122734	60295	62439	122611	59579	63031
42	-11.40	884521	355850	629505	914559	326595	587964	886411	343829	542582	896681	325932	570749
44RT	-14.87	1012476	453868	709139	1064011	409733	654279	1038750	392520	646229	1032641	353183	679458
481	-12.45	99323	61040	50647	104322	56899	47424	105821	57520	48301	104020	54666	49354
482	13.49	44385	20602	17796	41114	21853	19260	38670	20602	18068	39038	20602	18436
483	15.70	26885	14560	8104	24636	15767	8869	24712	16389	8323	25149	16705	8444
484	-1.15	206694	111572	97507	205457	109704	95753	203078	108433	94646	204825	108123	96702
485	5.45	31493	11819	17958	30813	12192	18621	30410	11934	18476	31050	12619	18431
486	1.00	22295	14933	7139	22140	15011	7129	22294	15094	7200	22288	15060	7227
487O	-2.33	123197	51049	75016	124556	50189	74367	123270	49980	73289	123287	48796	74491
493	-0.55	42267	16150	26348	42162	15871	26290	42397	16098	26299	41923	15626	26297
511	19.05	253504	98886	106315	232057	110309	121749	250375	103828	146548	246248	117935	128313
512	18.73	84449	32580	36053	77545	36189	41356	84387	32822	51565	83746	36402	47344
513	0.75	522937	238676	280358	521999	237432	284567	522841	241100	281742	522362	240158	282204
514	12.83	94497	32905	49473	88324	34748	53576	92904	33080	59824	91537	35587	55950
521C	-0.29	682952	215051	469853	685838	215066	470773	687443	216587	470855	683421	211118	472303
523	-2.24	287275	121910	171802	284855	117979	166876	287233	120598	166634	287130	116442	170688
524	-2.53	458845	207086	263363	461551	203653	257898	459698	202991	256707	459726	200037	259688
525	-19.08	79481	67006	27643	84144	60169	23975	80207	56335	23872	80936	56368	24568
531	5.42	1806964	496025	1213004	1775802	513350	1262452	1796219	547537	1248682	1792357	532016	1260341
532R	39.74	222857	47111	87179	177103	61930	115172	186520	55913	130607	188526	70080	118446
5411	5.08	205707	50706	144547	201393	51545	149848	202598	51372	151226	203388	54508	148881
5412	6.70	684480	226582	412046	667479	233890	433589	683539	230505	453034	681656	250638	431017
5415	0.45	169285	58206	110314	168879	57528	111351	169035	58081	110954	169340	57142	112198
55	6.79	448154	170911	246810	437645	176472	261173	452124	205314	246811	446019	199169	246849
561	3.54	437435	136081	285890	431164	137693	293471	436990	147996	288994	436528	147678	288850
562	9.37	53367	22235	26132	51480	23412	28068	53363	22640	30723	53260	24317	28943
61	3.79	151438	62206	83488	148241	62350	85892	145065	61475	83590	147064	63114	83950
621	-5.32	523843	185239	366487	532304	176935	355369	524497	165897	358600	525637	165623	360014
622H	2.80	515417	203947	297064	508512	204677	303835	500789	202785	298004	509501	211076	298425
624	-6.81	102093	44159	64888	104173	41729	62444	104148	40174	63974	102943	38753	64190
711A	4.61	86025	33610	48453	83521	33929	49592	85924	36798	49126	85832	35980	49852
713	6.74	76325	27152	44029	74031	27902	46129	76294	31988	44306	76066	31085	44981
721	5.48	145680	48001	89693	142363	48859	93504	143578	50031	93547	143259	51230	92029
722	3.19	398878	179883	206259	390529	179557	210972	387258	178448	208811	393323	184386	208938
81	11.19	456778	188663	217002	433436	200654	232782	430960	200943	230017	444441	217784	226657
GFE	-0.55	82984	18854	64587	84589	18971	65617	84410	19823	64587	83743	19155	64588
GFG	0.00	599971	247099	352873	598731	244262	354469	596891	244018	352873	595806	242932	352874
GSLE	-3.55	161514	98910	68335	166032	98405	67627	165788	97452	68335	165856	97521	68335
GSLG	0.06	1301839	452108	848938	1299283	446531	852752	1299562	450624	848938	1297585	448643	848943
Sum	1.52	19180034	8398245	10490590	19003230	8410057	10593173	19141486	8484436	10657050	19120561	8491880	10628681

Table 3-b: Initial and Balanced Estimates for 69 Commodities
(Million dollars)

	A				B			C			D		
	Initial Commodity Estimates				Balanced Estimates (w=1/NV)			Balanced Estimates (w=1/var)			Balanced Estimates (w=1/se)		
Com ID	Initial Gap (%)	Gross outputs	intermediate inputs	final uses	Gross output	Inter. inputs	Final Uses	Gross output	Inter. inputs	Final Uses	Gross output	Inter. inputs	Final Uses
111C	2.58	214545	166578	42432	215963	174866	41098	214344	169311	45033	213949	170457	43492
113F	1.43	57268	57991	-1541	59552	59936	-384	60246	61226	-980	59224	59800	-577
211	-2.99	89280	185713	-93762	89015	179344	-90329	89062	182895	-93833	88971	182805	-93834
212	15.01	47266	39243	927	45408	43997	1410	47302	46353	948	47084	46131	953
213	2.45	33521	2303	30397	31155	2125	29030	33130	2303	30827	32599	2313	30286
22	-7.72	386593	245806	170630	402331	238977	163354	403798	243613	160185	397825	233687	164137
23	0.53	1032363	147684	879222	1049845	153400	896445	1059720	156975	902745	1050762	159093	891668
311F	0.22	569724	203885	364595	543440	202380	341059	569499	205223	364276	567332	205460	361873
313T	6.10	69113	46205	18695	65575	46855	18720	69103	47201	21902	68930	48266	20663
315A	2.58	44849	12601	31091	46199	12809	33391	44899	12593	32306	45005	12523	32482
321	1.49	88593	90514	-3238	88689	91613	-2923	88597	91800	-3202	88582	91699	-3117
322	0.52	149924	136436	12716	145273	135282	9991	149857	136878	12979	149316	136951	12365
323	-5.33	73844	75107	2673	75303	72966	2337	73843	71313	2530	73850	71592	2258
324	0.59	213910	160806	51833	223401	166311	57090	213956	166313	47644	214760	163396	51364
325	1.36	457643	323221	128200	444681	325420	119261	457530	323538	133992	456717	326687	130030
326	5.03	169821	147549	13724	162268	148907	13362	169786	155356	14430	169361	154227	15134
327	1.25	92058	93097	-2188	89570	92040	-2470	92055	94219	-2164	91948	93964	-2016
331	1.30	140243	161343	-22926	139633	163060	-23427	140242	162520	-22279	140200	162919	-22719
332	2.01	235607	224994	5880	231605	226269	5336	235598	229677	5921	235404	229186	6218
333	1.58	242097	90120	148147	235381	90221	145160	242033	91798	150235	241433	92021	149412
334	0.57	345213	200123	143107	308893	201072	107822	344804	200954	143850	341610	201191	140419
335	4.48	97555	65708	27473	100447	68553	31894	97558	69721	27837	97648	69257	28391
3361	0.65	458714	206260	249490	452607	206887	245720	458705	206058	252647	458502	205797	252705
3364	-0.48	160965	59827	101914	160179	59739	100440	160957	59605	101352	160524	59540	100984
337	-0.09	72924	21721	51266	69510	21030	48480	72918	21302	51615	72836	21262	51574
339	1.48	119847	47912	70161	119286	48128	71158	119847	48169	71678	119788	48757	71031
42	1.91	871529	364679	490238	892446	384460	507985	872245	365892	506353	870961	370841	500120
44RT	-0.63	908295	87275	826763	944216	89174	855043	934521	91732	842789	926838	92405	834433
481	-2.13	102369	36289	68257	107312	37667	69645	108869	36817	72052	107085	38438	68647
482	7.94	42289	28549	10384	39378	29027	10351	36893	26540	10353	37813	27502	10311
483	7.48	27482	3747	21680	25298	3769	21529	25311	3726	21585	25755	3906	21850
484	3.69	212125	118620	85676	210653	122771	87881	208505	122006	86499	210182	122956	87225
485	0.71	40313	19070	20958	39838	19203	20635	39229	18953	20276	39871	19400	20471
486	-0.10	22316	20398	1940	22151	20194	1957	22315	20375	1940	22303	20358	1945
487O	-3.23	117416	101290	19915	119081	99259	19822	117440	97525	19915	117439	97486	19953
493	-2.51	42698	42888	881	42663	41778	885	42829	41947	881	42367	41478	888
511	-0.75	164771	35313	130700	154674	35162	119512	161764	35281	126483	159461	34892	124570
512	4.06	84059	37189	43454	77237	36770	40467	83997	39976	44021	83385	39826	43559
513	0.15	437135	236263	200204	434940	236898	198043	437131	239444	197687	436967	239480	197488
514	3.79	93447	64078	25825	89073	63779	25294	92013	65483	26530	91285	65843	25443
521C	1.70	589117	386689	192389	587871	395322	192549	593618	402537	191082	589812	398145	191667
523	3.10	323928	200014	113865	319531	201329	118202	323873	204590	119282	323649	207681	115968
524	0.80	451911	260919	187365	454403	263383	191020	452750	261159	191591	452687	263864	188823
525	2.66	88019	5430	80244	91993	5800	86193	88090	5502	82588	88210	5984	82226
531	-0.57	1797000	522497	1284684	1768336	514929	1253407	1787116	530510	1256606	1784489	532509	1251981
532R	1.11	298796	184353	111118	263875	165831	98044	262642	161914	100728	265534	161215	104319
5411	1.52	205688	123506	79055	201388	123606	77783	202578	126045	76533	203372	128316	75056
5412	1.06	844388	758004	77452	821209	746289	74920	843255	766163	77091	840555	764269	76286
5415	2.31	267516	104808	156535	261780	106196	155584	267069	107975	159094	266133	110494	155638
55	0.15	440898	405397	34832	430711	397101	33610	444900	410117	34783	439075	404633	34442
561	1.30	444394	408041	30570	438301	408109	30192	443949	413339	30610	443464	413066	30397
562	0.71	60528	49067	11032	58963	48351	10612	60524	49607	10917	60462	49909	10552
61	0.70	197123	19384	176368	194197	19191	175006	193836	19280	174556	193761	18791	174970
621	0.02	545995	21057	524839	553593	21232	532361	546648	21535	525113	547740	21863	525877
622H	0.02	603794	1700	601994	597828	1701	596127	590253	1708	588545	598484	1772	596712
624	-0.16	102708	2053	100822	104685	2098	102587	104762	2060	102702	103537	2115	101422
711A	5.78	78295	36724	37044	75636	38176	37460	78021	39588	38433	77696	40306	37391
713	-0.41	110135	8815	101776	108418	8796	99622	109216	8823	100393	109236	9024	100212
721	0.16	101663	39911	61593	100189	40110	60079	101356	40004	61352	101206	41032	60174
722	1.97	470376	89783	371337	461880	91251	370629	458723	90572	368151	464690	94479	370210
81	0.25	568489	166989	400065	556751	165626	391125	542727	166376	376351	557902	164897	393005
GFE	-6.45	69471	65798	8151	70935	63084	7850	71015	62999	8016	70132	62143	7989
GFG	0.00	590653	0	590653	589318	0	589318	587529	0	587529	586476	0	586476
GSLE	0.03	52060	12184	39860	52333	12386	39947	52627	12199	40428	52660	12814	39846
GSLG	0.00	1042157	0	1042157	1039817	0	1039817	1036744	0	1036744	1038515	0	1038515
S005	2.10	0	-892	892	0	-888	888	0	-890	890	0	-892	892
S006	0.00	0	10556	-10556	5118	16253	-11135	5215	15593	-10378	5215	15990	-10776
S007	3.74	5215	107065	-102045	0	102697	-102697	0	102519	-102519	0	103665	-103665
S008	0.00	0	0	0	0	0	0	0	0	0	0	0	0
Sum	0.57	19180034	8398245	10671889	19003230	8410057	10593173	19141486	8484436	10657050	19120561	8491880	10628681

Table 4: Summary Statistics of Percentage Adjustments by Variable

Statistics	(w=1/NV))	(w=1/var(.))	(W=1/Stdv))
% Adjustment in Gross Output			
Mean	-1.54	-0.53	-0.66
Max	5.87	7.00	5.44
Min	-20.53	-16.31	-15.41
Stdv	4.40	3.34	2.92
% Adjustment in Intermediate Inputs			
Mean	1.03	1.27	2.45
Max	31.46	35.25	48.75
Min	-10.20	-25.68	-22.18
Stdv	6.56	8.34	11.29
% Adjustment in Value Added			
Mean	2.33	4.71	3.08
Max	32.11	49.82	35.87
Min	-13.27	-52.93	-42.51
Stdv	7.50	16.00	10.91
% Adjustment in Final Use by Category			
Mean	-0.37	3.46	1.47
Max	1.25	42.16	18.88
Min	-2.00	-0.74	-0.71
Stdv	1.09	11.67	5.28

Table 5-a: Estimates of Statistical Discrepancy by Industry (Million dollars)

	A		B			C			D		
	Initial VA Estimates		Estimated SD (w=1/NV)			Estimated SD (w=1/var)			Estimated SD (w=1/se)		
Ind ID	Initial VA Gap	Initial VA Gap (%)	Industry SD	Ind. SD as % of VA^0	VA' as % of GDP	Industry SD	Ind. SD as % of VA^0	Relative Variance	Industry SD	Ind. SD as % of VA^0	Relative Std. Error
111C	-7	-0.01	823	1.17	0.67	-3533	-5.31	0.48	-908	-1.31	0.70
113F	-8286	-48.85	-2857	-11.31	0.21	-3577	-16.50	0.01	-1778	-7.58	0.10
211	-5538	-10.19	-1955	-3.26	0.55	7046	10.53	0.39	2237	3.60	0.63
212	-2941	-10.57	-2611	-8.48	0.27	-1605	-5.50	2.82	-794	-2.65	1.68
213	5585	28.78	1694	12.26	0.15	4698	25.37	3.82	3265	19.11	1.95
22	-8595	-4.48	523	0.26	1.9	3021	1.48	0.99	-645	-0.32	1.00
23	-62781	-13.19	-21934	-4.07	4.88	-21312	-4.12	0.03	-6410	-1.20	0.17
311F	71287	32.13	15471	10.27	1.57	64960	30.14	7.15	42903	22.17	2.67
313T	6772	24.95	1058	5.20	0.2	6393	23.89	15.57	4013	16.46	3.95
315A	-3839	-23.06	-859	-4.19	0.19	-3275	-19.02	5.99	-1510	-7.95	2.45
321	-597	-2.01	-385	-1.27	0.28	-436	-1.46	2.89	-72	-0.24	1.70
322	14926	23.98	4136	8.74	0.49	12772	21.25	5.16	8830	15.72	2.27
323	1244	2.52	1834	3.81	0.47	906	1.85	1.41	701	1.44	1.19
324	-30988	-185.90	-4845	-10.17	0.4	-25224	-112.43	5.13	-20261	-73.95	2.27
325	34326	20.11	8096	5.94	1.36	24448	15.20	1.16	14747	9.76	1.08
326	13886	19.20	3015	5.16	0.58	10264	14.94	2.42	6338	9.79	1.55
327	5775	12.69	1497	3.77	0.39	5560	12.27	16.36	4101	9.35	4.05
331	2457	5.40	679	1.58	0.41	2444	5.38	13.92	2033	4.51	3.73
332	8346	7.59	2592	2.55	0.98	7641	6.99	8.17	4552	4.29	2.86
333	14311	13.58	3472	3.81	0.89	13273	12.72	9.63	8145	8.21	3.10
334	86518	55.46	17832	25.66	0.82	15395	18.14	13.92	75	0.11	3.73
335	-13695	-31.84	-4465	-7.87	0.49	-13235	-30.45	45.87	-9799	-20.89	6.77
3361	11370	9.01	2197	1.91	1.1	10477	8.36	2.42	7016	5.76	1.56
3364	519	0.81	725	1.15	0.6	1883	2.89	4.59	1773	2.72	2.14
337	8631	24.10	2156	7.93	0.28	3083	10.19	0.55	1505	5.25	0.74
339	-726	-1.17	-98	-0.16	0.59	-468	-0.75	4.54	125	0.20	2.13
42	-100835	-19.07	-41542	-6.60	5.55	-86923	-16.02	8.50	-58756	-10.29	2.92
44RT	-150531	-26.95	-54860	-7.74	6.18	-62909	-9.73	0.72	-29681	-4.37	0.85
481	-12365	-32.30	-3224	-6.36	0.45	-2346	-4.86	0.05	-1293	-2.62	0.21
482	5987	25.17	1464	8.23	0.18	272	1.50	0.01	640	3.47	0.11
483	4222	34.25	765	9.45	0.08	220	2.64	0.05	340	4.03	0.21
484	-2385	-2.51	-1754	-1.80	0.9	-2862	-3.02	0.37	-806	-0.83	0.61
485	1717	8.73	663	3.69	0.18	518	2.80	0.34	474	2.57	0.58
486	223	3.02	-10	-0.14	0.07	61	0.84	0.20	88	1.21	0.45
487O	-2868	-3.98	-649	-0.87	0.7	-1727	-2.36	2.45	-526	-0.71	1.57
493	-231	-0.88	-58	-0.22	0.25	-49	-0.19	0.41	-51	-0.19	0.64
511	48303	31.24	15434	14.52	1.15	40233	27.45	4.50	21998	17.14	2.12
512	15817	30.49	5304	14.71	0.39	15513	30.08	52.43	11292	23.85	7.24
513	3902	1.37	4208	1.50	2.69	1384	0.49	0.18	1846	0.65	0.42
514	12119	19.68	4103	8.29	0.51	10352	17.30	6.35	6477	11.58	2.52
521C	-1953	-0.42	919	0.20	4.44	1002	0.21	1.64	2450	0.52	1.28
523	-6437	-3.89	-4926	-2.87	1.58	-5168	-3.10	1.89	-1114	-0.65	1.38
524	-11604	-4.61	-5465	-2.08	2.43	-6656	-2.59	0.04	-3675	-1.42	0.20
525	-15168	-121.59	-3669	-13.27	0.23	-3771	-15.80	0.04	-3075	-12.52	0.20
531	97934	7.47	49448	4.08	11.92	35678	2.86	0.27	47337	3.76	0.52
532R	88567	50.39	27994	32.11	1.09	43428	33.25	0.34	31267	26.40	0.59
5411	10454	6.74	5300	3.67	1.41	6678	4.42	1.57	4333	2.91	1.25
5412	45852	10.01	21543	5.23	4.09	40988	9.05	7.33	18971	4.40	2.71
5415	764	0.69	1037	0.94	1.05	640	0.58	65.90	1884	1.68	8.12
55	30433	10.98	14363	5.82	2.47	0	0.00	0.00	39	0.02	0.00
561	15464	5.13	7581	2.65	2.77	3105	1.07	0.18	2961	1.02	0.42
562	5001	16.06	1936	7.41	0.26	4591	14.94	9.44	2812	9.71	3.07
61	5744	6.44	2403	2.88	0.81	102	0.12	0.00	462	0.55	0.04
621	-27883	-8.23	-11118	-3.03	3.35	-7887	-2.20	0.75	-6473	-1.80	0.87
622H	14406	4.63	6770	2.28	2.87	940	0.32	0.01	1361	0.46	0.09
624	-6953	-12.00	-2443	-3.77	0.59	-914	-1.43	0.12	-697	-1.09	0.35
711A	3963	7.56	1139	2.35	0.47	673	1.37	0.14	1399	2.81	0.38
713	5144	10.46	2100	4.77	0.44	277	0.63	0.00	952	2.12	0.06
721	7986	8.18	3811	4.25	0.88	3854	4.12	0.70	2336	2.54	0.84
722	12737	5.82	4713	2.28	1.99	2552	1.22	0.15	2679	1.28	0.39
81	51113	19.06	15780	7.27	2.2	13015	5.66	0.31	9655	4.26	0.56
GFE	-457	-0.71	1030	1.60	0.62	0	0.00	0.00	1	0.00	0.00
GFG	-1	0.00	1596	0.45	3.35	0	0.00	0.00	1	0.00	0.00
GSLE	-5732	-9.16	-709	-1.04	0.64	0	0.00	0.00	0	0.00	0.00
GSLG	793	0.09	3814	0.45	8.05	0	0.00	0.00	5	0.00	0.00
Sum	291200	2.78	102583		100	166460			138091		

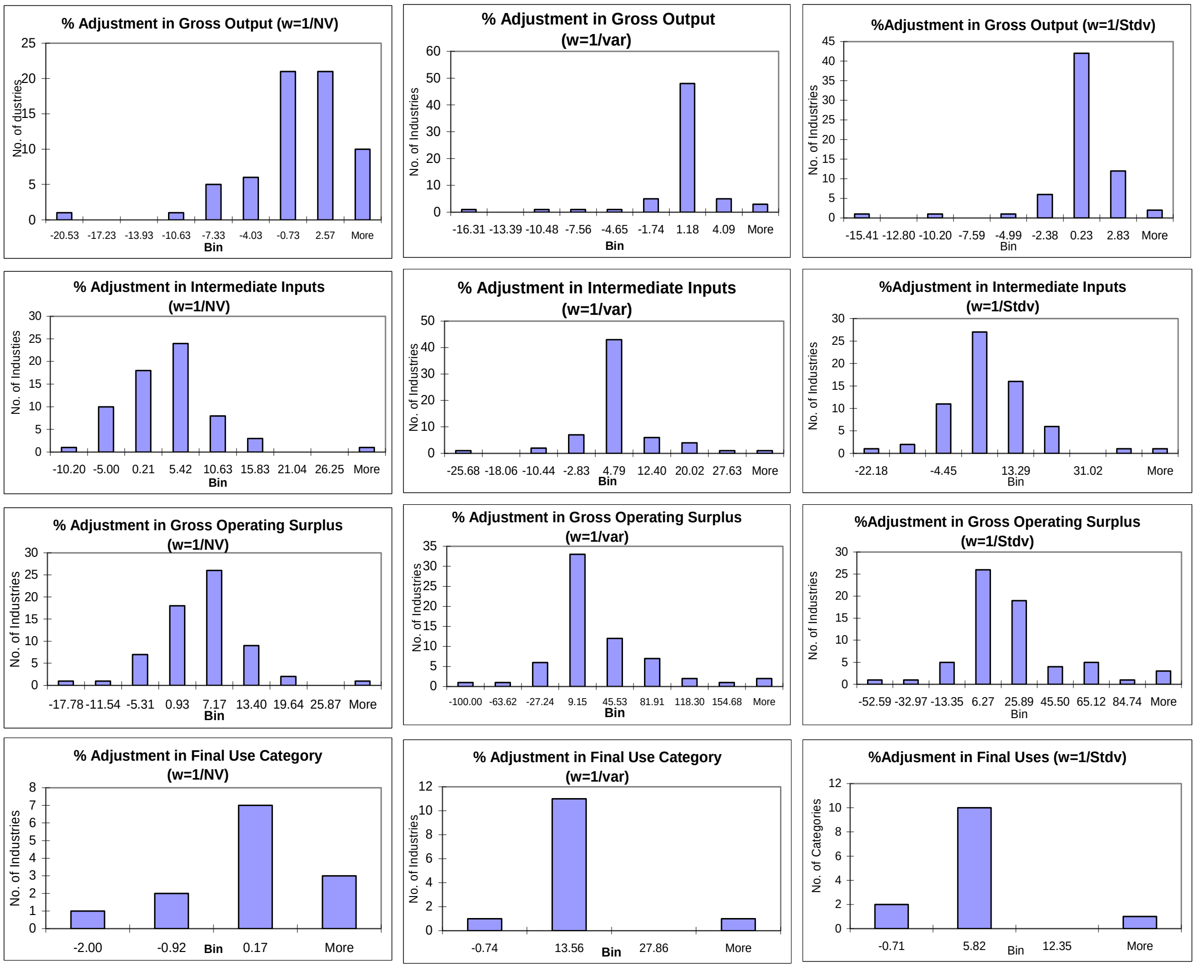
Table 5-b: Estimates of Statistical Discrepancy by Final Expenditure Category
(Billion dollars)

	Balanced Estimates (w=1/abs(a^0))				Balanced Estimates (w=1/var)			Balanced Estimates (w=1/se)			Reliability
Exp Cat. Code	$SD_d^E = y_d^* - y_d^0$	$SD_d^E\% = (y_d^* - y_d^0)/y_d^0$	$(SD_d^E/SD^E)\%$	$(y_d^*/GDP^*)\%$	$SD_d^E = y_d^* - y_d^0$	$SD_d^E\% = (y_d^* - y_d^0)/y_d^0$	$(SD_d^E/SD^E)\%$	$SD_d^E = y_d^* - y_d^0$	$SD_d^E\% = (y_d^* - y_d^0)/y_d^0$	$(SD_d^E/SD^E)\%$	$CV(y_d^E)(\%)$
F010	-42.5	-0.57	83.4	70.02	-41.1	-0.55	276.67	-49.92	-0.67	115.55	1.78
F020	-4.0	-0.25	7.86	15.11	23.5	1.44	-158.41	8.92	0.55	-20.65	5.52
F030	-0.3	-2.04	0.54	0.13	5.8	29.65	-38.85	2.58	15.88	-5.98	213.73
F040	-0.9	-0.23	1.69	8.56	-1.6	-0.41	10.53	-2.69	-0.71	6.23	0.25
F050	-1.1	-1.91	2.07	-12.40	0.0	0.00	0.01	-0.12	-0.22	0.28	0.21
F06C	-0.5	-0.23	0.93	3.58	-1.6	-0.75	10.53	-1.48	-0.71	3.44	20.00
F06I	-0.3	-0.84	0.53	0.53	0.0	0.04	-0.08	0.00	-0.02	0.01	2.41
F07C	-1.1	-0.23	2.17	1.97	-2.7	-0.55	18.24	-2.23	-0.45	5.17	20.00
F07I	0.8	1.12	-1.57	0.32	1.4	1.90	-9.2	1.22	1.70	-2.83	2.38
F08C	-1.2	-0.23	2.42	4.64	-2.7	-0.50	18.24	-1.41	-0.26	3.26	16.12
F08I	2.3	1.24	-4.52	0.69	3.9	2.05	-26.01	1.82	0.98	-4.22	9.64
F09C	-16.1	-1.81	31.65	5.01	0.0	0.00	-0.16	-0.30	-0.03	0.69	9.16
F09I	13.9	-1.04	-27.18	1.86	-0.2	0.02	-1.5	0.41	0.03	-0.94	6.24
Sum	-51.0	-0.54	100	100	-15.3		100	-43.21		100	

Table 5-c: Summary Statistics of Statistical Discrepancy
(Million Dollars)

Statistics	w=1/NV	w=1/var	w=1/SE
Distribution of Statistical Discrepancy by Industry			
Mean	1578	2561	2124
Max	49448	64960	47337
Min	-54860	-86923	-58756
Stdv	12914	19950	13362
Distribution of Statistical Discrepancy by Expenditure Category			
Mean	-705	-1141	-3324
Max	13859	23506	8924
Min	-16139	-41053	-49924
Stdv	6561	13816	14308

Figure 1: Histograms of Percentage Adjustments by Variable



Appendix A: NAICS Industry Codes and Industry Description

Industry	Industry description	Indcode	Industry description
111CA	Farms	486	Pipeline transportation
113FF	Forestry, fishing, and related activities	487OS	Other transportation and support activities
211	Oil and gas extraction	493	Warehousing and storage
212	Mining, except oil and gas	511	Publishing industries (includes software)
213	Support activities for mining	512	Motion picture and sound recording industries
22	Utilities	513	Broadcasting and telecommunications
23	Construction	514	Information and data processing services
311FT	Food and beverage and tobacco products	521CI	Federal Reserve banks, credit intermediation, and related activiti
313TT	Textile mills and textile product mills	523	Securities, commodity contracts, and investments
315AL	Apparel and leather and allied products	524	Insurance carriers and related activities
321	Wood products	525	Funds, trusts, and other financial vehicles
322	Paper products	531	Real estate
323	Printing and related support activities	532RL	Rental and leasing services and lessors of intangible assets
324	Petroleum and coal products	5411	Legal services
325	Chemical products	5412OP	Miscellaneous professional, scientific and technical services
326	Plastics and rubber products	5415	Computer systems design and related services
327	Nonmetallic mineral products	55	Management of companies and enterprises
331	Primary metals	561	Administrative and support services
332	Fabricated metal products	562	Waste management and remediation services
333	Machinery	61	Educational services
334	Computer and electronic products	621	Ambulatory health care services
335	Electrical equipment, appliances, and component	622HO	Hospitals and nursing and residential care facilities
3361MV	Motor vehicles, bodies and trailers, and parts	624	Social assistance
3364OT	Other transportation equipment	711AS	Performing arts, spectator sports, museums, and related activities
337	Furniture and related products	713	Amusements, gambling, and recreation industries
339	Miscellaneous manufacturing	721	Accommodation
42	Wholesale trade	722	Food services and drinking places
44RT	Retail trade	81	Other services, except government
481	Air transportation	GFE	Federal government enterprises
482	Rail transportation	GFG	Federal general government
483	Water transportation	GSLE	State and local government enterprises
484	Truck transportation	GSLG	State and local general government
485	Transit and ground passenger transportation		

Appendix B: Description of Quality Index in IO Accounts

θ	Description
1	2002 Economic Census or USDA data; regulatory data
2	Economic Census related Surveys, Business Expenditure Survey; SOI data; Economic Census data with adjustments
3	Non-Economic Census data from BEA or trade companies
4	Adjusted Census data, Misreporting, etc
5	Analysts judgments, other adjustments, data extrapolated from 1997 benchmark