Estimation Errors in Input-Output Tables and Prediction Errors in Computable General Equilibrium Analysis

Nobuhiro Hosoe

October 2013

Estimation Errors in Input-Output Tables and Prediction Errors in Computable General Equilibrium Analysis

October 3, 2013

Nobuhiro Hosoe[*]

Abstract

We used 1995-2000-2005 linked input-output (IO) tables for Japan to examine estimation errors of updated IO tables and the resulting prediction errors in computable general equilibrium (CGE) analysis developed with updated IO tables. As we usually have no true IO tables for the target year and therefore need to estimate them, we cannot evaluate estimation errors of updated IO tables without comparing the updated ones with true ones. However, using the linked IO tables covering three different years enables us to make this comparison. Our experiments showed that IO tables estimated with more detailed and recent data contained smaller estimation errors and led to smaller quantitative prediction errors in CGE analysis. Despite the quantitative prediction errors, prediction was found to be qualitatively correct. As for the performance of updating techniques of IO tables, a cross-entropy method often outperformed a least-squares method in IO estimation with only aggregate data for the target year but did not necessarily outperform the least-squares method in CGE prediction.

Keywords

Input-output tables; computable general equilibrium analysis; non-survey method; cross-entropy method

JEL Classification

D57; C68; C83

1.   Introduction

Input-output (IO) tables are one type of data essential for constructing the social accounting matrices (SAM) used in computable general equilibrium (CGE) modeling. They also give CGE models attractive features as a multi-sectoral model describing details of industrial activities useful for empirical policy analysis, such as trade, environment, and tax policies. However, the availability of IO tables is often limited, because IO tables with such details are costly to construct. IO tables for Japan are constructed regularly every five years after intensive works with the target year data for several years. Many countries with poorer statistical organizational capacity cannot afford to construct IO tables on a regular basis.

Such low availability of IO tables forces CGE modelers to use IO tables that are several years old. When data and models are too old to use for their analysis, CGE modelers have to update IO tables themselves with simpler methods and fewer data than those employed by professional statisticians. CGE modelers employ a so-called non-survey method to update new IO tables by replacing a part of old IO tables with the target-year data, which are often incomplete and sometimes inconsistent with each other. The updated IO tables inevitably suffer some estimation errors compared with true tables.

There are two main problems that CGE modelers face. One is that the updated IO tables may suffer estimation errors. (In connection with this issue, they might also be interested in finding methods of updating that can reduce estimation errors.) The other problem is prediction errors in CGE analysis caused by the estimation errors in IO tables.[1] Usually, we cannot examine the estimation errors in IO tables because we have no true IO tables for a target year and, thus, have to estimate them permitting some estimation errors. Without true IO tables or true CGE models, we cannot measure prediction errors of CGE analysis, either.

---

[1] Although CGE models are not used only for prediction, we refer to errors of simulation results in CGE analysis as prediction errors to avoid confusion with estimation errors in updated IO tables.

In the literature, Robinson et al. (2001) estimated stylized SAMs for Mozambique with two different matrix balancing methods: RAS and cross-entropy (CE) methods. They found that these estimated SAMs were similar in flow data but that the CE method was likely to estimate a SAM closer to prior values in input coefficients. Cardenete and Sancho (2004) did experiments estimating a regional SAM for Andalusia, Spain and found results similar to the ones by Robinson et al. (2001). Then, they simulated tax reforms with CGE models calibrated to their updated SAMs to compare their simulation results with each other. However, these two studies compared a table estimated with one method to a table estimated with another method, or a CGE simulation result based on an estimated SAM to another, not to a true SAM or a CGE simulation result based on a true IO table/SAM. They could not conclude anything about the accuracy of the estimated SAM or the performance of the matrix balancing methods.

As true IO tables were not available, Bonfiglio and Chelli (2008) randomly created "true" IO tables by a Monte Carlo method for their numerical experiments to examine the performance of various estimation methods. Real true tables have been used very rarely. Jalili (2000) did experiments by updating an IO table for the Former Soviet Union from 1966 to 1972 with various methods such as RAS and least-squares (LS) methods and compared them to the true table for 1972. Jackson and Murray (2004) did similar experiments by updating the US tables from 1966 to 1972 and from 1972 to 1977 with 10 different matrix balancing method and found no better methods than the RAS method overall. These studies focused on estimation errors of updated IO tables, but did not examine prediction errors in CGE analysis calibrated to these updated tables.

In this study, we used linked IO tables for 1995, 2000, and 2005 from the Ministry of Internal Affairs and Communications (2011) and measured (1) estimation errors of updated IO tables from 1995 or 2000 to 2005 by comparing them with the true IO table for 2005 and (2) prediction errors in CGE analysis caused by the estimation errors in the updated IO tables (Figure 1.1). We considered a LS method and a CE method among many

matrix balancing methods and two cases of rich and poor data availability for the target year in updating IO tables. Finally, we developed CGE models calibrated to these estimated and the true IO tables and made two numerical policy experiments to measure their prediction errors attributable to richness and age of information as well as matrix balancing methods. We found that the effect of richness and age of information used in updating IO tables was clear and straightforward but that the effect of matrix balancing methods was not.

Figure 1.1. Outline of the study
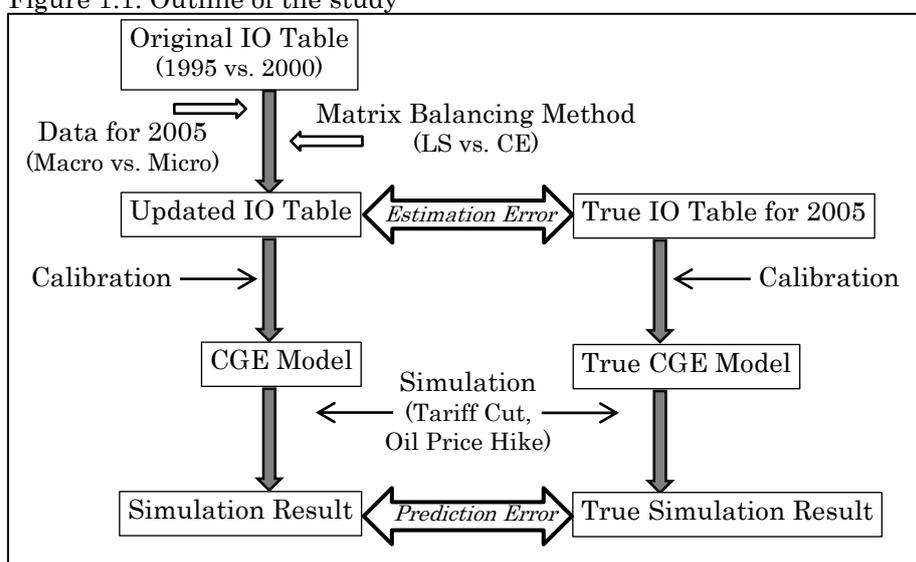
Our paper proceeds as follows. Section 2 discusses estimation methods and estimation errors of IO tables. Section 3 shows simulation results of CGE analysis to measure prediction errors. Section 4 concludes the paper, followed by an appendix demonstrating the robustness of CGE simulation results with respect to key trade elasticity.

2.   Estimation of IO Tables

2.1   Availability of Target Year Data and Settings of Prior Values

Let us update an old IO table $\left(IO_{u,v}^0\right)$ for 1995 or 2000 to a new one $\left(IO_{u,v}\right)$ for a target year of 2005 by partly replacing old data with the target year data. The IO table

$\left(IO_{u,v} \right)$ can be subdivided into a few sub-matrixes,

$$\left(IO_{u,v} \right) = \begin{pmatrix} \left(X_{i,j} \right) & \left(F_{i,f} \right) \\ \left(Y_{y,j} \right) & \mathbf{0} \end{pmatrix},$$

where

$\left(X_{i,j} \right)$: intermediate input from industrial sectors i to j,

$\left(Y_{y,j} \right)$: value added of the y-th factor used by the j-th sector, and

$\left(F_{i,f} \right)$: final demand by the f-th user purchased from the i-th industrial sector.

The updated tables must satisfy the row-sum and column-sum consistency for each industry i:

$$\sum_{j} X_{j,i} + \sum_{y} Y_{y,i} = \sum_{u} IO_{u,i} = \sum_{u} IO_{i,u} = \sum_{j} X_{i,j} + \sum_{f} F_{i,f} \qquad \forall i \qquad (2.1)$$

If additional information is available for some cells in the new IO table, those values are fed into the estimation process by pinning down these cell values with constraints. In this study, the available information was retrieved from the true IO table for 2005 $\left(IO_{u,v}^{2005} \right)$ for ease of comparison.[2] If the final demand of the i-th good by the f-th user $F_{i,f}^{2005}$ is available for each cell, we can impose a constraint as follows.

$$F_{i,f} = F_{i,f}^{2005} \qquad \forall i, f \qquad (2.2)$$

Similarly, if we know the y-th value-added component of the j-th sector $Y_{y,j}^{2005}$, we can impose a constraint as follows.

$$Y_{y,j} = Y_{y,j}^{2005} \qquad \forall y, j \qquad (2.3)$$

---

[2] In reality, there are various factors that can cause deviations of data in the compiled true IO tables from true data, such as measurement errors in the original data and matrix balancing done for row-column consistency. However, we simplified our discussion by assuming the data in the IO table for 2005 to be true.

In contrast, while these microeconomic data may be less likely to be available, we can more often obtain macroeconomic data. That is, if we know the final demand of the f -th user in total (e.g., total household expenditure) $\sum_i F_{i,f}^{2005}$, we can impose a constraint, which is looser than (2.2), as follows.

$$\sum_i F_{i,f} = \sum_i F_{i,f}^{2005} \qquad \forall f \qquad (2.4)$$

We can consider a similar constraint for the total of the y -th value added component $\sum_i Y_{y,i}^{2005}$, which is looser than (2.3), as follows.

$$\sum_i Y_{y,i} = \sum_i Y_{y,i}^{2005} \qquad \forall y \qquad (2.5)$$

We may well conjecture that the signs of cell values in the old tables are still kept in the target year and impose a sign condition:

$$sign(IO_{u,v}) = sign(IO_{u,v}^0) \quad \forall u, v \qquad (2.6)$$

We can also conjecture the level of cell values (prior values). For example, if we assume that input patterns are stable over time, we can compute an input coefficient for industries or expenditure share for final demand $a_{u,v}$ as follows.

$$a_{u,v} = \frac{IO_{u,v}^0}{\sum_u IO_{u,v}^0} \qquad \forall u, v \qquad (2.7)$$

By combining this coefficient/share $a_{u,v}$ with the IO table margin data $\sum_u IO_{u,v}^{2005}$, we can update the prior values as follows.

$$IO_{u,v}^0 = a_{u,v} \sum_u IO_{u,v}^{2005} \qquad \forall u, v \qquad (2.8)$$

We can estimate another type of prior value. When we know all the cell values in the value-added matrix $(Y_{y,j})$ and the final demand matrix $(F_{i,f})$ in addition to the column totals for the j -th industrial sector $\sum_u IO_{u,j}^{2005}$ as assumed for (2.2) and (2.3), we

need to estimate only the intermediate input matrix $\left(IO_{i,j}\right)$. By estimating an input coefficient $a_{i,j}$ computed from the old table data $\left(IO^0_{i,j}\right)$ as

$$a_{i,j} = \frac{IO^0_{i,j}}{\sum_i IO^0_{i,j}} \qquad \forall i, j \tag{2.9}$$

and combining $a_{i,j}$ with the known column total $\sum_u IO^{2005}_{u,j}$, we can compute prior values for this sub-matrix $\left(IO_{i,j}\right)$ as follows.

$$IO^0_{i,j} = a_{i,j} \sum_i IO^{2005}_{i,j} \qquad \forall i, j \tag{2.10}$$

2.2   Matrix Balancing Methods

Because prior values are updated partially and hence do not satisfy the constraints for the row-sum and column-sum balance (2.1) and the additional constraints (2.2)–(2.5), we need to conduct matrix balancing to recover their balance and consistency with the control totals. Let us formulate this adjustment as a constrained minimization problem. If we use a LS method, its minimand is

$$W = \sum_{u,v} \left(\frac{IO_{u,v}}{IO^0_{u,v}} - 1 \right)^2 \tag{2.11}$$

If we, instead, use a CE method, it is defined as

$$W = \sum_{u,v} IO_{u,v} \ln \left(\frac{IO_{u,v}}{IO^0_{u,v}} \right) \tag{2.12}$$

(Golan et al. (1996)).[3]

In our computational implementation, we excluded terms with priors $IO^0_{u,v}$ equal to zero in the minimands. In addition, when we used a CE method, we modified the minimand (2.12) by taking absolute values for negative prior values and using an arbitrary

---

[3] Miller and Blair (2009, Ch.7) suggested alternative minimands for this problem.

small value δ to treat cases where the estimated cell values $IO_{u,v}$ became zero, as suggested by Robinson and El-Said (2000):[4]

$$W = \sum_{u,v} |IO_{u,v}| \ln\left(\frac{IO_{u,v} + \delta}{IO_{u,v}^0 + \delta}\right) \qquad (2.12')$$

We updated tables from 1995 or 2000 to 2005 by minimizing deviations of new cell values from the priors (2.8) or (2.10) as defined by (2.11) or (2.12') subject to the row-sum and column-sum constraint (2.1) and the sign condition (2.5) as well as constraints describing the additional microeconomic and macroeconomic data (2.2)–(2.5). In our experiments, we considered eight cases with variations in richness of additional data, age of original IO tables, and minimands to examine their contributions to accuracy of estimates (Table 2.1).

Table 2.1: Estimation Methods

Case	Additional Data and Constraints[a]	Prior Values	Original IO Table	Minimand[b]
LS_Macro(1995)	Macroeconomic data (2.4), (2.5)	(2.8)	1995	LS (2.11)
LS_Micro(1995)	Microeconomic data (2.2), (2.3)	(2.10)	1995	LS (2.11)
CE_Macro(1995)	Macroeconomic data (2.4), (2.5)	(2.8)	1995	CE (2.12')
CE_Micro(1995)	Microeconomic data (2.2), (2.3)	(2.10)	1995	CE (2.12')
LS_Macro(2000)	Macroeconomic data (2.4), (2.5)	(2.8)	2000	LS (2.11)
LS_Micro(2000)	Microeconomic data (2.2), (2.3)	(2.10)	2000	LS (2.11)
CE_Macro(2000)	Macroeconomic data (2.4), (2.5)	(2.8)	2000	CE (2.12')
CE_Micro(2000)	Microeconomic data (2.2), (2.3)	(2.10)	2000	CE (2.12')

[a] In all the cases, we assumed that the column totals for all the industrial sectors j $\sum_u IO_{u,j}^{2005}$ were known and imposed constraints for the row-sum and column-sum consistency (2.1) and signs (2.6).
[b] LS: least-squares, CE: cross-entropy.

We aggregated the 108-sector linked IO tables to 18-sector tables for simplicity of our subsequent CGE experiments in Section 3 (Table 2.2). We fed the detailed or aggregate data appearing in the true 18-sector table for 2005 about capital and labor income, indirect taxes, household and government consumption, investment, exports, imports, and import

---

[4] In our study, we assumed δ = 10E-10.

tariffs, which were typically reported in such frequently used official statistics in Japan as the national account tables, the census of manufactures, and the trade statistics. We assumed that sectoral total input $\sum_u IO_{u,j}^{2005}$ was known in all cases.

Table 2.2: Sectoral Aggregation

Sectors in Aggregated IO Tables	Abbreviations	Sector Codes in the Original 108-Sector Tables
Industries (i, j)		
Agriculture	AGR	001–005, 011
Other Mining	MIN	006–007
Oil, Coal, and Natural Gas	OIL	008
Food	FOD	009–010, 012
Textiles and Apparel	TXA	013–014, 032
Wood, Paper, and Printing	WPP	015–019
Chemical	CHM	020–027, 030–031
Oil and Coal Products	P_C	028–029
Pottery	POT	033–036
Steel	STL	037–040
Metal	MET	041–044
Machinery and Other Manufacturing	MAN	045–048, 062–063
Electric Equipment	EEQ	049–056
Transportation Equipment	TEQ	057–061
Electricity	ELY	069
Town Gas	TWG	070
Transportation	TRS	078–085
Other Services	SRV	065–068, 071–076, 086–108
Value Added (y)		
Capital	CAP	113–115(row code)
Labor	LAB	111–112(row code)
Indirect Tax	IDT	116–117(row code)
Final Demand (f)		
Household	HOH	111–112(column code)
Government	GOV	113–114(column code)
Investment	INV	115–117(column code)
Exports	EXP	120–121
Imports	IMP	125
Import Tariffs	TRF	126–127

## 2.3   Estimation Errors of Updated Input-Output Tables

We measured (1) the deviations of updated tables $\left(IO_{u,v}\right)$ from the prior matrix $\left(IO_{u,v}^0\right)$ and (2) their estimation errors as deviations from the true IO table for 2005 $\left(IO_{u,v}^{2005}\right)$. The deviations of estimated tables from the prior values were mostly around 10%

in the LS cases but more extensive in the CE cases (Figure 2.1). The total of deviations from

the prior $IO_{u,v}^0$ were computed by such indicators as standardized total percentage error

(STPE), mean absolute difference (MAD), Theil's U_2, root mean squared error (RMSE), and

mean absolute percentage error (MAPE) (Figure2.2).[5] If we assumed that true values

should be close to the priors, we could interpret those deviations from the priors as

"estimation errors," which indicate accuracy of our estimates, but we would find the results

counterintuitive. That is, the rich information cases (Micro) appear to perform more poorly

than the poor information cases (Macro) in Figure 2.2. These counterintuitive results

originated from the nature of their minimization problems: The constraints in the rich

information cases (2.2)–(2.3) were stricter than the constraints (2.4)–(2.5) in the poor

information cases. Even if we assumed oppositely that the true table should be quite

different from the prior matrix, the large deviations, however, would not immediately

indicate better accuracy. Only comparison with the true table can tell us accuracy of our
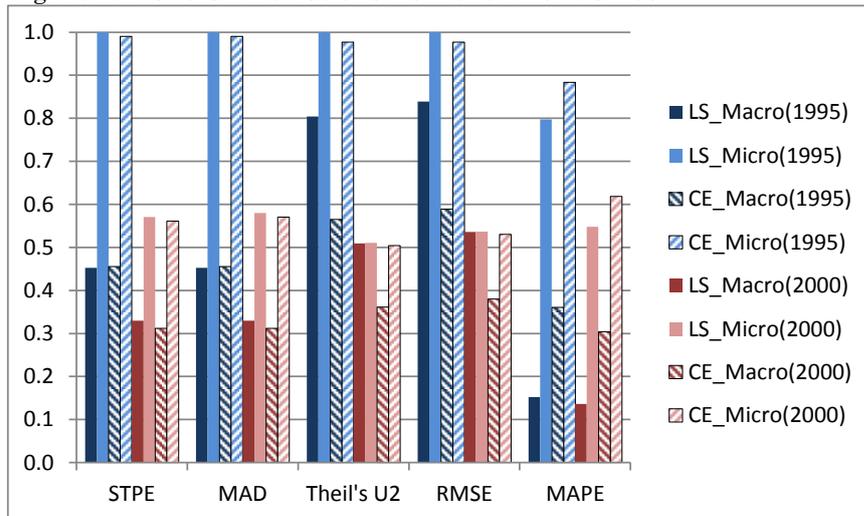
updated tables.

---

[5] They are defined as $STPE = \sum_{u,v}\left|IO_{u,v} - IO_{u,v}^0\right| \Big/ \sum_{u,v} IO_{u,v}^0$, $MAD = \frac{1}{u \cdot v}\sum_{u,v}\left|IO_{u,v} - IO_{u,v}^0\right|$,

$U_2 = \sqrt{\sum_{u,v}\left(IO_{u,v} - IO_{u,v}^0\right)^2} \Big/ \sqrt{\sum_{u,v}\left(IO_{u,v}^0\right)^2}$, $RMSE = \sqrt{\frac{1}{u \cdot v}\sum_{u,v}\left(IO_{u,v} - IO_{u,v}^0\right)^2}$, and

$MAPE = \frac{1}{u \cdot v}\sum_{u,v}\left|IO_{u,v} / IO_{u,v}^0 - 1\right|$ to compute deviations from the priors $IO_{u,v}^0$. We can replace

$IO_{u,v}^0$ with $IO_{u,v}^{2005}$ to compute the accuracy of the new tables. In all these indicators, smaller values

indicate smaller deviations/better accuracy.

Figure 2.1: Deviations from Prior Values [%]
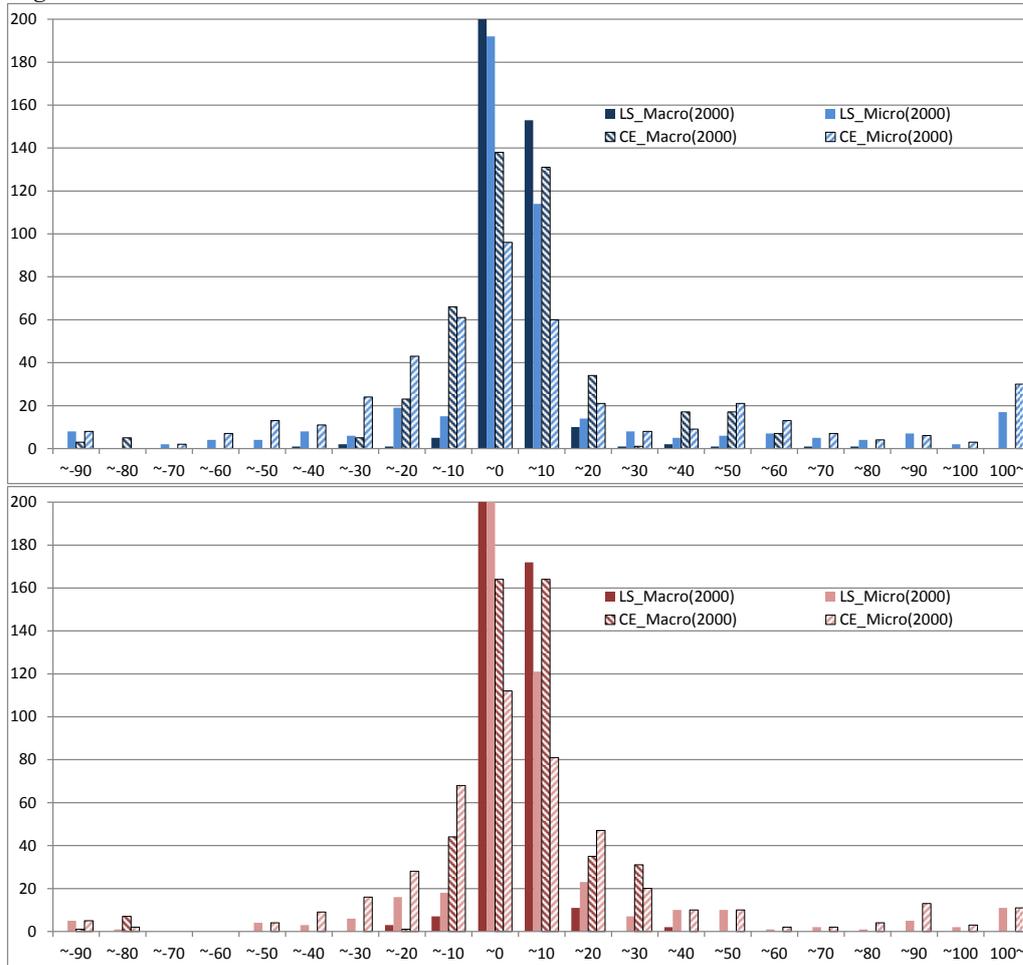
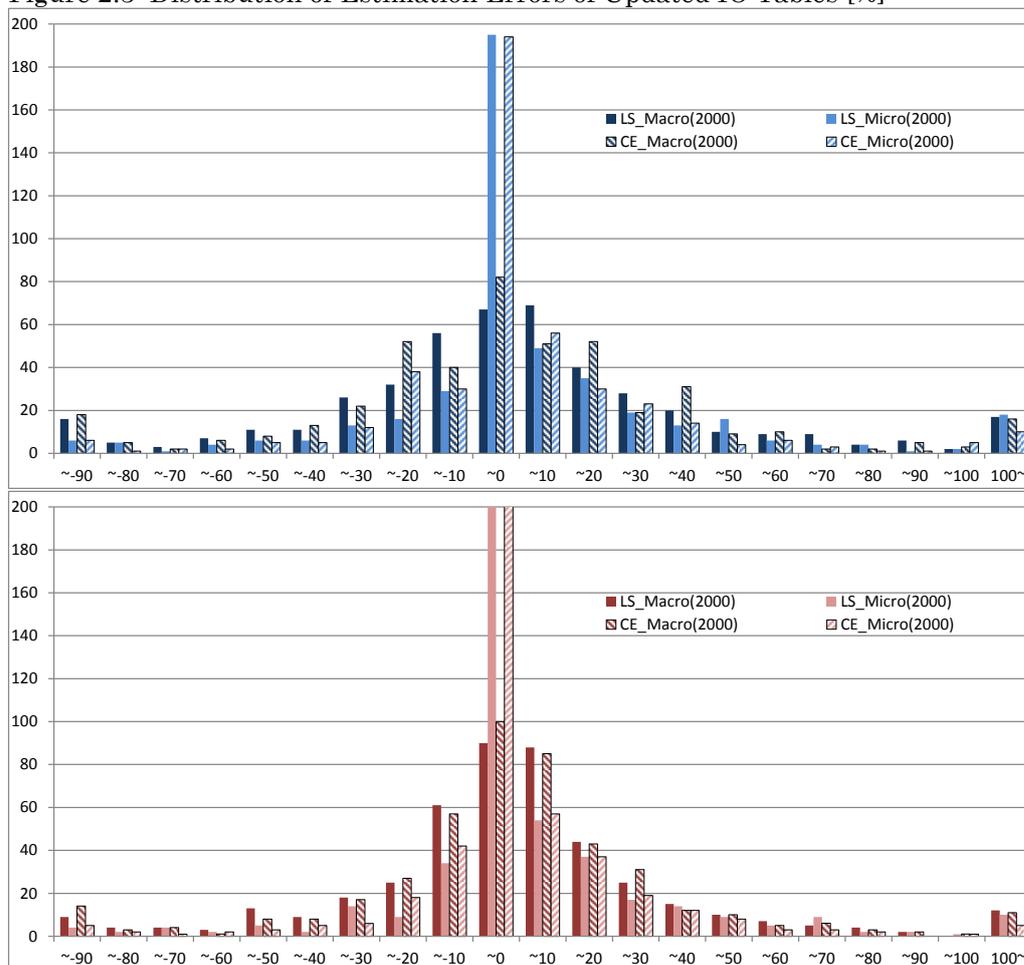Note: Deviations of cells with priors equal to zero are not considered.

Figure 2.2: Overall Deviations from the Prior Values

Note: All the indicators but MAPE are normalized so that their maximum values are equal to 1.00.

We measured the true accuracy of the updated tables by comparing them with the true table for 2005 (Figure 2.3). The cases based on the table for 2000 were likely to achieve better accuracy than the cases based on the table for 1995. The rich information cases (Micro) were likely to outperform the poor information cases (Macro) (Figure 2.4).

Figure 2.3: Distribution of Estimation Errors of Updated IO Tables [%]

Note: Estimation errors of true cell values equal to zero are not considered.

Figure 2.4: Overall Estimation Errors
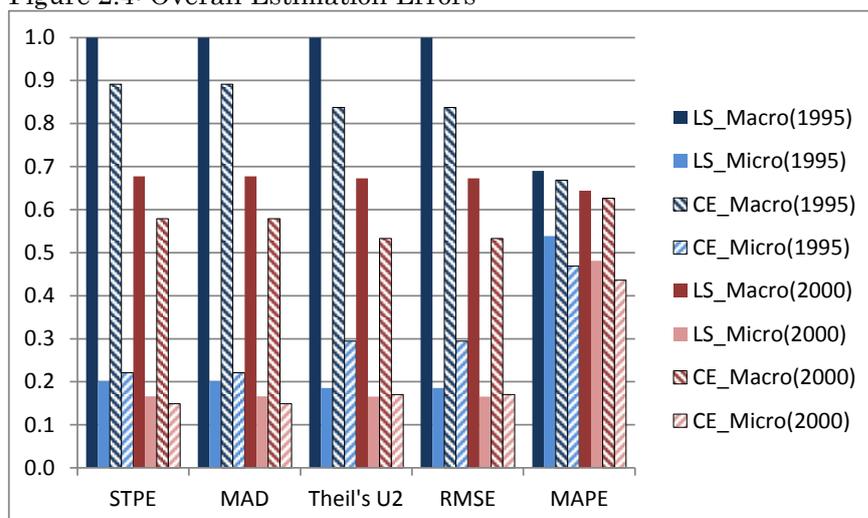
Note: All the indicators but MAPE are normalized so that their maximum values are equal to 1.00.

In contrast to these straight-forward results about richness and age of information fed for updating IO tables, we need a careful examination of the performance of the matrix balancing methods (Figure 2.4). In the poor information cases (Macro), the CE method was likely to generate more accurate estimates. In the rich information cases (Micro), the LS method outperformed the CE method by the four indicators (all sectors but MAPE) in the cases with the IO table for 1995 and did so only by Theil's U_2 and RMSE in the cases with the table for 2000. The relative performance of these two methods was not clear-cut.

Regarding the magnitude of errors, MAPE indicates sizable average estimation errors per cell reaching 44–69%. These large errors may be carried over to CGE analysis and result in large prediction errors. We examine the significance of these estimation errors transmitted into the prediction errors of CGE analysis in the next section.

3. Prediction Errors in CGE Analysis

3.1 A CGE Model and Simulation Scenarios

To examine prediction errors in CGE analysis, we used a static small open single-country CGE model by Hosoe et al. (2010) calibrated to SAMs constructed with those

IO tables estimated in the previous section and the actual direct tax data in the national account tables. [6] We modified the original CGE model to treat negative values in IO tables because the updated and the true IO tables for 2005 carried negative values in household consumption of other mining (MIN) and steel (STL) and in some of government consumption and investment uses. We set the household consumption of these two goods as exogenous.[7] We also assumed that all the government consumption and investment uses were exogenous while lump-sum direct taxes and savings filled the gaps in their budget constraints.

There were two simulation scenarios considered. Scenario 1 assumed a uniform cut of import tariff rates by 2% points for all the 14 tradable goods (i.e., all but ELY, TWG, TRS, and SRV).[8] Scenario 2 assumed exogenous hikes of international prices (in USD) of oil, coal, and natural gas (OIL) and petroleum and coal products (P_C) by 100%.

We used those eight IO tables estimated in Section 2, as well as true IO tables for

[6] As Hosoe et al. (2010) demonstrated, we can construct a SAM with only an IO table and information about direct tax payments. The amount of direct taxes was estimated 38,285.5 billion Japanese yen, consisting of current taxes and net social contributions earned by the general government in 2005 (Economic and Social Research Institute (2012)).

[7] In the IO table for 1995, the household consumption of other mining goods (MIN) was positive; therefore, the updated tables based on this one also had positive consumption of this good due to the sign condition (2.6). In contrast, as the IO tables for the other two years showed negative household consumption of this good, the estimated tables based on them showed negative consumption. For simplicity, we assumed that consumption of both other mining (MIN) and steel (STL) goods in the CGE model was exogenous in all the cases.

[8] We did not assume a common scenario of abolition of import tariffs, because estimated tariff rates implied by the estimated IO tables for the base run were different and, thus, would make the impact of tariff abolition different among different cases.

1995 (True(1995)) and for 2000 (True(2000)).[9] We compared their CGE simulation results based on these ten IO tables to the simulation result based on the true IO table for 2005 (True(2005)). We set the Armington's (1969) elasticity at two for all the sectors.[10]

3.2   Prediction Errors in Simulation Results

In Scenario 1 (uniform tariff cut), the output changes predicted by the CGE models calibrated to these ten IO tables were consistent for all the sectors but MAN with the prediction by the true CGE model (True(2005)) (Figure 3.1). Prediction accuracy significantly differed among these ten cases depending on the IO table estimation methods. The cases with IO tables updated with the poor information (Macro) and the naïve cases with the true but old IO tables (True(1995) and True(2000)) clearly showed poor predictability. Regarding the matrix balancing methods, the CE method combined with the poor information (Macro) performed better than the LS method in IO table estimation (Figure 2.4) but poorer in CGE analysis (Table 3.1). In the rich information cases (Micro), as the IO estimation errors did not differ much by the matrix balancing methods, neither did the CGE prediction errors.

---

[9] We adjusted the scaling of these two true IO tables for 1995 and 2000 according to the GDP growth among these three periods. The GDP in 2005 was larger than that in 1995 by 0.5% and that in 2000 by 2.5%. This scaling-up process, however, changed results only in levels but not in rates or ratios.

[10] The Appendix presents the results of sensitivity analysis where the Armington (1969) elasticity was perturbed by 30% upward and downward.

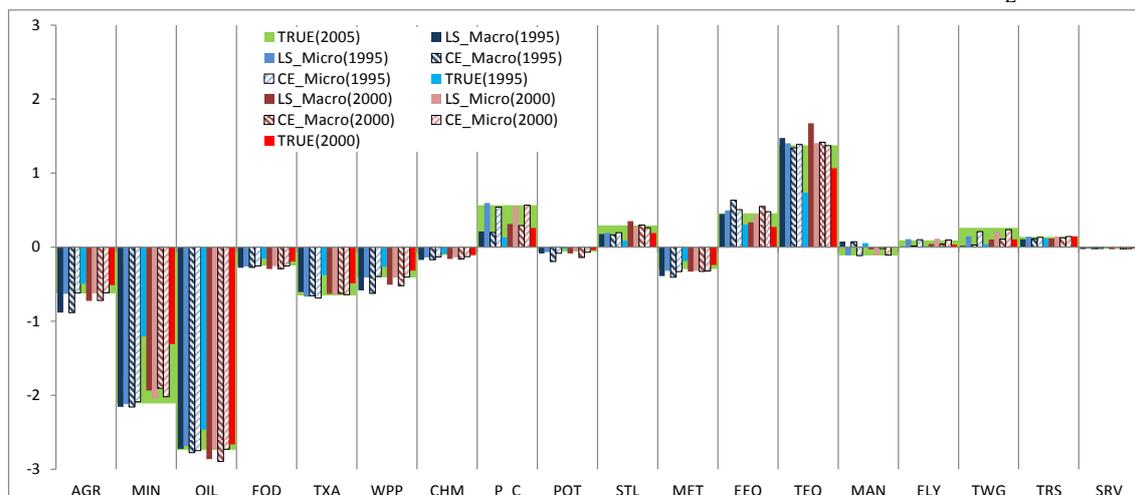Figure 3.1: Result of Scenario 1 (Uniform Tariff Cuts) [Changes from the Base ($\frac{\Delta Z}{Z}$), %]

Table 3.1

18-Sector Simple Mean of Absolute Prediction Errors of Sectoral Output Changes ($\frac{\Delta Z}{Z}$) [%]

Case	Scenario 1	Scenario 2
LS_Macro(1995)	41.1	65.7
LS_Micro(1995)	10.4	29.4
CE_Macro(1995)	53.3	61.2
CE_Micro(1995)	10.9	27.3
True(1995)	51.8	69.9
LS_Macro(2000)	28.9	38.2
LS_Micro(2000)	7.2	22.8
CE_Macro(2000)	32.4	35.5
CE_Micro(2000)	6.0	23.2
True(2000)	33.4	44.2

Note: Absolute Prediction Error Rate=$100 \left| \frac{\Delta Z}{Z} \middle/ \frac{\Delta Z^{true}}{Z^{true}} - 1 \right|$.

While Scenario 1 assumed a uniform shock to all the tradable sectors, Scenario 2 assumed exogenous shocks on international prices only for the two oil-related sectors. The results of Scenario 2 showed that prediction of output changes was generally correct but qualitatively erroneous for a few sectors (Figure 3.2). One sector is the textiles and apparel sector (TXA) in all ten cases. The others were the steel (STL) and the pottery sectors (POT) in CE_Macro(1995). The change of chemical output (CHM) was correctly predicted in quality but suffered a large prediction error in quantity. The prediction errors attributable to age of the original IO tables (1995 vs. 2000) and richness of additional information (Macro vs.

Micro) were found similar to those found in Scenario 1 (Table 3.1). Regarding the matrix balancing methods, the CE method performed better than the LS method in all cases but CE_Micro(2000).

Figure 3.2: Result of Scenario 2 (Petroleum Price Hike) [Changes from the Base ($\frac{\Delta Z}{Z}$), %]
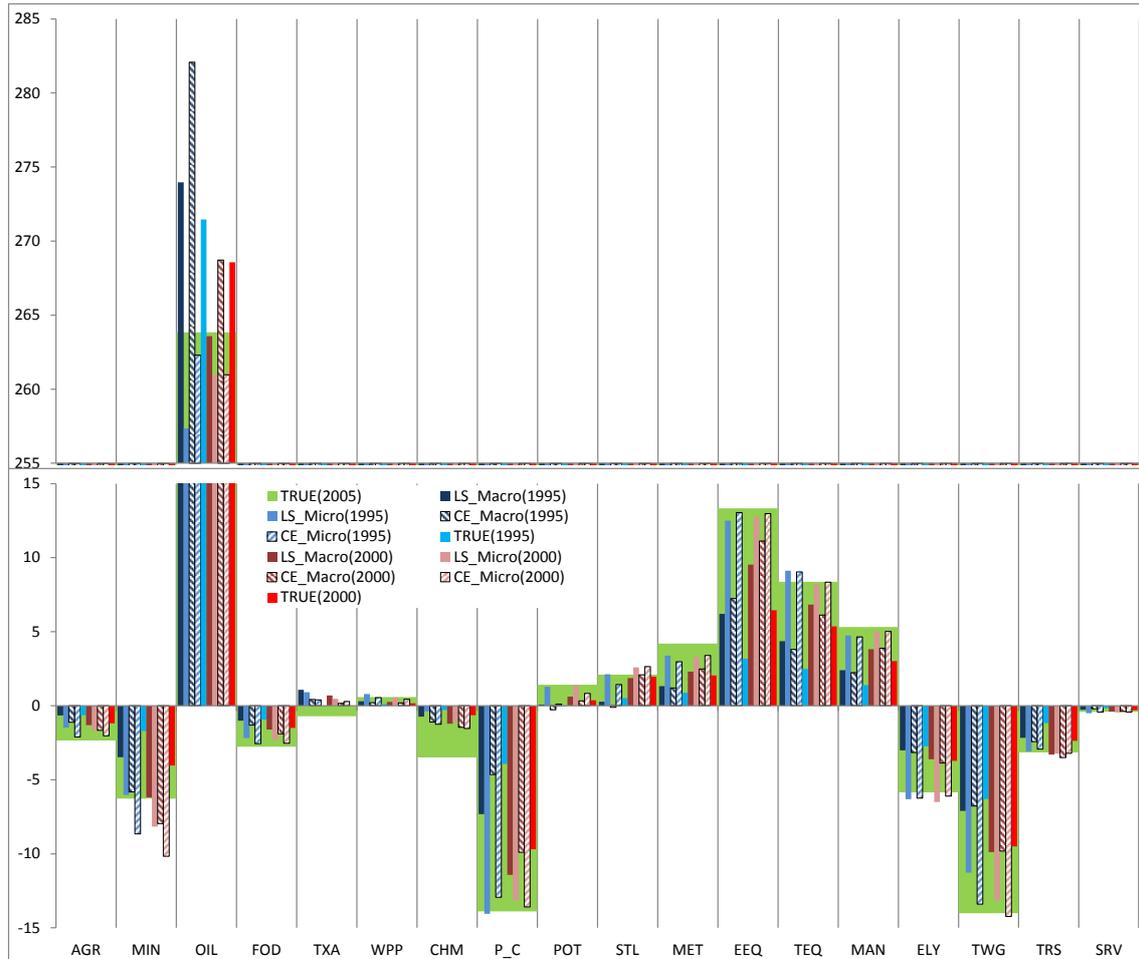

Throughout these two simulation exercises, prediction of the welfare impact measured with the Hicksian equivalent variations was qualitatively correct (Figure 3.3). The naïve case of True(1995) always led to the quantitatively poorest prediction, which was a half or one-third of the true impact. The other naïve case of True(2000) showed prediction errors comparable to those in Macro(1995).

When only macroeconomic data were available, CGE analysis based on the young

IO table for 2000 was found performing better. In contrast, when microeconomic data were fed into the IO table estimation process, estimates of welfare impact were made almost perfectly, irrespective of age and richness of data and matrix balancing methods (Figure 3.3).

Figure 3.3: Welfare Impact in Scenario 1 (Left) and Scenario 2 (Right) [Unit: bil. JPY]
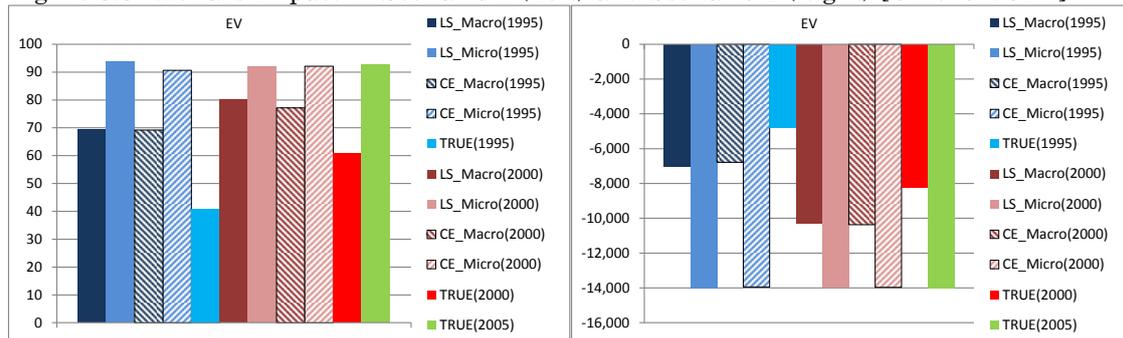
Note: Measured by Hicksian equivalent variations.

4.   Concluding Remarks

In this paper, we studied the effect of estimation methods on estimation errors of updated IO tables and prediction errors in CGE analysis using the linked IO tables for 1995, 2000, and 2005. Our results suggested that younger IO tables with richer additional information were likely to reduce estimation errors in IO tables. Smaller estimation errors in IO tables would also reduce prediction errors in CGE analysis. Even if we used IO tables suffering more or fewer estimation errors, the predicted signs of output changes were correct in most CGE simulations.

The effect of matrix balancing methods was complex. In the poor information cases with only macroeconomic data, the CE method successfully reduced estimation errors in IO tables. In the rich information cases with microeconomic data, the CE method sometimes worsened the accuracy of IO table estimates–though the penalty was not found to be significant. Moreover, as the results of Scenario 1 (uniform tariff cut) demonstrated, the CE method combined with macroeconomic data reduced the estimation errors in IO tables but

increased the prediction errors in CGE analysis.

In our experiments updating IO tables, we used data reported in the true IO table for 2005 assuming these target year data were all true. However, if the target year data suffer measurement errors, they may lead to larger estimation errors in updated IO tables and thus to larger prediction errors in CGE analysis. In this sense, the estimation errors and the prediction errors demonstrated here should be considered as their minimum values. In this case, it may be useful to employ an updating technique that explicitly considers measurement errors in available information as proposed by Robinson et al. (2001).

For better accuracy of IO table estimation and CGE analysis, more recent and more detailed data are requested but rarely or insufficiently available. CGE modelers, who are not necessarily professional statisticians, often have difficulty in this regard. Our results suggest that CGE modelers can make qualitatively correct predictions even if their IO tables are old or suffer errors compared with the true target year IO tables. This approach would be useful, especially for practical purposes, when we assess the impact of, for example, free trade agreements, whose details can be frequently revised and updated in the process of their negotiations.

References

Armington, P. (1969) "A Theory of Demand for Products Distinguished by Place of Production," *IMF Staff Papers* 16(1), 159–178.

Bonfiglio, A., Chelli, F. (2008) "Assessing the Behaviour of Non-Survey Methods for Constructing Regional Input-Output Tables through a Monte-Carlo Simulation," *Economic Systems Research* 20(3), 243–258.

Cardenete, M. A., Sancho, F. (2004) "Sensitivity of CGE Simulation Results to Competing SAM Updates," *Review of Regional Studies* 34(1), 37–56.

Economic and Social Research Institute (2012) *Annual Report on National Accounts for 2012*, Cabinet Office, the Government of Japan.

Golan, A., Judge, G., Miller, D. (1996) *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, Wiley.

Hosoe, N., Gasawa, K., Hashimoto, H. (2010) *Textbook of Computable General Equilibrium Modelling*, Palgrave Macmillan.

< http://www.gams.com/modlib/libhtml/stdcge.htm> (Accessed on August 27, 2013)

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

Jalili, A. R. (2000) "Evaluating Relative Performances of Four Non-Survey Techniques of Updating Input-Output Coefficients," *Economics of Planning* 33, 221–237.

Miller, R. E., Blair, P. D. (2009) *Input-Output Analysis: Foundations and Extensions*, Second Edition, Cambridge.

Ministry of Internal Affairs and Communications (2011) *Heisei 7-12-18-nen Setsuzoku-sangyo-renkan-hyo [Linked Input-Output Tables for 1995, 2000, and 2005]*, the Government of Japan, Tokyo.

< http://www.stat.go.jp/data/io/071217.htm> (Accessed on June 16, 2013)

Robinson, S., El-Said, M. (2000) "CESAM.GMS: Cross Entropy SAM Estimation," [GAMS computer code]. <http://www.gams.com/modlib/libhtml/cesam.htm> (Accessed on August 24, 2013)

Robinson, S., Cattaneo, A., El-Said, M. (2001) "Updating and Estimating a Social Accounting Matrix Using Cross Entropy Methods," *Economic Systems Research* 13(1), 47–64.

Appendix: Sensitivity Analysis of CGE Results

Following the usual practice of CGE analysis, we conducted a sensitivity analysis regarding our simulation results with respect to a key parameter of Armington's (1969) elasticity by perturbing it by 30% upward and downward. The results of Scenario 1 were found to be robust irrespective of the assumed parameter values (Figure A.1). The results of Scenario 2 led to a similar conclusion about output changes for all sectors, except chemical (CHM), pottery (POT), and steel (STL). Their predicted output changes were qualitatively erroneous in some cases with old or poor information (Figure A.2). The perturbation little affected the conclusion regarding their average prediction errors (Table A.1). One exception was found regarding the performance of matrix balancing methods in Scenario 1. While the LS method outperformed the CE method in Micro(1995) in the central and lower elasticity cases, the conclusion was found to be opposite in the higher elasticity case.

Figure A.1: Result of Scenario 1 (Uniform Tariff Cuts) [Changes from the Base ($\frac{\Delta Z}{Z}$), %]

30% Lower Armington Elasticity Case
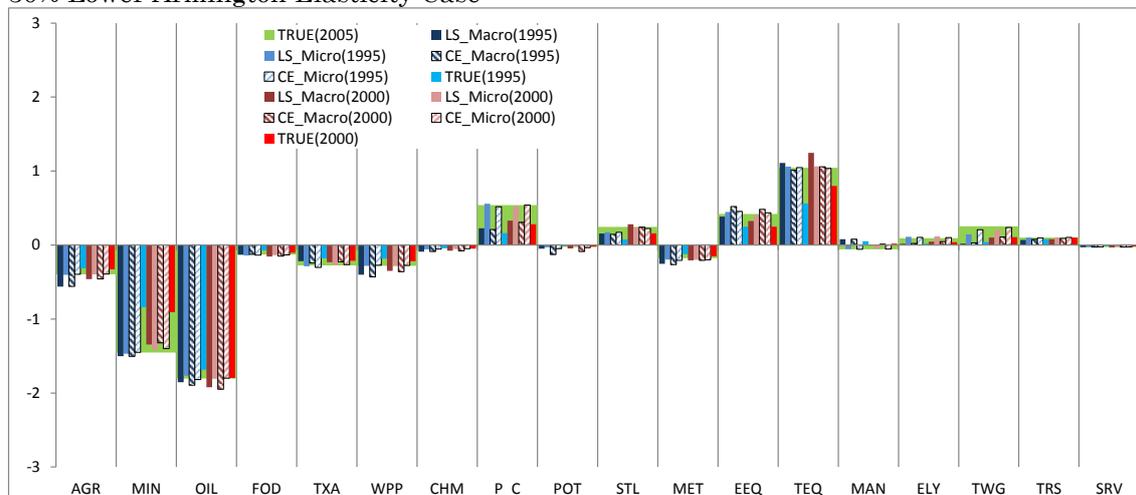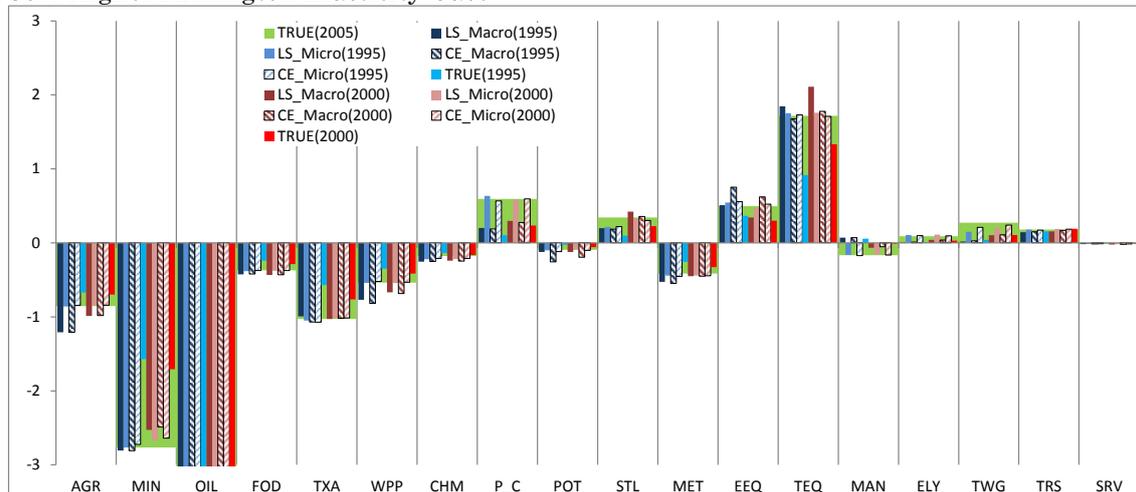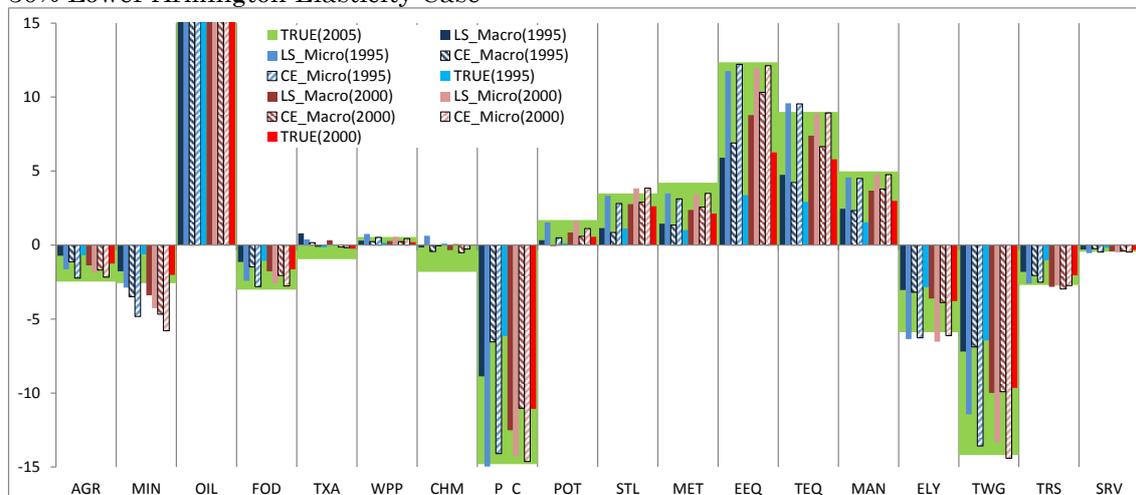

30% Higher Armington Elasticity Case

Figure A.2: Result of Scenario 2 (Petroleum Price Hike) [Changes from the Base ($\frac{\Delta Z}{Z}$), %]

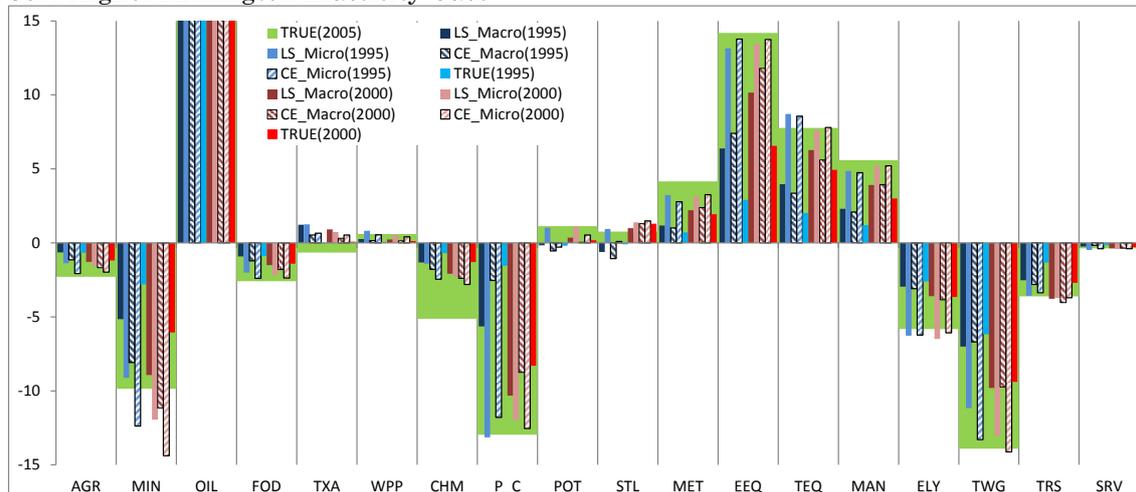30% Lower Armington Elasticity Case

30% Higher Armington Elasticity Case

Table A.1: 18-Sector Simple Mean of Absolute Prediction Errors of Sectoral Output Changes ($\frac{\Delta Z}{Z}$) [%]

Elasticity	Scenario 1			Scenario 2		
	Lower	Central	Higher	Lower	Central	Higher
LS_Macro(1995)	56.1	41.1	37.9	58.7	65.7	75.2
LS_Micro(1995)	14.5	10.4	9.8	26.8	29.4	34.5
CE_Macro(1995)	75.0	53.3	48.8	56.4	61.2	74.2
CE_Micro(1995)	17.7	10.9	9.6	25.5	27.3	34.3
True(1995)	53.4	51.8	53.6	67.5	69.9	74.6
LS_Macro(2000)	38.9	28.9	27.6	36.9	38.2	43.1
LS_Micro(2000)	10.6	7.2	6.6	21.2	22.8	28.6
CE_Macro(2000)	47.3	32.4	29.8	36.7	35.5	41.4
CE_Micro(2000)	10.4	6.0	5.1	23.0	23.2	29.4
True(2000)	35.4	33.4	34.9	42.8	44.2	49.8

Note: Absolute Prediction Error Rate=$100\left|\frac{\Delta Z}{Z}\middle/\frac{\Delta Z^{true}}{Z^{true}}-1\right|$. The central case results are those shown in Table 3.1.