Supplemental Information for the Presentation "A Balanced System of Industry Accounts for the U.S. and Structural Estimation of Statistical Discrepancy"

Baoline Chen

Researcher, NEA Directorate

1. The GLS account reconciliation model and computation of solutions (Section 3 of the presentation)

2. Figure 1: Percentage adjustments in gross output, intermediate inputs, income-by-industry, and adjustments from company-to-establishment conversion (Section 4 of the presentation)

3. Tables of results (with a few more industries) (Section 5 of the presentation)

4. Partial references (Section 1 of the presentation)

5. Acknowledgement

3. A GLS Method of Accounts Reconciliation

The objective here is to reconcile the Input-Output and Income-by-industry data with expenditure based benchmark GDP. Because the expenditure based GDP estimate was from the 2003 comprehensive revision, it was considered the correct measure of GDP. Therefore, the initial estimates of final demand by expenditure category, exports, and imports were not to be adjusted. The mathematical problem is to minimize the reliability weighted sum of squares of adjustments of initial estimates in all components of gross output, intermediate inputs, and value-added of all industries and commodities, subject to accounting constraints and restrictions.

Let x, z, v denote estimates of gross output, intermediate input, and value-added. Let wx, wz and wv denote reliabilities of the corresponding initial estimates, where reliabilities are measured by variances of initial estimates. Let y, e and m denote final demand by expenditure category, exports and imports. Let Y^E and Y^I denote aggregate GDP and GDI. Formally, the reconciliation problem is

$$(1)\quad \text{Min } S_{\{x,z,v)} = \sum_{i=1}^{65}\sum_{k=1}^{69}\frac{(x_{ik}-x_{ik}^0)^2}{wx_{ik}} + \sum_{i=1}^{65}\sum_{k=1}^{69}\frac{(z_{ik}-z_{ik}^0)^2}{wz_{ik}} + \sum_{i=1}^{65}\sum_{f=1}^{3}\frac{(v_{if}-v_{if}^0)}{wv_{if}},$$

$$(2)\qquad \text{s.t.}\quad \sum_{k=1}^{69}x_{ik} - \sum_{k=1}^{69}z_{ik} - \sum_{f=1}^{3}v_{if} = 0, \qquad i = 1, \ldots, 65,$$

(3) $$\sum_{i=1}^{65} x_{ki} - \sum_{i=1}^{65} z_{ki} - \sum_{d=1}^{11} y_{kd}^{o} - e_{k}^{o} + m_{k}^{o} = 0, \qquad k = 1, \ldots, 69,$$

(4) $$\sum_{i=1}^{65}\sum_{f=1}^{3} v_{if} = \sum_{k=1}^{69} \sum_{d=1}^{11} y_{kd}^{o} - e_{k}^{o} + m_{k}^{o},$$

with the initial conditions satisfying

(5) $$\sum_{k=1}^{69} [\sum_{d=1}^{11} y_{kd}^{o} - e_{k}^{o} + m_{k}^{o}] = Y^{E0},$$

(6) $$\sum_{i=1}^{65}\sum_{f=1}^{3} v_{if}^{o} = Y^{I0}.$$

Letters i, k, f and d are indexes for industry, commodity, component of value-added and category of final demand, and superscript “0” indicates initial estimates.

The accounting constraint (2) states that for each industry, final estimates of intermediate inputs and value-added must sum up to final estimate of industry gross output. Constraint (3) says that for each commodity, final estimates of commodities used as intermediate inputs and commodities sold as final demand must sum up to final estimate of commodity output. Constraint (4) states value-added estimates from 65 industries must

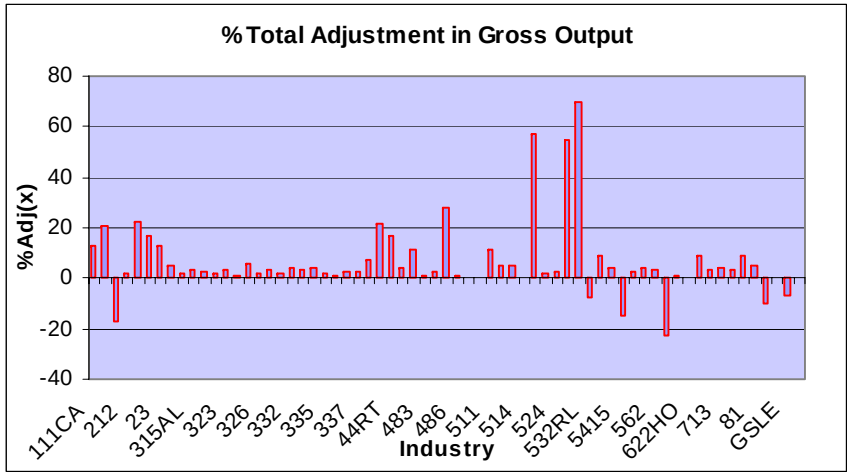
sum up to total GDP, removing aggregate statistical discrepancy. Equation (5) and (6) are initial conditions indicating that initial estimate of total GDP differs from initial estimate of total GDI. The difference between the two, Y^{E0} - Y^{I0}, is the aggregate statistical discrepancy.

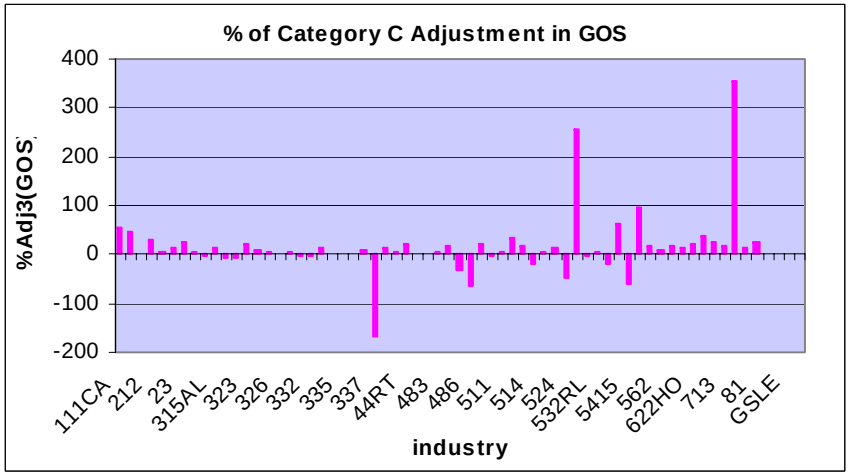
The system of accounts described here consists of 100062 variables and 135 accounting constraints. Because the 11 final uses, exports and imports are held as fixed, there are 9165 final estimates to be solved for. The reconciliation model is solved with an optimization software package GAMS using its CPLEX solver, a powerful tool for handling large linear or quadratic programming models. One reason that reconciliation of disaggregated system of accounts has not received wide attention since its inception by Byron (1978) was because of the computer technology available at the time and the large memory requirement of reconciliation. Under the current technology of computer and software, the system of accounts described can be successfully reconciled in about .5 second.

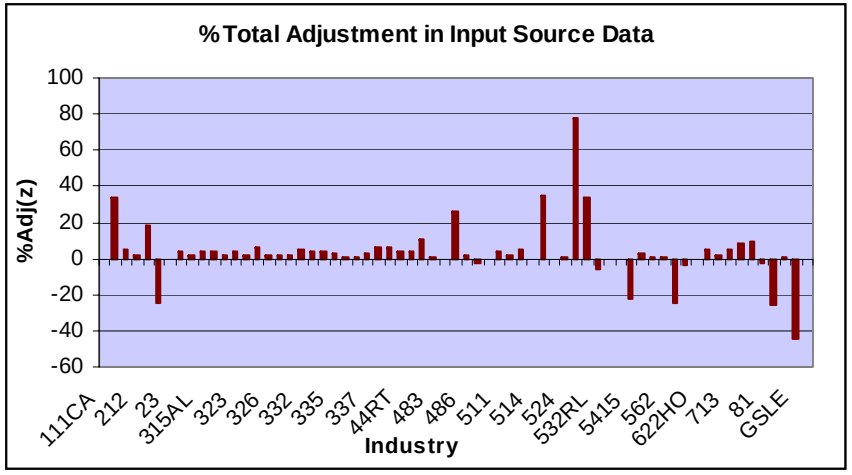
4. Measure Reliability of Initial Data

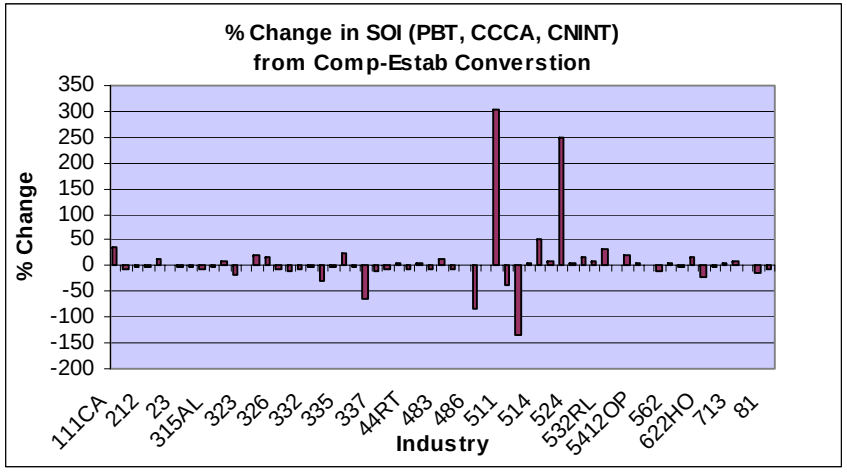
To correct errors such as double-counting, misallocation, misreporting, misspecification and omission, various types of adjustments are made to the source data. Figure 1 provides a glimpse of adjustments relative to initial estimates.

Figure 1: Percentage Adjustment in Gross Output, Total Inputs and Some Items of Value Added Source Data from Income-by-Industry









5. Balanced Estimates and Distribution of Statistical Discrepancy

Table 1: Initial and Balanced Estimates for 65 Industries
(Millions of dollars)

	Initial Estimates					Balanced Estimates w = var(est^0)					Balanced Estimates w = abs(est^0)			
Pub code	x_i^0	z_i^0	v_i^0(GDI)	$[x_i^0-(z_i^0+v_i^0)]$%		x_i*	z_i*	v_i*	xi*-zi*-vi*		x_i'	z_i'	v_i'	x_i'-z_i'-v_i'
111CA	241952	153810	88142	0.00		241952	153810	88142	0		244496	154618	89878	0
322	149062	97640	50943	0.32		149175	96499	52676	0		149786	98193	51593	0
335	109172	67362	78029	-33.18		109341	65819	43522	0		118336	55055	63281	0
44RT	830070	313587	585081	-8.26		835062	301269	533793	0		844738	290864	553874	0
514	47220	16953	18587	24.74		47219	19356	27863	0		42507	19636	22870	0
531	1260014	318624	883180	4.62		1256817	339677	917141	0		1248366	328743	919623	0
532RL	176438	31943	73375	40.31		165254	32011	133242	0		141976	42670	99306	0
...	...	...	...	...		...	...	...	...		...	...	...	...
Sum	15217582	6917468	8257803	-3.44		15201130	6896787	8304344	0		15184319	6879975	8304344	0

Note: x_i, z_i are from 1997 benchmark I-O database, and v_i are from Income-by-industry data prior to allocation of SD. Data were not published.

Table 2: Initial and Balanced Estimates for 69 Commodities
(Millions of dollars)

	Initial Estimates				Balanced Estimates w = var(est^0)					Balanced Estimates w = abs(est^0)			
Com Code	$z_k^0 + y_k^0$	x_k^0	$[x_k-(z_k+y_k)]\%$		x_k^*	z_k^*	y_k^0	$x_k^*-z_k^*-y_k^0$		x_k'	z_k'	y_k^0	$x_k'-z_k'-y_k^0$
22	335178	335214	0.01		338234	185978	152256	0		340426	188170	152256	0
324	174267	173626	-0.37		173745	117568	56177	0		185043	128865	56177	0
3364OT	150707	148435	-1.53		149094	63909	85185	0		148853	63668	85185	0
42	736666	736429	-0.03		744859	364826	380034	0		743952	363919	380034	0
481	113654	124418	8.65		114560	54558	60002	0		120992	60990	60002	0
487OS	89435	83861	-6.65		83875	67689	16186	0		88136	71950	16186	0
531	1227035	1227089	0.00		1225686	383026	842660	0		1219163	376503	842660	0
GFG	459378	459378	0.00		459378	0	459378	0		459378	0	459378	0
...	...	...	...		...	...	...	...		...	...	...	...
Total	15221812	15217582	-7.197		15201130	6896787	8304344	0		15184319	6879975	8304344	0

Note: x_k and z_k are from benchmark I-O database and are not published, and y_k are from 2003 benchmark revision.

Table 5: Estimates of SD by Industry Based on Relative Reliability and Relative Size of Initial Estimates

	Initial Gap		Estimated SD by Industry (w = var(est^0))					Estimated SD by industry (w = abs(est^0))			
Pub code	[(x_i^0-z_i^0) – v_i^0]		SDi'	SDi*/SD	var(v)/ var(x-z)	vi*/GDP		SDi'	SDi'/SD	vi/(xi-zi)	v i'/GDP
211	-16910		-14695	-31.57	4.05	0.53		-8847	-19.01	1.41	0.6
311FT	24823		9749	20.95	0.32	1.67		8240	17.7	0.84	1.65
324	-43164		-42410	-91.13	82.03	0.29		-13775	-29.6	2.83	0.64
334	32578		17389	37.36	0.44	1.94		11607	24.94	0.81	1.87
335	-36219		-34507	-74.14	1304.39	0.52		-14747	-31.69	1.87	0.76
481	-9494		-3409	-7.33	0.04	0.62		-5271	-11.32	1.21	0.59
531	58210		33961	72.97	0.82	11.04		36443	78.3	0.94	11.07
532RL	71120		59868	128.63	0.09	1.6		25932	55.72	0.51	1.2
721	5312		5310	11.41	418.42	0.91		3537	7.6	0.93	0.89
...	...		...	...	...	...		...	...	...	...
Sum	42311		46541	1		1		46541	1		1

Reference

1. Beaulieu, J.J. and Bartelsman, E.J. (2004), "Integrating Expenditure and Income Data: What to do with the Statistical Discrepancy?" Unpublished paper, Board of Government of the Federal Reserve System and Free University, Amsterdam.

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

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

4. Ploeg, F. van der (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), p. 169-194.

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

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

7. Weale, M. (1992), "Estimation of Data Measured with Error and Subject to Linear Restrictions," Journal of Applied Econometrics, Vol. 7, No. 2, p. 167-174.

8. Weale, M. (1988), "The Reconciliation of Values, Volumes and Prices in the National Accounts," Journal of the Royal Statistics Society, Series A, Vol. 151 (1), p. 211-221.

Acknowledgement:

This project has received strong support from NEA and Industry Accounts management since the proposal was submitted. During the time I have worked on the project, I have received assistance and support from staff members from different divisions. Some have provided data needed for the project, some with whom I have had helpful discussions about the project, and some have helped process data or with my SAS programming questions. I would like to express my appreciation to the following people,

Nation Economic Account: Brent Moulton, Brian Moyer, Brooks Robinson, Steve Andrews, Greg Keys, Kurt Kunze, Nicole Mayerhauser, Julia Nguyen, Marshall Reinsdorf, Charles Robinson, Dave Wasshausen.

Industry Accounts: Sumiye Okubo, Ann Lawson, Karen Horowitz, George Smith, Felicia Candela, Tom Howells, Sarah Mattingly, Doug Meade, Gabe Medeiros, Mark Planting, Dylan G. Rassier, Carol Robbins, Conrad E. Roesch, Erich Strassner, Mary Streitwieser, Bob Yuskavage.

Regional Division: Bettina Aten, Zoe Ambargis.

Office of Chief Statistician: Ruth A. Bramblett.

Census Bureau: Stacy Cole.

IRS: Kimberly A. Henry, Thomas Petska, Valerie Testa, Ahmed Z. Yahia.

My special thanks go to Karen Horowitz and George Smith. George has helped obtain and process data on sampling errors from the IRS. He also provided Income-by-industry data and helped analyze adjustment data on the income side of the accounts. Karen Horowitz has taught me everything I understand about I-O account, explained the data problems in the 1997 data set, and always provided timely answers to my questions. She provided large I-O data sets from the databank and helped analyze adjustment data on the I-O side of the accounts. I would also like to express my sincere thanks to my former colleague Zhi Wang for bringing this subject to me and for his help with computer programs in the early stage of the project.