

Implementing a Reconciliation and Balancing Model in the U.S. Industry Accounts*

Thomas F. Howells III
Edward T. Morgan
Dylan G. Rassier[†]
Conrad E. Roesch

U.S. Department of Commerce
Bureau of Economic Analysis
Industry Accounts
Washington, DC

June 7, 2007

(Very preliminary: Please do not cite without permission.)

* We thank Baoline Chen for contributing motivation and knowledge to this project through her previous study to reconcile and balance the 1997 industry accounts (Chen, 2006). We also thank Karen Horowitz, Douglas Meade, Mark Planting, and George Smith for providing early direction, leadership, and advice to the project. We express special appreciation to Nicholas Empey for his support and assistance with the benchmark database.

This paper has been prepared for presentation at the 2007 NBER Conference on Research in Income and Wealth in Cambridge, MA, July 16 – 17, 2007. We thank Baoline Chen, Sumiye Okubo, George Smith, Erich Strassner, and Mary Streitwieser for helpful comments. All errors and omissions are our own.

The views expressed in this paper are solely those of the authors and not necessarily those of the U.S. Bureau of Economic Analysis or the U.S. Department of Commerce.

[†] Corresponding author: Dylan Rassier, U.S. Bureau of Economic Analysis, Industry Accounts (BE-40), 1441 L St. NW, Washington DC 20230; e-mail dylan.rassier@bea.gov; phone 202-606-9575; fax 202-606-5366.

Abstract

As part of the U.S. Bureau of Economic Analysis' integration initiative (Yuskavage, 2000; Moyer et al., 2004a, 2004b; Lawson et al., 2006), the Industry Accounts Directorate is drawing upon the Stone method (Stone et al., 1942) and Chen (2006) to reconcile the gross operating surplus component of value-added from the 2002 expenditure-based benchmark input-output accounts and the 2002 income-based gross domestic product-by-industry accounts. The objective of the reconciliation is to use information regarding the relative reliabilities of underlying data in both the benchmark input-output use table and the gross domestic product-by-industry accounts in a balanced input-output framework in order to improve intermediate input estimates and gross operating surplus estimates in both accounts. Given a balanced input-output framework, the Stone method provides a tool for balancing the benchmark use table in addition to improving estimates.

This paper presents continuing work by the Industry Accounts Directorate to develop and implement the reconciliation and balancing model. The paper provides overviews of the benchmark use table and gross domestic product-by-industry accounts, including features of external source data and adjustment methodologies that are relevant for the reconciliation. In addition, the paper presents the empirical model that the Industry Accounts Directorate is building and briefly describes the technology used to solve the model. Preliminary work during development of the model shows that reconciling and balancing a large system with disaggregated data is computationally feasible and efficient in pursuit of an economically accurate and reliable benchmark use table and gross domestic product-by-industry accounts.

JEL Codes: C02, C13, C67, C88, E10

Keywords: mathematical methods; estimation; input-output models; other computer software; aggregative models

1. Introduction

In a series of papers beginning with Stone et al. (1942), Stone (1961, 1968, 1970, 1975, 1976, and 1984) advocates a generalized least squares framework to improve the economic and statistical accuracy of independent estimates of national income and expenditures based on relative reliabilities of underlying data used to construct the estimates. Researchers have subsequently revised the Stone method to facilitate its implementation (Byron, 1978; van der Ploeg, 1982, 1984), but federal agencies responsible for producing national economic accounts have generally not implemented the method. The primary reason for this has been a lack of technology that is typically required to solve the complex systems of equations faced by federal agencies. Another reason has been a lack of information regarding the relative reliabilities of underlying data used to construct national accounting estimates. While some agencies have resolved the latter challenge with subjective measures of relative reliabilities (Mantegazza and Pisani, 2000; Moyer et al., 2004a, 2004b), a lack of adequate technology has until recently continued to prevent the implementation of the Stone method (Nicolardi, 2000; Tuke and Aldin, 2004). In a recent study, Chen (2006) addresses both challenges by building an empirical model based on the Stone method and incorporating statistical measures of relative reliability in the model to reconcile and balance estimates in the U.S. industry accounts.

As part of the U.S. Bureau of Economic Analysis' (BEA's) integration initiative (Yuskavage, 2000; Moyer et al., 2004a, 2004b; Lawson et al., 2006), the Industry Accounts Directorate is building an empirical model based on the Stone method and Chen (2006) in order to reconcile the gross operating surplus component of value-added from the 2002 expenditure-based benchmark input-output accounts and the 2002 income-based gross domestic product

(GDP)-by-industry accounts.¹ In addition to reconciling gross operating surplus, the model provides a tool for balancing the 2002 benchmark use table.²

The objective of the reconciliation is to use information regarding the relative reliabilities of underlying data that are used to estimate intermediate inputs in the benchmark use table and gross operating surplus in the GDP-by-industry accounts in a balanced input-output framework in order to improve intermediate input estimates and gross operating surplus estimates in both accounts. In the benchmark use table, gross operating surplus is derived as a residual between the industry estimates of gross output and intermediate inputs, compensation, and taxes. Intermediate inputs in the benchmark use table are derived from external source data and BEA adjustment methodologies.³ In the GDP-by-industry accounts, gross operating surplus is derived from the industry estimates of the gross domestic income (GDI) components of the national income and product accounts as well as BEA adjustment methodologies. Similar to gross operating surplus in the benchmark use table, intermediate inputs in the GDP-by-industry accounts are derived as a residual between the industry estimates of gross output and value-added. The reconciliation sets a new best level for gross operating surplus by industry, which will be published in the 2002 benchmark use table.⁴

¹ Note that the term *GDP* typically refers to expenditure- or production-based estimates. However, estimates for the GDP-by-industry accounts are derived from the industry estimates of the income-based gross domestic income estimates in the national income and product accounts. Thus, GDP-by-industry accounts are truly income based.

² The U.S. benchmark input-output accounts are composed of several tables, including the make table, the use table, and supplementary tables (Lawson et al., 2002). This paper focuses on the benchmark use table, which shows commodities consumed by industries and final users as well as value-added contributed by industries. The benchmark make table has been balanced separately using a different methodology. The supplementary tables are outside the scope of this paper.

³ Likewise, gross output, compensation, and taxes in the benchmark use table are derived from external source data and BEA adjustment methodologies. However, the reconciliation model only adjusts intermediate inputs and gross operating surplus, so this paper focuses only on the external source data and BEA adjustment methodologies used to derive intermediate inputs and gross operating surplus.

⁴ In contrast, a prior reconciliation based on a different model set a best level for *value-added* in the 1997 benchmark use table for use in the GDP-by-industry accounts (Moyer et al., 2004a, 2004b).

An empirical model based on the Stone method has five advantages over past models used by the Industry Accounts Directorate. First, the Stone method yields a model that is transparent. The method is widely researched and familiar to national economic accounting agencies as well as national statistical agencies. In addition, models based on the Stone method are easy to describe to users. Second, the Stone method is built on a firm statistical foundation. In particular, the method uses information on the relative reliabilities of initial underlying data from which estimates are derived to make adjustments to the initial estimates in a balanced accounting framework. Thus, a model based on the Stone method yields a set of fully balanced accounts with adjustments that reflect the reliability of the initial data (Chen, 2006). Third, the least squares framework of the Stone method guarantees that adjustments made to initial estimates are as small as necessary to remove discrepancies between estimates and to satisfy the accounting constraints of the model. In this way, the method yields final estimates that are consistent with the economic concepts upon which the accounts are built. Fourth, the Stone method yields a model that is flexible (Chen, 2006). The model can be implemented with aggregated data or disaggregated data. The objective function can be easily altered to include the reliability-weighted squared differences of additional variables. Constraints can be easily removed or imposed. Finally, the Stone method yields a model that is replicable. If no changes are made to data that are introduced to the model, the model yields a duplicate set of results. Alternatively, updated data can be introduced to the model without requiring any substantial changes to the model.

This paper presents continuing work by the Industry Accounts Directorate of the BEA to develop and implement an empirical model based on the Stone method and Chen (2006) to reconcile gross operating surplus estimates in the 2002 benchmark use table and 2002 GDP-by-

industry accounts as well as to balance the benchmark use table. In this spirit, the paper presents the empirical model and briefly describes the technology used to solve the model. Preliminary work during development of the model shows that reconciling and balancing a large system with disaggregated data is computationally feasible and efficient in pursuit of an economically accurate and reliable benchmark use table and GDP-by-industry accounts.

Section 2 of the paper discusses the benchmark use table, and section 3 discusses the GDP-by-industry accounts. In addition to providing overviews of each product, each section discusses features of external source data and BEA adjustment methodologies that are relevant for the reconciliation. Section 4 presents the empirical model as well as a discussion of the construction of reliability measures in the absence of a complete variance-covariance matrix required by theory to obtain an efficient estimator. Section 5 briefly describes the technology used to solve the model, and section 6 discusses the next steps in the implementation strategy.

2. The U.S. Benchmark Input-Output Accounts

The benchmark input-output accounts are prepared quinquennially and provide a comprehensive picture of the flows of goods and services across all consuming industries and final use categories. The accounts are presented in a series of tables. The use table incorporates a complete, balanced set of economic statistics and presents a full accounting of industry and final use transactions. In addition, the use table provides gross output estimates for industries and commodities, value-added estimates by industry, intermediate consumption by industry, a cross check for the variety of data used to estimate the national income and product accounts, and estimates of final consumption for the national income and product accounts that incorporate the best data available and are in balance with industry inputs and outputs. Chart 1 provides an overview of the structure of the benchmark use table.

Intermediate input estimates in the benchmark use table are derived using expense data from the Census Bureau, data from non-Census sources, and BEA adjustment methodologies. Approximately 71 percent of the value of intermediate inputs comes from Census expense data, 22 percent of the value comes from non-Census data, and 7 percent of the value comes from BEA adjustment methodologies.

Expense data from Census for 2002 are more comprehensive and improved than were data available for previous benchmark use tables in the manufacturing sectors (NAICS 31 – 33) and the service sectors (NAICS 42 – 81).⁵ Within the manufacturing sectors, nineteen categories of expenses are available from the 2002 Economic Census.⁶ Within the service sectors, nineteen categories of expenses are available from the 2002 Business Expenses Survey (BES).⁷ Given this depth of coverage in the manufacturing sectors and service sectors, initial gross operating surplus estimates in the benchmark use table should be much improved in those sectors both before and after the reconciliation because gross operating surplus is calculated as a residual: industry gross output less intermediate inputs, compensation, and taxes.

⁵ However, intermediate input estimates continue to be derived entirely or partially from non-Census data and BEA adjustment methodologies for the following sectors: agriculture, forestry, fishing, and hunting (NAICS 11), mining, utilities, and construction (NAICS 21-23), transportation and warehousing (NAICS 48-49), information (NAICS 51), finance and insurance (NAICS 52), real estate (NAICS 53), and education (NAICS 61).

⁶ Expense categories from the 2002 Economic Census for manufacturing sectors include the following: (1) accounting, auditing, and bookkeeping, (2) advertising, (3) communications, (4) computer services, (5) contract labor, (6) depreciation, (7) electricity, (8) fuels, (9) legal services, (10) management, consulting, and administrative services, (11) materials consumed, (12) other expenses, (13) payroll and fringe benefits, (14) refuse removal, (15) re-sales, (16) rental of buildings and other structures, (17) rental of machinery and equipment, (18) repairs and maintenance of buildings and/or machinery, and (19) taxes and license fees.

⁷ Expense categories from the 2002 BES for service sectors include the following: (1) accounting, auditing, and bookkeeping, (2) advertising, (3) communications, (4) contract labor, (5) data processing and other purchased computer services, (6) depreciation and amortization charges, (7) insurance, (8) lease and rental payments, (9) legal services, (10) packaging and containers, (11) printing, (12) management, consulting, and administrative services, (13) materials and supplies not for resale, (14) other expenses, (15) payroll and fringe benefits, (16) repair and maintenance, (17) taxes and license fees, (18) transportation, shipping, and warehouse services, and (19) utilities. In addition to payroll from the 2002 BES, payroll is available from the 2002 Economic Census, which is considered more reliable than the 2002 BES. Thus, the benchmark data include payroll from the 2002 Economic Census.

Expense categories provide a basis for each industry's input category control estimates in the benchmark data.⁸ However, adjustments to the input category control estimates similar to adjustments that are made to derive each industry's estimate of gross output are required in order to maintain equality among industry inputs and outputs. In particular, adjustments are made for non-employer expenses, misreporting and non-filing, and auxiliary services.⁹

The Economic Census only collects detailed data on outputs and inputs for establishments with employees and payroll. Thus, BEA makes an adjustment to capture outputs and inputs for non-employer establishments. Census provides receipts for non-employers derived from administrative data, which are used to estimate the gross output of non-employers.¹⁰ For each industry, the ratio of inputs to outputs is assumed to be the same for employers and non-employers. Thus, expenses for non-employers are estimated in a given industry by allocating total receipts for non-employers based on the ratio of inputs to outputs for employers in the industry, with the residual being allocated to gross operating surplus. Given the reasonable underlying assumption and the reliable source data for non-employers, the adjustment for non-employer expenses is considered to be more reliable than other adjustments in the benchmark data but less reliable than data from the Economic Census.

Unlike non-employer establishments, small employers are included in estimates published for the Economic Census. However, similar to non-employers, Census derives estimates for small employers from administrative data. Administrative data for non-employers

⁸ Input category control estimates are used to guide transactions for industries' purchases of detailed items for intermediate uses.

⁹ While other adjustments are made to source data for intermediate inputs, these three adjustments capture the majority of the dollar value of adjustments made in the benchmark data.

¹⁰ Census provides receipts for non-employers at a more aggregated level than receipts for employer establishments. To distribute receipts for non-employers to the appropriate industry level, BEA uses a ratio of receipts for small-employer establishments by industry to total receipts for all small employers from the Economic Census. However, these adjustments do not affect the data for the reconciliation because gross output is fixed in the reconciliation model.

typically include individual income tax returns, and administrative data for small employers typically include individual, partnership, and corporate income tax returns. Given these administrative data, BEA adjusts estimates for non-employers and small employers for misreporting and non-filing.¹¹ A misreport results when a tax return is filed with incomplete or incorrect information. Non-filing results when a business or individual who earns income fails entirely to file a return. An adjustment for misreporting is based on data from two Internal Revenue Service (IRS) programs: the Taxpayer Compliance Measurement Program (TCMP) and the TCMP-Information Returns Program (TCMP-IRP).¹² The most recent year for which data are available from a TCMP study is 1988. An adjustment for non-filers is based on data from an exact-match study conducted by Census.¹³ The most recent year for which data are available from an exact-match study at the time the 2002 benchmark use table is being composed and the 2002 GDP-by-industry accounts were published is 1996.¹⁴ In addition, data from the TCMP, TCMP-IRP, and exact-match studies are provided at a more aggregated level than the industry level in the benchmark use table. Given the infrequencies with which the TCMP, TCMP-IRP, and exact-match studies are conducted and the need to approximate the industry distribution of the adjustments, adjustments based on data from the studies are considered less reliable than other adjustments.

¹¹ According to Internal Revenue Service compliance studies, three components contribute to a tax gap between the amount taxpayers should pay and the amount taxpayers actually pay in a timely manner: non-filed returns, underreported income, and underpaid taxes (Brown and Mazur, 2003).

¹² Discontinued in the early 1990s, the TCMP was an audit program designed to study compliance patterns and levels of misreporting among sole proprietors from a stratified random sample of individual tax returns. The TCMP-IRP was designed to compare the results of the TCMP audits to information returns (e.g., W-2, 1099-INT, etc.) filed with the IRS in order to capture misreporting that TCMP auditors failed to find.

¹³ An exact-match study compares records from the Current Population Survey (CPS) to records from the IRS in order to identify and estimate non-filed income for individuals who report income in the CPS but do not file a return with the IRS.

¹⁴ Data for more recent years will be incorporated in the next comprehensive revision of the national income and product accounts and will also be included in the revised version of the 2002 benchmark use table used for the comprehensive revision of the annual industry accounts.

BEA makes an adjustment for auxiliary services in order to allocate sector-level expense data to the industries served by an auxiliary.¹⁵ For a given type of auxiliary, such as central administrative offices, Census provides expense data tabulated for the sector served by the auxiliary. For each industry, the ratio of industry-level payroll to sector-level payroll is assumed to be the same as industry-level expense to sector-level expense. The adjustment draws upon this assumption and industry-level payroll data to allocate the sector-level expense data to the industries served. Additions to intermediate inputs in a given industry served are offset by reductions to gross operating surplus in that industry because gross output is not affected by the adjustment. The adjustment for auxiliary services is considered more reliable than the misreporting and non-filing adjustment but less reliable than the adjustment for non-employer expenses because of the source data and the relatively strong assumption regarding payroll and other expenses.

Before gross operating surplus in the benchmark use table is reconciled with the GDP-by-industry accounts, all benchmark estimates are finalized and reviewed to ensure the data submitted to the model are as complete as possible and consistent with the economic concepts upon which the benchmark use table is built. In particular, total gross industry output over all industries must equal total gross commodity output over all commodities, controls for the sum of intermediate inputs and value-added components must equal industry gross output, and commodity gross output must be as fully allocated as possible to intermediate inputs. Likewise, the commodity distributions of final use categories are reconciled with the national income and product accounts, Census-based compensation estimates are scaled to match total compensation over all industries in the national income and product accounts, and the industry distribution of

¹⁵ An auxiliary is an establishment that provides to other establishments services that may not be part of the economic activities of which the serviced establishment is primarily engaged. These services are typically support and administrative services but can also include professional services.

taxes is made equal to those in the national income and product accounts. In addition, estimates for intermediate inputs, trade margins, and transportation costs must be non-negative. Furthermore, input-output coefficients and the ratio of value-added by industry to total GDP are examined for unreasonable changes from those in the 1997 benchmark use table. Finally, the industry estimates of initial residual gross operating surplus from the benchmark data are compared with the industry estimates of gross operating surplus from the GDP-by-industry accounts, and total value-added from the GDP-by-industry accounts is compared to total final uses as reconciled with the national income and product accounts.

3. The U.S. Gross Domestic Product-by-Industry Accounts

The GDP-by-industry accounts focus on the industry shares of value-added to total GDP as well as contributions of industries to GDP growth and provide an annual time series of estimates for gross output, intermediate inputs, and value-added by industry (Moyer et al., 2004a). Intermediate inputs in the GDP-by-industry accounts are derived as a residual between the industry estimates of extrapolated gross output and value-added. Gross output in the GDP-by-industry accounts is calculated using annual survey data to extrapolate best-level gross output detail from the make table in the most recent benchmark input-output accounts. Best-level estimates of the compensation and tax components of value-added in the GDP-by-industry accounts are derived from the GDI components of the national income and product accounts, and these best-level estimates are used directly. Gross operating surplus in the GDP-by-industry accounts is derived by extrapolating best-level estimates of gross operating surplus from the most recent benchmark use table using the sum of annual GDI-based components of gross operating surplus as an indicator.¹⁶ In addition to being utilized as annual indicators, the GDI-

¹⁶ Best-level estimates of gross operating surplus from the most recent benchmark use table are derived from a prior reconciliation to set best-level value-added (Moyer et al., 2004a, 2004b). In particular, value-added in the GDP-by-

based estimates of gross operating surplus are used as inputs for the reconciliation model. Thus, estimates of 2002 gross operating surplus used in the reconciliation model are different from estimates of gross operating surplus published in the 2002 GDP-by-industry accounts.

GDI-based estimates of gross operating surplus for private, non-farm industries are derived using data from the IRS's Statistics of Income Division (SOI), data from non-SOI sources, and BEA adjustment methodologies.¹⁷ Approximately 47 percent of private, non-farm industries' gross operating surplus comes from the SOI data, 14 percent comes from the misreporting and non-filing adjustment, 22 percent comes from concept and coverage adjustments, and 17 percent comes from non-SOI data and related adjustments.¹⁸ In addition, the SOI data are collected and classified on a company basis. BEA makes an adjustment to convert these company-based data to an establishment basis, which is the production unit used for both the benchmark input-output accounts and the GDP-by-industry accounts. This adjustment shifts gross operating surplus from one industry to another with no impact on total gross operating surplus. Company-based and establishment-based industry estimates differ by an absolute average of 7.7 percent.

industry accounts was built from GDI components and BEA adjustment methodologies and reconciled with value-added from the 1997 benchmark use table. The resulting best-level value-added includes best-level estimates of compensation, taxes, and gross operating surplus, which are extrapolated in the GDP-by-industry accounts. In order to fit the sum of best-level estimates of compensation, taxes, and gross operating surplus with best-level value-added, estimates of gross operating surplus were adjusted as necessary. Thus, best-level estimates of gross operating surplus from the most recent benchmark use table are a residual determined from best-level estimates of compensation, taxes, and value-added. Gross operating surplus is composed of the following components: corporate profits before tax, corporate capital consumption allowance, corporate net interest, corporate inventory valuation adjustment, business transfer payments, proprietors' income, non-corporate capital consumption allowance, non-corporate net interest, proprietors' inventory valuation adjustment, rent and royalties, rental income of persons, and surpluses of government enterprises.

¹⁷ Gross operating surplus estimates for farm and general government industries and the owner-occupied housing portion of gross operating surplus estimates in the real estate industry are not reconciled using the model presented in this paper. Thus, a detailed discussion of the data and methodologies used to prepare estimates for these industries is outside the scope of this paper.

¹⁸ NIPA tables 7.13, 7.14, 7.16, and 7.17 provide additional detail on the relationship between income components as reported by the IRS and as published by BEA.

Similar to the benchmark input-output accounts, BEA also makes a misreporting and non-filing adjustment for non-corporate business income tax-based source data in the GDP-by-industry accounts based on data from the TCMP, TCMP-IRP, and exact-match studies. Thus, the misreporting and non-filing adjustment is considered less reliable than other adjustments. The same assessment regarding the relative reliability of the adjustment is made in both the benchmark data and GDP-by-industry data.

Concept adjustments are a collection of adjustments designed to convert tax accounting-based concepts in the SOI data to economic accounting-based concepts consistent with national accounts. Concept adjustments include the removal of capital gains and dividends from business income. Coverage adjustments are designed to include the activities of entities that contribute to gross domestic product but are not required to file a return with the IRS. Coverage adjustments include adding income earned by Federal Reserve banks and imputing net income for owner-occupied housing. While concept and coverage adjustments are considered more reliable overall than the misreporting and non-filing adjustment, reliability varies by specific adjustment. Adjustments based on administrative data are considered second in reliability only to Economic Census data, and adjustments based on survey data are considered third in reliability of this type of adjustment.

The company-establishment adjustment is limited to three of the income components of gross operating surplus derived from corporate income tax data: corporate profits before tax, corporate capital consumption allowance, and corporate net interest. The adjustment is made using a matrix of employment data from Census that relates employment by industry on a company basis and employment by industry on an establishment basis. By assuming that profits before tax, capital consumption allowance, and net interest are the same per employee for all

establishments performing the same activity, regardless of the company to which that establishment belongs, the matrix converts the industry distribution of each income component from a company basis to an establishment basis. An employment matrix based on Economic Census data is used to adjust the corporate income tax data available for the reconciliation. Due to the weakness of the assumptions that underlie this adjustment methodology, the company-establishment adjustment is considered less reliable than most concept and coverage adjustments but more reliable than the misreporting and non-filing adjustment.

Similar to the benchmark estimates, gross operating surplus estimates from the GDP-by-industry accounts are reviewed and finalized before they are submitted to the reconciliation model. In particular, gross operating surplus levels and the ratio of value-added by industry to total GDP is compared with corresponding measures in the benchmark use table. In addition, the initial GDI-based gross operating surplus estimates are reviewed for reasonability within the context of an annual time series.

4. Reconciling Gross Operating Surplus in a Balanced Input-Output Framework

4.1. Empirical Model

The objective is to reconcile in a balanced input-output framework the expenditure-based residual gross operating surplus estimates by industry in the 2002 benchmark use table with the gross operating surplus estimates derived from GDI-based external source data and BEA adjustment methodologies in the 2002 GDP-by-industry accounts. Based on an adaptation of Chen (2006), the objective is satisfied with a constrained optimization model that minimizes the reliability-weighted sum of squared differences between initial and final estimates subject to a system of accounting constraints.¹⁹ In particular, estimates of expenditure-based final use

¹⁹ Chen (2006) builds an empirical model based on the Stone method and incorporates statistical measures of relative reliability in the model to reconcile and balance estimates in the U.S. industry accounts. While Chen's

categories in the benchmark use table are considered highly reliable and are not allowed to adjust.²⁰ In addition, industry gross output, commodity gross output, and the compensation component of value-added, all in the benchmark use table and all derived from the 2002 Economic Census, are considered to be highly reliable and are not allowed to adjust.²¹ Likewise, initial estimates of the industry distribution and the total tax component of value-added are derived from the same data as the GDP-by-industry accounts and so are not allowed to adjust. Thus, the objective function in the model is constructed to adjust intermediate input estimates in the benchmark use table and value-added component estimates in both the GDP-by-industry accounts and the benchmark use table.²² In addition to identifying fixed elements of the benchmark use table, the system of constraints incorporates the accounting identities upon which the table is built. In this way, the model provides a tool for balancing the benchmark use table in addition to improving intermediate input estimates and gross operating surplus estimates in the table and the GDP-by-industry accounts.²³

More formally, let z and v denote intermediate inputs and value-added, respectively. Let w^z and w^v denote reliability measures for initial estimates of intermediate inputs and value-added,

(2006) model adjusts gross output, intermediate inputs, and all components of value-added, the model being implemented by the Industry Accounts Directorate only allows intermediate inputs and the gross operating surplus component of value-added to adjust.

²⁰ Fixler and Grimm (2005) show that expenditure-based GDP estimates are reliable over time based on the small magnitude of revisions.

²¹ In order to maintain consistency with other BEA programs, the compensation component of value-added in the GDP-by-industry accounts is derived using the same data as used for the industry distributions of the GDI components in the national income and product accounts. In particular, compensation in the GDP-by-industry accounts is derived using data from the Bureau of Labor Statistics. However, in order to maintain Census-based relationships in the benchmark use table, the industry distribution of the compensation component of value-added in the table is derived from Census compensation data and scaled to total compensation in the national income and product accounts. Thus, total compensation is the same but the industry distribution of compensation is different in the benchmark use table and GDP-by-industry accounts.

²² Although the objective function specifies all value-added components (i.e., compensation, taxes, and gross operating surplus) as variables, compensation and taxes are fixed as noted.

²³ In addition to the stated fixed elements of the benchmark use table, gross operating surplus estimates for farm and general government industries and the owner-occupied housing portion of gross operating surplus estimates in the real estate industry are fixed because they are reconciled using a different methodology.

respectively. Let x , y , Y^E , and Y^I denote gross output, final uses, aggregate expenditure-based GDP, and aggregate income-based GDI, respectively. Subscript i indexes industries, k indexes commodities, d indexes final use categories (including imports and exports), and h indexes the components of value-added (i.e., compensation, taxes, and gross operating surplus).²⁴ Superscript 0 denotes an initial value.

The model is described mathematically as follows:

$$\text{Min } S_{\{z, v\}} = \sum_i \sum_k \frac{(z_{i,k} - z_{i,k}^0)^2}{w_{i,k}^z} + \sum_i \sum_h \frac{(v_{i,h} - v_{i,h}^0)^2}{w_{i,h}^v} \quad (4.1.1)$$

Subject to:

$$\sum_d y_{k,d} - \sum_d y_{k,d}^0 = 0 \quad \forall k \quad (4.1.2)$$

$$x_i - x_i^0 = 0 \quad \forall i \quad (4.1.3)$$

$$x_k - x_k^0 = 0 \quad \forall k \quad (4.1.4)$$

$$v_{i,compensation} - v_{i,compensation}^0 = 0 \quad \forall i \quad (4.1.5)$$

$$v_{i,taxes} - v_{i,taxes}^0 = 0 \quad \forall i \quad (4.1.6)$$

$$x_i - \sum_k z_{i,k} - \sum_h v_{i,h} = 0 \quad \forall i \quad (4.1.7)$$

$$x_k - \sum_d y_{k,d} - \sum_i z_{k,i} = 0 \quad \forall k \quad (4.1.8)$$

²⁴ The Industry Accounts Directorate maintains multiple indices of industries and commodities. The most aggregated industry index comprises 65 industries, and the most aggregated commodity index comprises 69 commodities. The most disaggregated industry and commodity indices comprise nearly 1,000 industries and commodities, respectively. Each of the most aggregated industries and commodities comprise a mutually exclusive subset of the most disaggregated industries and commodities. The benchmark input-output database contains data for the most disaggregated industries and commodities. The GDP-by-industry database contains data for the most aggregated industries and no data for commodities. If the most disaggregated data are introduced to the model to reconcile industry gross operating surplus estimates from the benchmark input-output database and the GDP-by-industry database, the model requires separate indices and constraints to identify the most disaggregated industries included in the most aggregated industries. However, the choice of implementation strategy does not affect the specified model. Thus, for simplicity, this exposition focuses on a single index for industries and commodities.

$$Y^E - \sum_i \sum_h v_{i,h} = 0 \quad (4.1.9)$$

Initial conditions:

$$\sum_k \sum_d y_{k,d} - Y^E = 0 \quad (4.1.10)$$

$$\sum_i x_i^0 - \sum_i \sum_k z_{i,k}^0 \neq Y^E \quad (4.1.11)$$

$$x_i^0 - \sum_k z_{i,k}^0 \neq \sum_h v_{i,h}^0 \quad \forall i \quad (4.1.12)$$

$$x_k^0 - \sum_d y_{k,d}^0 \neq \sum_i z_{k,i}^0 \quad \forall k \quad (4.1.13)$$

$$\sum_i v_{i,gos}^0 - Y_{gos}^{I,0} = 0 \quad (4.1.14)$$

$$\sum_i x_i = \sum_k x_k \quad (4.1.15)$$

$$Y^{I,0} \neq Y^E \quad (4.1.16)$$

Constraints (4.1.2) to (4.1.6) identify fixed values for final use categories, industry gross output, commodity gross output, compensation, and taxes, all in the benchmark use table, respectively.²⁵ Equation (4.1.7) constrains the sum of final estimates for intermediate inputs used by an industry and value-added for the industry to equal gross output for the industry in the benchmark use table. Equation (4.1.8) constrains the sum of final estimates for a commodity used by industries as an intermediate input and sold for final uses to equal gross output for the

²⁵ Industry gross output and commodity gross output in the benchmark use table come from the balanced make table of the benchmark input-output accounts. Given the constraints in equations (4.1.3) and (4.1.4), the distribution of commodities produced by industries (i.e., the make table) is not affected by the model.

commodity in the benchmark use table. Constraint (4.1.9) indicates that final total value-added equals aggregate expenditure-based GDP.²⁶

The initial condition in equation (4.1.10) indicates that aggregate expenditure-based GDP is set exogenously by the total commodity distribution of final uses determined through the reconciliation with the national income and product accounts. Equation (4.1.11) indicates that total value-added is not initially equal to aggregate expenditure-based GDP. Equation (4.1.12) indicates that the sum of initial estimates for intermediate inputs used by an industry and value-added for the industry do not equal gross output for the industry in the benchmark use table. Equation (4.1.13) indicates that the sum of initial estimates for a commodity used by industries as an intermediate input and sold for final uses does not equal gross output for the commodity in the benchmark use table. The condition in equation (4.1.14) indicates that initial estimates of the gross operating surplus component of value-added come from the income-based GDI data. Equation (4.1.15) identifies initial equality between total industry gross output and total commodity gross output in the benchmark use table.²⁷ The initial condition in equation (4.1.16) identifies the statistical discrepancy between aggregate expenditure-based GDP and aggregate income-based GDI.²⁸

²⁶ Recall that the industry distribution and the total tax component of value-added is the same between the benchmark use table and the GDP-by-industry accounts. In contrast, the industry distribution of the compensation component of value-added is different but total compensation is the same between the benchmark use table and the GDP-by-industry accounts. Thus, while total final value-added equals aggregate expenditure-based GDP in both the benchmark use table and the GDP-by-industry accounts, the industry distribution of value-added will be different in both accounts.

²⁷ Recall that industry gross output and commodity gross output in the benchmark use table come from the balanced make table of the benchmark input-output accounts.

²⁸ At the most aggregated industry level and commodity level (i.e., 65 industries and 69 commodities), the model contains 4,550 variables to be solved and 135 constraints to be satisfied. At the most disaggregated industry level and commodity level (i.e., 987 industries and 987 commodities), the model contains 975,156 variables to be solved and 2,040 constraints to be satisfied. However, at the most disaggregated level, the intermediate input portion of the benchmark use table is not fully populated. Thus, the number of variables to be solved decreases accordingly. At the time this paper is being written, the Industry Accounts Directorate plans to introduce fully disaggregated data to the model. If this strategy proves impractical or computationally infeasible, the alternative strategy is to introduce aggregated data to the model in a first-stage reconciliation of gross operating surplus and impose the adjusted gross

4.2. Reliability Weights

The optimal estimator requires a variance-covariance matrix to weight the adjustments made to the initial estimates. In practice, the variance-covariance matrix for initial estimates comprising the benchmark use table and the GDP-by-industry accounts is unavailable. However, information regarding the relative reliabilities of initial estimates is used to construct a variance measure for each estimate in the benchmark data and GDP-by-industry data. In addition, given the lack of information regarding correlations among the initial estimates, covariance measures are assumed to be zero. While this assumption results in an estimator that is less than efficient, the inefficiency is less than may be introduced if the correlations are incorrectly determined (Byron, 1978; Nicolardi, 2000).

Following Chen (2006), the noted variance measures are constructed using coefficients of variation where available or internally-assigned reliability indicators where coefficients of variation are unavailable.²⁹ Recall that estimates in the benchmark use table and GDP-by-industry accounts are derived from two sources of information: 1) external source data and 2) BEA adjustment methodologies. External source data are subject to sampling and non-sampling errors. Coefficients of variation measure sampling errors and are available for source data provided to BEA by SOI and Census.³⁰ BEA adjustment methodologies are designed and implemented to correct non-sampling errors in external source data. However, the adjustments

operating surplus results from the first stage as constraints in a second-stage balancing with more disaggregated data (Chen, 2006).

²⁹ Coefficients of variation are defined as the standard error of an estimate divided by the estimate.

³⁰ In particular, coefficients of variation are available from SOI for approximately 47% of the total gross operating surplus value in the GDP-by-industry accounts. Likewise, coefficients of variation are available from Census for the expense data from the Business Expenses Survey and purchased services from the Annual Survey of Manufactures portion of the Census of Manufactures. Purchased services include the following expense categories: (1) accounting, auditing, and bookkeeping, (2) advertising, (3) communications, (4) computer services, (5) legal services, (6) management, consulting, and administrative services, (7) other expenses, (8) refuse removal, (9) repairs and maintenance of buildings and/or machinery, and (10) taxes and license fees. Coefficients of variation are available from Census for approximately 26% of the total intermediate input value in the benchmark use table.

are themselves subject to non-sampling errors. Thus, BEA analysts assign reliability indicators to individual adjustment records based on a rubric required to ensure a consistent application of the reliability indicators across industries and adjustments common to both the benchmark use table and the GDP-by-industry accounts.³¹

Table 1 contains a general summary of the rubric used by analysts to assign reliability indicators to individual records.³² As shown in Table 1, internally-assigned reliability indicators range from 1 to 5 in one-unit increments, with 1 indicating the most reliable estimates (e.g., Census-based estimates), 5 indicating the least reliable estimates (e.g., misreporting adjustments), and intermediate values indicating intermediately reliable estimates (e.g., concept adjustments). In this way, internally-assigned reliability indicators provide an indication of the percentage error or relative reliability of the related estimate, similar to coefficients of variation.

Coefficients of variation range from 0 to 1 with a theoretically infinite number of intermediate values.³³ In order to use internally-assigned reliability indicators where coefficients of variation are unavailable, a function is applied to internally-assigned reliability indicators to generate an equivalent scale of reliability across all indicators of reliability (i.e., internally-assigned reliability indicators and coefficients of variation). The function maps internally-assigned reliability indicators, denoted θ , according to an exponential parameter, denoted α , as follows:

³¹ In addition, using the noted rubric, BEA analysts assign reliability indicators to all detailed estimates in the benchmark data, including those derived from external source data, regardless of the availability of coefficients of variation.

³² In addition to reviewing estimates from the benchmark data and GDP-by-industry data before the data are submitted to the model, internally-assigned reliability indicators are reviewed for consistency of application across industries, type of adjustment, and data set (i.e., benchmark data and GDP-by-industry data) prior to implementing the model.

³³ Coefficients of variation greater than 1 are limited to 1. In the GDI-by-industry data, 1.8% of the coefficients of variation are greater than 1. In the benchmark input-output data, .05% of the coefficients of variation are greater than 1.

$$f(\theta) = \left(\frac{\theta - 1}{4} \right)^\alpha. \quad (4.2.1)$$

If $\alpha = 1$, internally-assigned reliability indicators are mapped linearly from 0 to 1 in increments of 0.25. If $\alpha > 1$, internally-assigned reliability indicators are mapped from 0 to 1, assuming that less-reliable estimates are exponentially less reliable than more reliable estimates. If $\alpha < 1$, internally-assigned reliability indicators are mapped from 0 to 1, assuming that reliability decreases at a diminishing rate.³⁴ In the end, a reliability indicator exists for an individual record and includes a coefficient of variation where available or an internally-assigned reliability indicator mapped by equation (4.2.1) where a coefficient of variation is unavailable.³⁵ The reliability indicators range from 0 to 1, with 0 indicating the most reliable estimates and 1 indicating the least reliable estimates.³⁶

Upon generating a reliability indicator for each record, a variance measure is constructed for each initial estimate by squaring the product of the initial estimate and the related reliability indicator, denoted q . The following equations summarize the construction of a variance measure, w^z or w^v , for each initial estimate of intermediate inputs, z^0 , and value-added, v^0 , in equation (4.1.1):

$$w^z = (q \times z^0)^2 \quad (4.2.2)$$

and

$$w^v = (q \times v^0)^2. \quad (4.2.3)$$

³⁴ At the time this paper is being written, evidence has not been generated regarding an appropriate value for α . However, given a lack of any evidence to support $\alpha \neq 1$, internally-assigned reliability indicators will be mapped assuming $\alpha = 1$.

³⁵ Given the use of fully disaggregated data in the model, this approach yields a variable with either a coefficient of variation or a mapped internally-assigned reliability indicator for each related estimate. Chen's (2006) approach requires aggregating coefficients of variation and linearly mapped internally-assigned reliability indicators in order to use data at the most aggregated industry level and commodity level (i.e., 65 industries and 69 commodities).

³⁶ This method assumes that the best estimates with internally-assigned reliability indicators are as good as the best estimates with coefficients of variation, and the worst estimates with internally-assigned reliability indicators are as bad as the worst estimates with coefficients of variation.

Thus, according to equation (4.1.1), adjustments in the model are weighted by the reciprocal of the variance measure.

5. Technology

As noted in the introduction, recent technology makes implementation of the reconciliation and balancing model possible. In particular, the reconciliation model is solved using the CPLEX solver in the Generalized Algebraic Modeling System (GAMS). GAMS is a flexible optimization software package designed to handle large, complex mathematical programming problems. CPLEX is a GAMS solver with solution algorithms for linear, quadratically constrained, and mixed-integer problems. In particular, CPLEX includes simplex and barrier algorithms. The CPLEX solver automatically chooses the optimal combination of algorithms to efficiently solve the particular model specified. Alternatively, GAMS also allows users to adjust tuning parameters in order to set algorithmic options. The reconciliation model is solved with the combination of algorithms chosen by GAMS.

6. Next Steps

As noted in sections 2 and 3, the data used to construct the benchmark use table and GDP-by-industry accounts are subject to review and finalization before they are submitted to the reconciliation model. At the time this draft is being written, the benchmark data are not ready for review and finalization. Thus, a set of reconciled gross operating surplus estimates is currently unavailable. However, preliminary work during development of the model shows that the model is computationally feasible and efficient with a set of fully disaggregated data.

Upon review and finalization of the benchmark data, the model is implemented and the resulting estimates for intermediate inputs and gross operating surplus are subject to a follow-up review to ensure the estimates are consistent with the economic concepts upon which the

benchmark use table and GDP-by-industry accounts are built. In particular, the industry distribution of reconciled gross operating surplus is compared to the industry distribution of initial residual gross operating surplus estimates in the benchmark data and the industry distribution of initial gross operating surplus estimates in the GDP-by-industry accounts. Likewise, input-output coefficients and the ratio of final value-added by industry to total GDP are examined for unreasonable changes from the 1997 benchmark use table. In addition, initial estimates for intermediate inputs, trade margins, and transportation costs in the benchmark data are compared to final estimates and final estimates are reviewed for negative values. Furthermore, the net operating surplus components of reconciled gross operating surplus estimates are divided by produced assets from BEA's fixed asset accounts to review the dispersion of rates of return across industries. Finally, reconciled gross operating surplus estimates are examined for reasonability within the context of an annual time series.

References

- Brown, Robert E. and Mark J. Mazur. 2003. "IRS's Comprehensive Approach to Compliance Measurement" paper presented at the National Tax Association Spring Symposium.
- Byron, Ray P. 1978. "The Estimation of Large Social Account Matrices" *Journal of the Royal Statistical Society, Series A*, 141(3), 359-367.
- Chen, Baoline. 2006. "A Balanced System of Industry Accounts for the U.S. and Structural Distribution of Statistical Discrepancy" BEA working paper WP2006-8.
- Fixler, Dennis and Bruce Grimm. 2005. "Reliability of the NIPA Estimates of U.S. Economic Activity" *Survey of Current Business*, 85(2), 8-19.
- Lawson, Ann M.; Kurt S. Bersani; Mahnaz Fahim-Nader; and Jiemin Guo. 2002. "Benchmark Input-Output Accounts of the United States, 1997" *Survey of Current Business*, 82(12), 19-43.
- Lawson, Ann M.; Brian C. Moyer; Sumiye Okubo; and Mark A. Planting. 2006. "Integrating Industry and National Economic Accounts: First Steps and Future Improvements" in *A New Architecture for the U.S. National Accounts*; edited by Dale W. Jorgenson, J. Steven Landefeld, and William D. Nordhaus; The University of Chicago Press, Chicago.
- Mantegazza, Susanna and Stefano Pisani. 2000. "Present Practices and Future Developments" paper presented at the 13th International Conference on Input-Output Techniques, University of Macerata, Italy.
- Moyer, Brian C.; Mark A. Planting; Mahnaz Fahim-Nader; and Sherlene K. S. Lum. 2004a. "Preview of the Comprehensive Revision of the Annual Industry Accounts: Integrating the Annual Input- Output Accounts and Gross-Domestic-Product-by-Industry Accounts" *Survey of Current Business*, 84(3), 38-51.
- Moyer, Brian C.; Mark A. Planting; Paul V. Kern; and Abigail M. Kish. 2004b. "Improved Annual Industry Accounts for 1998-2003: Integrated Annual Input-Output Accounts and Gross- Domestic-Product-by-Industry Accounts" *Survey of Current Business*, 84(6), 21-57.
- Nicolardi, Vittorio. 2000. "Balancing Large Accounting Systems: An Application to the 1992 Italian I-O Table" paper presented at the 13th International Conference on Input-Output Techniques, University of Macerata, Italy.
- Ploeg, Frederick van der. 1982. "Reliability and the Adjustment of Sequences of Large Economic Accounting Matrices" *Journal of the Royal Statistical Society, Series A*, 145(2), 169-194.

- Ploeg, Frederick van der. 1984. "Generalized Least Squares Methods for Balancing Large Systems and Tables of National Accounts" *Review of Public Data Use*, 12(1), 17-33.
- Stone, Richard. 1961. *Input-Output and National Accounts*; Organization for European Economic Cooperation, Paris.
- Stone, Richard. 1968. "Input-Output Projections: Consistent Prices and Quantity Structures" *L'Industria*, 2, 212-224.
- Stone, Richard. 1970. *Mathematical Models of the Economy and Other Essays*; Chapman and Hall, London.
- Stone, Richard. 1975. "Direct and Indirect Constraints in the Adjustment of Observations" in *Nasjonalregnskap, Modeller og Analyse: Essays in Honour of Off Aukrust*; Statistisk Sentralbyrå, Oslo.
- Stone, Richard. 1976. "The Development of Economic Data Systems" in *Social Accounting for Development Planning with Special Reference to Sri Lanka*; edited by G.Pyatt et al.; Cambridge University Press, Cambridge.
- Stone, Richard. 1984. "Balancing the National Accounts: The Adjustment of Initial Estimates – A Neglected Stage in Measurement" in *Demand, Equilibrium and Trade: Essays in Honor of Ivor F. Pearce*; edited by A. Ingham and A.M. Ulph; St. Martin's Press, New York, 191-212.
- Stone, Richard; James E. Meade; and David G. Champernowne. 1942. "The Precision of National Income Estimates" *The Review of Economic Studies*, 9(2), 111-125.
- Tuke, Amanda and Vanna Aldin. 2004. "Reviewing the Methods and Approaches of the UK National Accounts" *Economic Trends*, 47-57.
- Yuskavage, Robert E. 2000. "Priorities for Industry Accounts at BEA" paper presented at the November 17 meeting of the BEA Advisory Committee.

Chart 1: Overview of the Structure of the Benchmark Use Table

		Industries						Final Uses							Total Commodity Output	
		Construction	Manufacturing	Trade	Information	Government	Other Services	Total Intermediate Inputs	Personal Consumption Expenditures	Private Equipment and Software	Changes in Private Inventories	Exports	Imports	Government		GDP
Commodities	Construction															
	Manufacturing															
	Trade															
	Information															
	Government															
	Other Services															
	Total Intermediate Inputs															
Value-Added	Compensation															
	Taxes															
	Gross Operating Surplus															
	Total Value-Added															
Total Industry Output																

Table 1: General Rubric for Assigning Reliability Indicators to Underlying Data

<i>Reliability Indicator</i>	<i>Source of Underlying Data</i>
1	<ul style="list-style-type: none">• Economic Census data with no adjustments
2	<ul style="list-style-type: none">• Economic Census data with adjustments• Survey data with no adjustments• Concept and coverage adjustments based solely on administrative data• Non-employer adjustments
3	<ul style="list-style-type: none">• Survey data with adjustments• Trade association data• Concept and coverage adjustments based on survey data• Auxiliary service adjustments
4	<ul style="list-style-type: none">• Company-establishment conversion adjustment• Adjustments based on a combination of analyst judgment and external source data
5	<ul style="list-style-type: none">• Misreporting adjustments• Adjustments based solely on analyst judgment