Int. Journal of Economics and Management 14 (3): 401-415 (2020)

IJEM
International Journal of Economics and Management
Journal homepage: http://www.ijem.upm.edu.my

Disaggregating Economic Micro-Data under Limited Sectoral Breakdown

ANITHA ROSLAND[a], MUHAMMAD DAANIYALL ABD RAHMAN[a*],
MOHD YUSOF SAARI[a, b], CHAKRIN UTIT[c] AND UMI ZAKIAH NORAZMAN[c]

[a]*School of Business and Economics, Universiti Putra Malaysia, Malaysia*
[b]*Centre for Future Labour Market Studies, Menara PERKESO, Malaysia*
[c]*Institute of Tropical Agriculture and Food Security, Universiti Putra Malaysia, Malaysia*

ABSTRACT

Limited sectoral breakdown or unmatched sectoral details is some of recurring issues in input-output impact assessment analysis. This issue will refrain researchers from obtaining comprehensive and detailed results, which also lead to some valuable information loss and limited policy recommendation. This study examines the robustness of economic micro-data disaggregation when a limited sectoral breakdown is available to the data users. Specifically, we test a disaggregation technique on national income and employment survey-based data set namely the Household Income Survey (HIS) 2005 and compute the multiplier indicators for the respective data, using Input Output Table 2005. Our results show that the technique can be used to produce consistent results with some limited caveats to avoid result misinterpretations.

JEL Classification: C67, D57
Keywords: Disaggregation; input-output analysis; multiplier effect

Article history:
Received: 19 October 2020
Accepted: 28 November 2020

* Corresponding author: Email: daaniyall@upm.edu.my

INTRODUCTION

Input-output analysis has been widely used to conduct impact assessment and policy evaluation. The strength of the input-output model in linking supply and demand components in a complex production network enables a comprehensive analysis to be undertaken, and the same time, helps to answer plethora of issues, encompassing economic, social and environment research areas. As part of the streams of input-output analysis application, Saari et al. (2016) use input-output price model to simulate the potential impacts of minimum wages regulation on major ethic groups as the repercussion of possible labour cost surge in response to price increase. To examine the impact of oil price shocks on the Malaysian economy, Maji et al. (2017) combine input-output and econometric model in an integrated methodological framework, which allows them to assess the shocks on macroeconomic variables such as tax revenues, employment, labour income and gross domestic product (GDP).

Most interestingly, input-output analysis can also be integrated with other economic, socio-economic and environmental indicators such as household's income (Hassan et al., 2016; Saari et al., 2015; Zhang et al., 2017; Abd Rahman et al., 2019), employment (Alsamawi et al., 2014a; Alsamawi et al., 2014b), biodiversity (Lenzen et al., 2012) and greenhouse gas emissions (Oita et al., 2016; Owen, 2017; Steen-Olsen et al., 2014; Zhang et al., 2015). The data integration and extended analysis can be carried out by supplementing the standard monetary input-output data with the indicators data through, a so-called "satellite accounts".

Such data integration and extended analysis would of course provide meaningful assessment to the researchers as well as to the policy makers because, for example, affixing an employment indicator to the input-output model can help to explain to what extent does final demand (domestic or exports) of an industry has created job(s) *directly* to the industry itself and *indirectly* to the other industries within the national economic system. For instance, Yen et al. (2015) use input-output table and sectoral employment data to compare employment multiplier effects of private and public higher education institutions in Malaysia. Meanwhile, Mohamad Akhir et al. (2018) investigate the impact of *batik* industry to the national employment, by augmenting the industry in the Malaysia input-output table. To a greater extent and put it in a global context, using a global multi-region input-output model (MRIO) [1], Alsamawi et al. (2014a) and Alsamawi et al. (2014b) utilize employment satellite account, distinguished by more than 180 countries, to analyse employment and inequality footprints embodied in a country imported goods and services[2].

Above studies have shown that incorporating the micro-data and input-output data in one analytical framework is truly useful, not only at the national level, but also at the global level. However, limited sectoral breakdown of a survey-based micro-data, of which the statistical institution tends to aggregate the information for general public use[3], hinders the users from performing detailed analysis using the information-rich dataset. This imposes challenging efforts for the users to harmonize the industrial classification of input-output data and micro-data, if a more detailed analysis is desired.

In such cases, users have to aggregate the input-output data in order to match micro-data industrial classification before performing their analysis, for which highly aggregated sectoral breakdown may be insufficient to yield accurate results (Steen-Olsen, 2014). For example, Yen et al. (2015) aggregate 94 input-output table sectors into 16 sectors in order to fit the workforce data that only available at 16 sectoral details. In that regards, Lenzen (2011) argues that the data aggregation may not only lead to loss of valuable information but may also lead to significant errors as well, if disaggregated sector characteristics have large variances to which both conditions could lead to 'aggregation-biased' (see also, Kymn, 1990; Lenzen et al., 2004; Su et al., 2010). For instance, if a small 'Highway, Bridge and Tunnel Operation Services' sector and a large 'Telecommunication' sector are aggregated into 'Transport, Storage and Communication', where most of the employment may be associated with smaller 'Highway, Bridge and Tunnel Operation Services' sector, it is

[1] Until now, there is a number of works has been devoted to develop the inter-country input-output data, where all countries national input-output data were compiled in a single large-scale database (see for example: Abd Rahman et al., 2017; Dietzenbacher et al., 2013; Lenzen et al., 2013).

[2] The employment satellite account was constructed using labour force survey from various countries and years that are mostly available from International Labour Organization (ILO) website. See also http://www.worldmrio.com for global MRIO data and various economic, socio-economic and environment indicators.

[3] The standard classification for national data collection is following to the Malaysia Standard Industrial Classification (MSIC: 2000, 2008), which contain up to 5-digit industrial category (more than 1000 industries). However, the most detailed data would not have been published, even if with special data request, due to data confidentiality.

ended up yielding a grossly underestimated employment intensity for 'Highway, Bridge and Tunnel Operation Services', and a grossly overestimated employment intensity for 'Telecommunication'.

Having this conundrum in mind, the main aim of this paper is to examine the robustness of economic micro-data disaggregation when a limited sectoral breakdown is presence. Figure 1 shows the main idea of this paper and illustrates the flow of disaggregating the broad economic sectors. Suppose that there are only three broad economic sectors for employment data published by the statistical institution, namely Agriculture, Manufacturing and Services. This limited information would not give further information on the employment characteristics of the detailed sectors within the respective broad sector categories. Thus, the economic impact analysis using input-output modelling would also be producing aggregated overview of the employment structure of the whole economic system, overlooking specific impact of labour-intensive sub-sectors within the broad sector categories. On the other hand, this would also end up into overestimating employment intensity of the capital-intensive sub-sectors. We further compute the effect of this issue in the Methodology section by using hypothetical example.

To detail out the broad sector categories (e.g., from Agriculture into Paddy, Palm Oil and Live stocks, as shown in (Figure 1), we adopt disaggregation technique proposed by Lenzen (2011), who has proven that the technique is reliable and robust for environmental data disaggregation by employing generic input-output datasets and Monte Carlo analysis. Specifically, we apply the similar technique to disaggregate income and employment micro-data from broad sectors into detailed sectors, before calculating multiplier indicators for each of the data using input-output modelling. Then, we compare the results with the actual income and employment multipliers at detailed sectors from which the Household Income Survey (HIS) micro-data for income and employment are extracted. On the other extent, we are unable to locate other studies in the literature that apply the technique on economic survey-based data set, which support the novelty of our work.

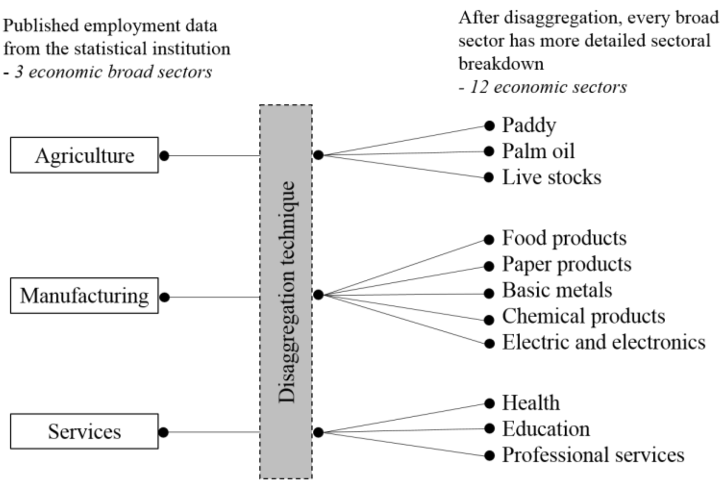

Figure 1 Flow of work to disaggregate broad economic sector categories to more detailed sectoral breakdown

The contributions of our work are manifested in two areas. First, our work could support the application of the disaggregation technique for time-series sectoral data. This is because the survey-based micro-data was published periodically, while for example, employment and income data at the aggregated level, are available on a time-series basis. In fact, the use of time-series input-output data has been widely used to conducting year-to-year growth analysis (see Los et al., 2015), exploiting various micro-data variables, such as gender, ethnicity, skill levels and types of occupation (see Saari et al., 2016; Saari and Pei, 2013). Second, more detailed analysis at sectoral level can be performed, preventing the loss of important information when aggregated data is used. Given that the robustness of the results is statistically justified, our findings recommend for more potential applications on impact assessment and policy evaluation studies using input-output analysis.

We organise our work as follows. In Section 2, we present the standard input-output model and disaggregation technique, followed by some selected robustness checking measures. We show the results in Section 3 and draw the conclusion in Section 4.

METHODOLOGY AND DATA SOURCES

In this section, we firstly discuss the standard input-output model by paying specific attention on calculating multiplier indicators, which will be the basis for our result comparison. Next, we present the disaggregation techniques, followed by some selected analytical tools for robustness checking. Finally, we describe the data sources.

Methodology

Input-Output Analysis: The Multiplier Effect

Multiplier effect is one of standard indicators calculated using the input-output model in a way to measure the direct and indirect effect of increase in a desired indicator (i.e. employment, income, value-added) induced by each additional unit of final demand for a particular sector. The multiplier analysis is started from basic input-output linear equation as follows:

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{Y} \tag{1}$$

where $\mathbf{x}$ is the vector of final output; $\mathbf{A}$ ($\mathbf{A} = \mathbf{Z}\hat{\mathbf{x}}^{-1}$) stands for the domestic input coefficient matrix of input-output transaction matrix $\mathbf{Z}$, which is based on $N \times N$ matrix and a hat indicates its diagonalised matrix (non-zeros elements in the diagonal part and zeros elsewhere); and, $\mathbf{Y}$ represents the final demand vector. Eq. (1) can be further re-arranged to form into fundamental input-output identity introduced by Leontief (1936):

$$\mathbf{x} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y} = \mathbf{L}\mathbf{Y} \tag{2}$$

where $\mathbf{I}$ is the identity matrix with $N \times N$ dimension (ones in the diagonal part and zeros elsewhere) and $\mathbf{L} = (\mathbf{I} - \mathbf{A})^{-1}$ is the Leontief inverse, measuring the output of an industry must generate that is necessary to satisfy a unit of final demand from other industry.

Finally, the multiplier, $\mathbf{m} = \mathbf{v}(\mathbf{I} - \mathbf{A})^{-1} = \mathbf{v}\mathbf{L}$ where $\mathbf{v} = \mathbf{V}\hat{\mathbf{x}}^{-1}$ is a $1 \times N$ vector containing sectoral production factor usage (**V**) per unit of total output. In this study, the **V** is referred to income and employment, where the multipliers will be derived from these two factors.

The Effect of Disaggregated Versus Aggregated Multipliers: A Hypothetical Example

In this sub-section, we explain the extent to which the effect of aggregated income multiplier would limit analytical assessment using input-output modelling. To do so, we employ a hypothetical example of 3×3 and 12×12 input-output tables, denote as **G** and **G*** respectively, where **G** is the aggregated matrix of **G***. Both matrices come with their respective final demand and value-added matrices, and employment vectors. Similar employment multiplier calculation (as explained in previous section) is applied but for the sake of limited space, the input-output tables and employment multiplier results can be obtained in the Appendix.

In a nutshell, our result shows that aggregation of employment data causes limited assessment of the impact of employment multiplier on the economy. As refer to Table A4, our argument holds, for example, when Sector A has 0.13 unit of employment multiplier. This implies that Sector A, as a whole, generates 0.13 unit of employment to the economy as a result of a unit increase of final demand for Sector A product. This result is relatively low compared with employment multiplier generated by Sector B (i.e. 0.15) that obviously gives more impact on the economy by creating more jobs. Therefore, for policy makers, promoting Sector B, as a whole, via private or government investment could be better off than Sector A, in term of generating employment opportunities. But, this also becomes a misleading interpretation because Sector A could have more prospective sub-industries that can be promoted as well, which is hindered due to limited sectoral multiplier indicators. Thus, this requires for more detailed data to provide better explanation.

In Table A3, the employment data is more disaggregated to match detailed input-output tables as Sector A has more detailed sectoral breakdown, which are A1, A2 and A3 and likewise for Sector B and C. Table A5 shows the disaggregated employment multipliers. Apparently, each of the sectors has different employment multiplier magnitude, which to some extent, even larger than their respective aggregated employment multiplier outcomes (compare employment multipliers in Table A4 with Table A5). For example, Sector A1 has employment multiplier of 0.19 compared with Sector A that only generates 0.13. This result shows that the aggregated data would leverage the 'true' sectoral multiplier effect as a result of loss of some important information, especially due to data aggregation. This condition will affect the policy decision making differently as more detailed data allows policy makers to implement cost-effective and targeted investment.

Data Disaggregation Technique

Lenzen (2011) proposed a data disaggregation technique to overcome the recurring issue of different economic classifications, which typically impedes data harmonisation in the input–output analysis. Although this approach was introduced for disaggregating environmental data to match sectoral classification in the input–output database, our work extends this technique to disaggregate data from the official economic survey (i.e. HIS), especially in the presence of limited sectoral breakdown of the raw data. The standard formulation for this approach, or so-called the 'map' matrix, is written as follows:

$$\mathbf{M} = (\widehat{\mathbf{C}\hat{\mathbf{x}}_p\mathbf{1}_p})^{-1}\mathbf{C}\hat{\mathbf{x}}_p = \hat{\mathbf{x}}_q^{-1}\mathbf{C}\hat{\mathbf{x}}_p \tag{3}$$

where $\mathbf{C}$ is a binary concordance matrix sized p x q. Let $p > q$, so that the columns of $\mathbf{C}$ contain the disaggregated classification. $\mathbf{x}_p$ is a row vector containing the p-classed proxy variable to be used for prorating, $\hat{\mathbf{x}}_p$ is the diagonal matrix corresponding to $\mathbf{x}_p$, $\mathbf{x}_q = \mathbf{C}\hat{\mathbf{x}}_p\mathbf{1}_p$, is the q-classed representation of $\mathbf{x}_p$ where $\mathbf{1}_p$ is an p-classed summation vector, and a hat means diagonalised matrix, where the non-zero element on the diagonal part of the matrix and zero elsewhere.

For illustration of this technique, we form the ground for disaggregating the HIS income data for year 2005 and for the employment data set, the similar steps can be methodically replicated. Note that we have the most detailed HIS income and employment data at 120 sectors for 2005, meaning that the HIS classification is perfectly matched with the 2005 input-output table sectoral details. We use the detailed data as a benchmark in testing the technique's robustness to disaggregate the economic survey data. In particular, the disaggregation can be done in a number of steps:

1. Aggregate the actual income data (i.e., HIS, 2005), denoted as $\mathbf{L}$ with dimension 120×1, into randomly q broad sectors vector (termed as 'aggregation factor'), denoted as $\mathbf{K}$ with dimension $q \times 1$.
2. Create a concordance matrix; $\mathbf{C}$ with dimension $q \times 120$, indicating that sector $i \in q$ corresponds to sector $j \in p$. Each row must have at least one non-zero entry; thus, each aggregated sector in n corresponds to at least one sector in m.
3. We use total output (TO), value-added (VA), and compensation of employee (CE) from the 2005 input–output table as a disaggregator weight proxy row vector, $\mathbf{x}$ with dimension 1×120, respectively.
4. Compute the map matrix, $\mathbf{M}$, as in Eq. (1) with dimension $q \times 120$.
5. Multiply $\mathbf{M}^T$ to $\mathbf{K}$ and obtain $\mathbf{K}^*$. Superscript T refers to matrix transposition.
6. Calculate the multiplier, $\mathbf{m}$ for both $\mathbf{L}$ and $\mathbf{K}^*$, followed by calculating differences between these using selected analytical tools (see Section 2.1.3)
7. Repeat step (1) to (6) using different aggregated income data.
8. Apply similar procedures on the employment data.

We use three arbitrary aggregation factors (f) on the actual data: 1) four broad sectors; 2) one-digit MSIC 2008 classification with 17 sectors; and 3) ISIC revision three with 35 sectors. The aggregation factors are supplemented in the Table A1.

It is important to note that the use of disaggregator weight proxy applied in step 3 is subject to the availability of data. In our case, the use of total output (TO), value-added (VA), and compensation of employee (CE) comes with two justifications. First, these are the indicators that are publicly available and easily accessed by the users. Second, these indicators can be directly observed in the input-output table and available at most detailed sectoral breakdown, which allows users to obtain corresponding classification between external data and input-output table. The usage of IO information for data disaggregation has been applied by Lenzen et al. (2017).

Multiplier Analytical Checking

According to Lahr (2001), the measure of distance and association serve two purposes: 1) to measure the ability of models to produce accurate results and 2) to determine the statistical significance of the difference between the actual and the estimated data. To compare the sectoral employment and income multiplier indicators obtained from the actual and estimated data, we adopt several measures. There are multiple measures have been used in the literature for comparing input-output matrices (see Geschke et al., 2014; Knudsen and Fotheringham, 1986; Lenzen et al., 2009; Wiebe and Lenzen, 2016), however there is no single statistical test for assessing the accuracy of two corresponding matrices (Butterfield and Mules, 1980). Therefore, for our work we consider five measures to examine the distances between the multipliers. Suppose $\mathbf{m}^{(1)}$ and $\mathbf{m}^{(2)}$ are two vectors of total multipliers of equal length, q. The norms used for the comparison are listed in
Table 1:

Table 1 Distance norms to measure multiplier differences

Name	Abb.	Formula	Reference
Mean absolute difference	MAD	$\frac{\sum_j(\lvert m_j^{(1)} - m_j^{(2)}\rvert)}{n}$	Geschke et al., 2014; Miller and Blair, 2009; Saari et al., 2014; Wiebe et al., 2016
Root mean squared error	RMSE	$\frac{\sqrt{\left[\sum_j(m_j^{(1)} - m_j^{(2)})^2\right]}}{n}$	Geschke et al., 2014; Wiebe et al., 2016
Pearson's correlation coefficient	1-CORR	$1 - \frac{\sum_j\left[\left(m_j^{(1)} - \bar{M}^{(1)}\right)\left(m_j^{(2)} - \bar{M}^{(2)}\right)\right]}{\sqrt{\sum_j\left(m_j^{(1)} - \bar{M}^{(1)}\right)^2 . \sum_j\left(m_j^{(2)} - \bar{M}^{(2)}\right)^2}}$	Gallego and Lenzen, 2009; Geschke et al., 2014; Wiebe et al., 2016
Isard-Romanoff similarity index	DSIM	$\frac{1}{n}\sum_{j=1}^{n}\frac{\lvert m_j^{(1)} - m_j^{(2)}\rvert}{\lvert m_j^{(1)}\rvert + \lvert m_j^{(2)}\rvert}$	Gallego et al., 2009; Steen-Olsen et al., 2016
$\boldsymbol{\chi^2}$ distribution of absolute difference	CHI	$\sum_j\frac{\left(m_j^{(1)} - m_j^{(2)}\right)}{m_j^{(2)}}$	Gallego et al., 2009

Note: n = no. of sectors; $m_j^{(1)}, m_j^{(2)}$= cell-wise element of respective multiplier indicator; $\bar{\mathbf{M}}^{(1)}, \bar{\mathbf{M}}^{(2)}$= mean of the respective multiplier vectors.

The MAD provides measure of the direct relative absolute difference between the $\mathbf{m}^{(1)}$ and $\mathbf{m}^{(2)}$. The MAD metric takes a value from 0 to 1, implying the closer MAD is to zero, the higher similarity between $\mathbf{m}^{(1)}$ and $\mathbf{m}^{(2)}$. Likewise, RMSE, DSIM and CHI carry similar interpretation as that of MAD metric, but values for MAD are always larger than or equal to those for RMSE (Geschke, 2014)[4]. We take reciprocal value of the correlation coefficient (1-CORR), as similarly applied by Gallego et al. (2009), Geschke et al. (2014) and Wiebe et al. (2016), so that the interpretation would be similar to that of the other norms (low 1-CORR value determines small difference, while high 1-CORR value corresponds oppositely).

Data Sources

The analysis in this paper makes use of Malaysia input-output table for year 2005 and income and employment data extracted from the Household Income Survey (HIS) for similar reference year. Both data were compiled by the Department of Statistics Malaysia (DOSM) in accordance with the Malaysia Standard Industrial Classification (MSIC). In term of sectoral coverage, both data comprise of 120 production sectors, encompassing agriculture, manufacturing and services industries. As the main of aim this paper is to analyse the extent to which the disaggregated multiplier estimates are reliable under the condition of limited sectoral

[4] RMSE is also known as EMD or Euclidean metric distance (Geschke et al., 2014; Wiebe et al., 2016).

breakdown, the HIS data is then being aggregated into several arbitrary aggregation factor (as previously discussed in the Methodology section).

RESULT AND DISCUSSION

This section presents the results obtained from our analysis. To view outcomes of the simulation, we first visualize the calculated multiplier magnitudes from actual HIS income and employment data against the estimated multipliers from different aggregation factors and proxy weights for both datasets. To investigate in depth the reliability of the calculated multipliers, we explain the outcomes from the analytical norms for each of the multipliers.

Visualizing the Differences in Multipliers Estimates

Visual comparison (see Lenzen et al., 2013; Abd Rahman et al., 2017; Lenzen et al., 2017) is important to give some early indications of the reliability of the technique used in disaggregating economic survey-based micro data. Each of the data points in Figure 1 represents respective income multiplier magnitude obtained from iterative procedures as described in step-by-step procedures above, which was executed based on 4 aggregation factors and 3 proxy weights (total output (TO), value-added (VA), compensation of employee (CE)). Therefore, there are 24 iterations have been made in this analysis, including both income and employment multiplier estimates. For ease of interpretation, we refer 'actual multipliers' to the magnitudes obtained from actual HIS disaggregated data at 120 sectors, whereas 'estimated multipliers' are the results calculated from the pre-defined arbitrary aggregation factors.

At the macro view, we find that there are minimal deviations between the estimated multipliers and the actual multipliers, particularly on the left hand side of the figure. However, as we move to the downstream sectors (to the right of the figure), we observe rather volatile patterns. For example, in sector 108 (Rental and Leasing; see data points within the rectangular shape), the dispersion is quite large, where significant difference is portrayed by the estimated multiplier from 4 sectors aggregation factor using CE proxy weights (upper side of the rectangle: refer to "×" symbol).

On the other hands, by using similar proxy weights (see data points within the dotted circle), a closer estimated multiplier is obtained when more disaggregated information is realized (that is disaggregating 35 sectors aggregation factor into 120 sectors using CE proxy weights: represented by "+" symbol). As such, richer information would minimize the relative deviation of both multipliers.

An important intermediate conclusion could be drawn from these results are performing disaggregation into a more detail data is plausible and able to produce better results under the condition of: 1) the raw data (i.e. HIS data in our case) is available at the utmost detailed level as to avoid crude approximation using the disaggregation tool, and 2) suitable proxy weight is used to closely represent the respective industry economic structure.

With respect to the comparison between the visualization of the income multipliers (in Figure 2) and employment multipliers (in Figure 3), it is found that, by and large, performing disaggregation technique on micro-data is likely to produce consistent multipliers. To be exact, based on the sectoral count, at least 95% of the estimated multipliers are consistent to their actual multipliers.

Other than that, our results are close to estimation by Ahmad Fuad and Puasa (2011) who calculate income multiplier for national key economic areas (NKEA) using 2005 input-output data. For example, they find that the income multiplier for Education sector is 0.58, while our findings are ranged from 0.55 to 0.57 that vary depending on the initial sectoral aggregation factors and proxy weights used.

For the employment multiplier, there are some significant deviations for a few sectors. For example, Sector 1 (Paddy), the actual employment multiplier is 0.286, while the estimated multipliers from all iterations are significantly lower than 0.100, meaning that, in this case, the employment multiplier for the paddy sector is somehow underestimated during the disaggregation procedure. Similar results are also obtained from other agricultural-based industries: Sector 2 (Food Crops) and Sector 3 (Vegetable), and manufacturing-based industries: Sector 22 (Grain Mills) and Sector 39 (Wooden and Cane Container). Further, Figure 4 plots the variations of these sector estimated multipliers to that of the actual multipliers for CE proxy weights at different aggregation factors (particularly at the first quadrant of the plot), whereas the rest of the estimated multiplier data points nudge closely to the 45 degree solid line, implying the closeness of actual and estimated multipliers.

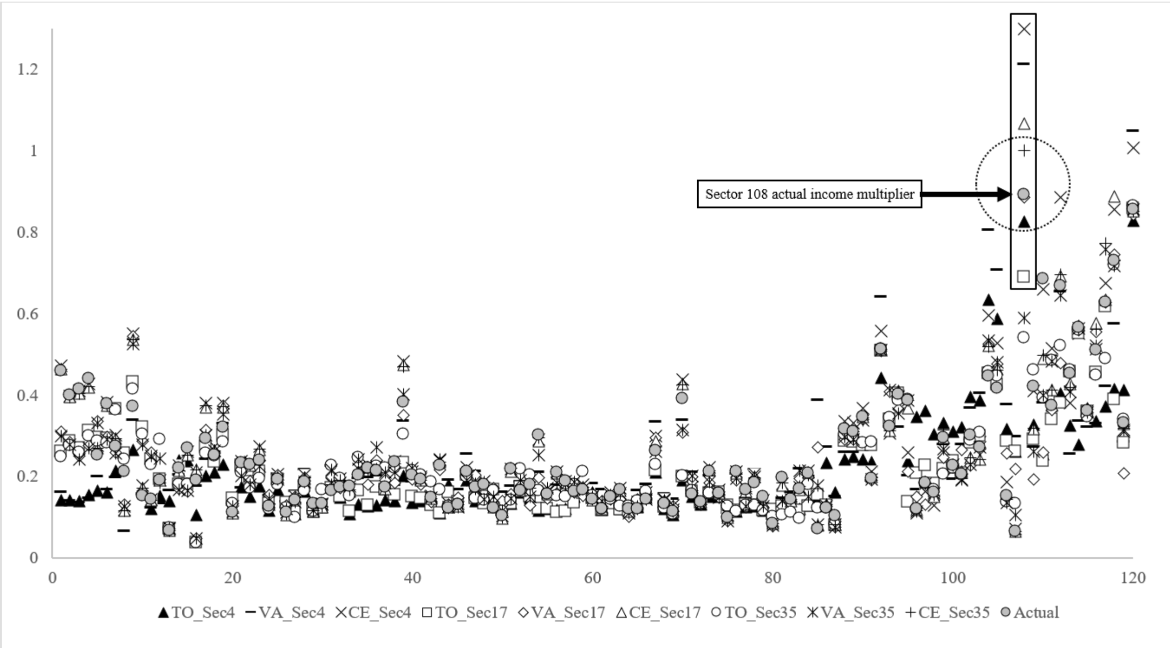

Figure 2 Income multipliers: Actual multipliers versus estimated multipliers at different aggregate factors and proxy weights

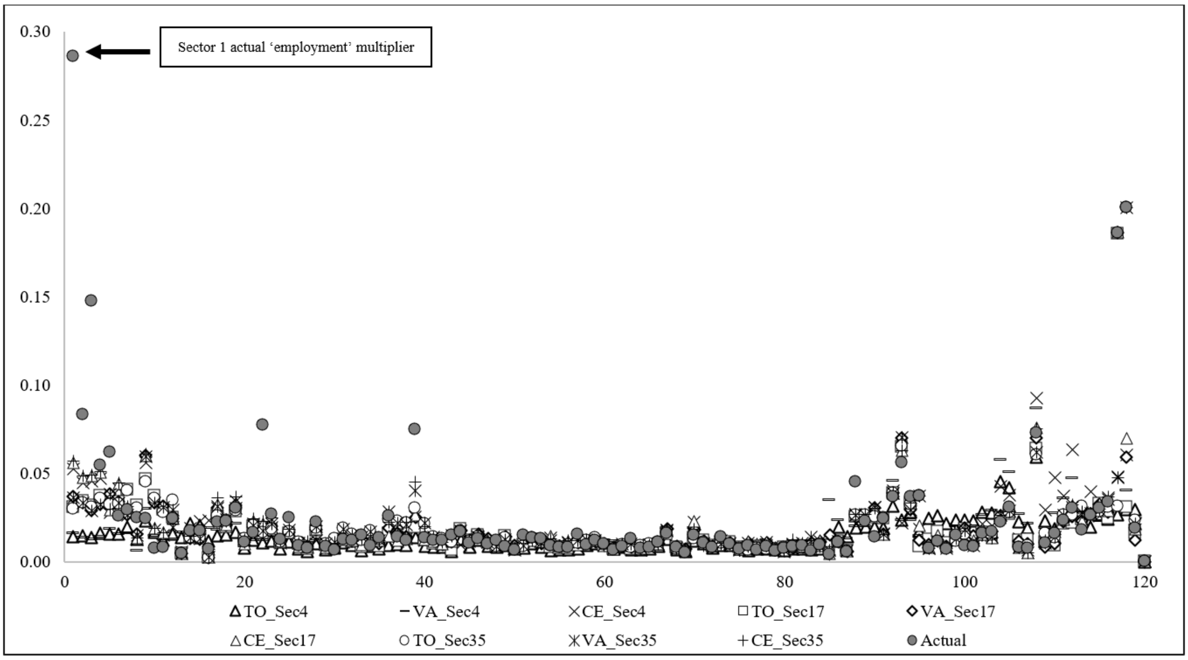

Figure 3 Employment multipliers: Actual multipliers vs. estimated multipliers at different aggregate factors and proxy weights for all 120 sectors

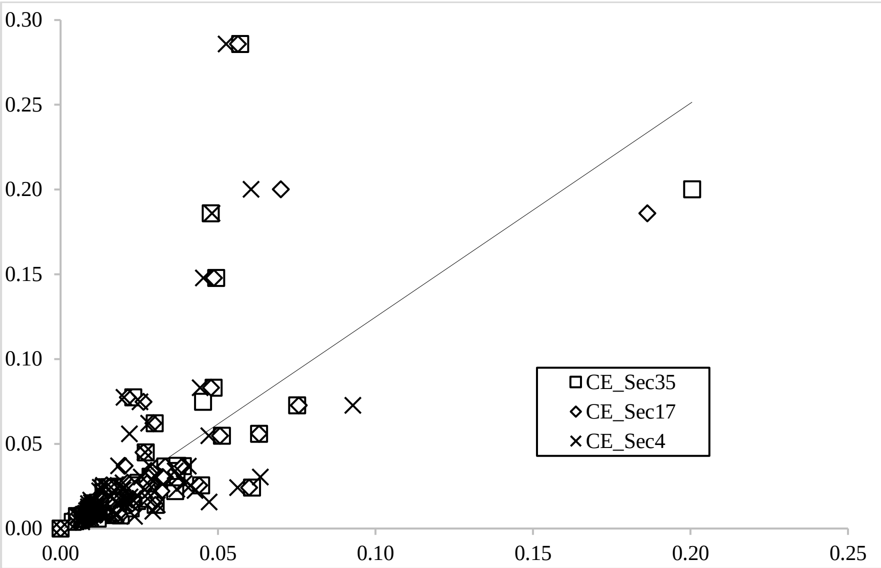

Figure 4 Selected actual versus estimated employment multipliers

Although there are some evidence of variability in the estimated employment multipliers, the main precautionary that need to be acknowledged is that the proxy weights used in this simulation is the monetary proxy weights, which to some extent conflicting to the real economic scenario of a particular sector. For example, agriculture sector, except for commodity products, is commonly associated to relatively lower output industry with rather labor intensive especially in an industrialized economic settings such as Malaysia. Therefore, adopting monetary proxy weight during disaggregation process could cause 'estimation bias' to the end results because the monetary values would underrepresent the nature of factor of production in the agricultural sector. However, in a different perspectives, comparing with all available proxy weights used in the simulations, the CE proxy weight outperforms the other proxy weights. In fact, this is within our expectation as the CE is the closest measure to represent labor composition of an industry.

The under- or over-representation condition can be improved if supporting information from other authorized agencies can be fitted in as the proxy weights. Note that the proxy weight vector does not necessarily come from a set of similar data set, but it also can be mixed up or 'hybridised' with other dataset as well in a way to produce the most disaggregated information. In other words, it can be gathered and constructed from different data sources. For example, Lenzen et al. (2017) combine the national input-output data, international commodity statistics (i.e. COMTRADE) and OECD statistics to construct the most detailed sectoral breakdown for the global MRIO database proxy weights. However, this procedure requires additional techniques that beyond the scope of this work.

Reliability of The Multipliers

Whilst the overall visualizations of the income and employment multipliers show some variability between actual and estimated multipliers, the existence of underestimated or overestimated multipliers, however, cannot be simply factored out. But such preliminary conceptions also should be supported with a battery of analytical measures in order to gauge statistical reliability and identify qualification of the analysis in way to grasping holistic accuracy of the estimations (Jensen, 1980). This procedure is widely applied in most input-output comparison studies (see references in
Table 1).

Table 2 Results of analytical norms for each aggregation factor and proxy weight

Aggregation factor	4 sector			17 sector			35 sector		
Proxy weight vector	TO	VA	CE	TO	VA	CE	TO	VA	CE
A. Income multiplier									
MAD	0.08	0.07	0.03	0.06	0.05	0.02	0.05	0.04	0.02
RMSE	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.00
1-CORR	0.27	0.24	0.03	0.17	0.12	0.03	0.11	0.09	0.02
DSIM	0.17	0.13	0.04	0.13	0.10	0.03	0.10	0.08	0.03
CHI	5.32	5.92	0.79	3.41	2.75	0.41	1.93	1.61	0.38
B. Employment multiplier									
MAD	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
RMSE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1-CORR	0.72	0.71	0.39	0.43	0.37	0.28	0.41	0.37	0.29
DSIM	0.23	0.24	0.16	0.15	0.15	0.12	0.13	0.13	0.11
CHI	1.38	1.55	1.04	0.93	0.91	0.64	0.87	0.86	0.64

Note: TO = Total output; VA=Value-added; CE= Compensation of employees.

Table 2 shows the result of analytical norms based on the comparison of actual and estimated income and employment multipliers using different aggregation factors and proxy weights. Based on these measures, there are three important findings that can be highlighted.

First, as we move from left to right of Table 2, we observe that all of the norms magnitudes for all iterations are approaching to zero, indicating there are shrinking differences between the actual and estimated income multipliers. Likewise, employment multipliers also portray similar trend changes. In addition, the correlation between each pair of norms is always positive and close to 0.99 to indicate the strength of the qualification analysis (see Table 3 and Table 4).

Second, in panel B of Table 2, it is noted that the employment multiplier shows less variability when comparing similar result's pattern as in the income multiplier (Table 2: Panel A). Apart from that, the magnitudes of RMSE are constant and close to zero at all iteration results, where combining the result of RMSE with other analytical norms signifies markedly the less dispersion between the actual and estimated employment multipliers.

From these reflections of the findings, we posit that as more detailed HIS data is known and used during the disaggregation procedure, the more accurate multipliers could be obtained, regardless of proxy weights particularly for employment data. In fact, the differences in multipliers estimates between 17-sector and 35-sector aggregation factor are relatively marginal to offset their reliability in the case of limited sectoral data.

Table 3 Coefficient of correlation between the distance norms for all iterations of income multiplier

	MAD	RMSE	1-CORR	DSIM	CHI
MAD	1.000	0.994	0.991	0.986	0.965
RMSE		1.000	0.982	0.974	0.963
1-CORR			1.000	0.970	0.981
DSIM				1.000	0.923
CHI					1.000

Table 4 Coefficient of correlation between the distance norms for all iterations of employment multiplier

	MAD	RMSE	1-CORR	DSIM	CHI
MAD	1.000	0.965	0.945	0.987	0.970
RMSE		1.000	0.987	0.974	0.977
1-CORR			1.000	0.962	0.972
DSIM				1.000	0.963
CHI					1.000

In contrast, a remarkable turnaround shall be expected if too aggregated employment data is perceived. As evident, there are significant drops of 1-CORR values particularly in the case of 4-sector employment multiplier when comparing with similar results for 17-sector and 35-sector aggregation factors, respectively. Otherwise, one would produce undesirable outcomes if extensively aggregated data is used to analyse economic indicators, which would eventually end up into result misinterpretations.

Other than that, it is also important to note that the income data is rather sensitive to the use of proxy weights, which warrants for a proper proxy selection. Still, the evidence from our finding suggests that the CE proxy weight outperforms the others, cut across all iterations. This finding is fundamentally independent of the size of the aggregation factors because this proxy weight does not influence the changes of analytical norms magnitude in any significant way, except for extremely broad sectors, as represented by the 4 broad sector category. It holds true for both income and employment multiplier statistical assessments. All in all, our findings in this section are consistent to validate our argument in the previous section and also in line to Lenzen (2011).

CONCLUSION

In this paper, we have applied a disaggregation technique for extricating aggregated sectoral breakdown of survey-based micro-data. The application of the technique enables data users to get more detailed 'estimated' industry data in the absence of actual data classification in order to perform comprehensive and extended assessment using the input-output analysis. To illustrate the method, we use total output (TO), value-added (VA), and compensation of employee (CE) as disaggregator weight proxies to disaggregate income and employment multiplier at detailed sectoral levels. Then, results of the multipliers are compared with the actual income and employment multipliers that derived from micro-data.

Key results show that applying the disaggregating technique on the income and employment survey-based micro data may produce consistent results. These findings are supported with statistical comparative measures that prove the reliability of the estimates. Therefore, from the policy makers viewpoint, analysis that based on detailed income and employment data is beneficial to provide a better sectoral level impact assessment because the multipliers generated from these estimates can inform them the potential cost and benefit of a particular policy measure, hence avoiding them from leveraging the sectoral impact that based on aggregated information. For example, employment multipliers can be used to formulate investment programmes that create more jobs and higher labour productivity for which not all sectors could demonstrate benefit from such policies. Thus, better understanding on the nature of the sector through detailed sectoral breakdown (identifying capital-intensive or labour-intensive sector) would allow the policy makers to design an appropriate measure at the targeted sector, which could not be realized when aggregated information used. Thus, we strongly argue that loss of information would be worse off than having less accurate estimated detailed information, particularly in the condition of limited sectoral breakdown of raw data.

Despite to the fact that it is preferable to acquire more detailed data for a better result, but it is also sufficient when a broader data is known. The disaggregation technique applied in this paper is proven to provide reliable estimates. However, it is important to note that the results obtained from the disaggregation technique are far from being perfect and needed to be interpreted cautiously. This is because the technique depends highly on the proxy weight to disaggregate the data. For example, the use of industry *value-added* as the proxy weight to disaggregate employment data would not be suitable compared with *compensation of employee*. In this case, *value-added* could overestimate the employment for the industry with higher value-added, even though the employment in that particular industry is relatively lower as a result of capital-intensive nature of production. However, this limitation is preventable if one could find additional secondary information, such as industrial administrative data, which can be used as alternative proxy.

REFERENCES

Abd Rahman, M. D., Los, B., Geschke, A., Xiao, Y., Kanemoto, K., and Lenzen, M. (2017) 'A Flexible Adaptation of the WIOD Database in a Virtual Laboratory', *Economic Systems Research*, *29*(2), pp. 187–208.

Abd Rahman, M. D., Saari, M. Y., Lenzen, M., and Malik, A. (2019) 'Skills and Ethnics Wage Inequalities Within the Global Value Chain: An Evidence from Malaysia', *Policy Studies*, pp. 1–20. https://doi.org/10.1080/01442872.2019.1657079

Ahmad Fuad, S. N., and Puasa, A. F. (2011) 'National Key Economic Area Multiplier Impact on Malaysian Economy: An Input-Output Analysis', *International Journal of Management Studies*, *18*(Special Issue), pp. 34–58.

Alsamawi, A., Murray, J., and Lenzen, M. (2014) 'The Employment Footprints of Nations', *Journal of Industrial Ecology*, *18*(1), pp. 59–70.

Alsamawi, A., Murray, J., Lenzen, M., Moran, D., and Kanemoto, K. (2014) 'The Inequality Footprints of Nations: A Novel Approach to Quantitative Accounting of Income Inequality', *PLos ONE*, *9*(10), pp. 1–10.

Butterfield, M., and Mules, T. (1980) 'A Testing Routine for Evaluating Cell by Cell Accuracy in Short-Cut Regional Input-Output Tables', *Journal of Regional Science*, *20*(3), pp. 293–310.

Dietzenbacher, E., Los, B., Stehrer, R., Timmer, M., and de Vries, G. (2013) 'The Construction of World Input-Output Tables in the WIOD Project', *Economic Systems Research*, *25*(1), pp. 71–98.

Gallego, B., and Lenzen, M. (2009) 'Estimating Generalized Regional Input–Output Systems: A Case Study of Australia', In M. Ruth and B. Davidsdottir (Eds.), *The Dynamic of Regions and Networks in Industrial Ecosystems* (pp. 55–88). Cheltenham, UK: Edward Elgar Publishing.

Geschke, A., Wood, R., Kanemoto, K., Lenzen, M., and Moran, D. (2014) 'Investigating Alternative Approaches to Harmonise Multi-Regional Input–Output Data', *Economic Systems Research*, *26*(3), pp. 354–385.

Hassan, A. A. G., Saari, M. Y., Utit, C., Hassan, A., and Harun, M. (2016) 'Estimating the Impact of GST on Cost of Production and Cost of Living in Malaysia', *Malaysian Journal of Economics*, *50*(2), pp. 15–30.

Jensen, R. C. (1980) 'The Concept of Accuracy in Regional Input-Output Models', *International Regional Science Review*, *5*(2), pp. 139–154. Retrieved from http://irx.sagepub.com/content/5/2/139.abstract

Knudsen, D. C., and Fotheringham, A. S. (1986) 'Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modelling', *International Regional Science Review*, *10*(2), pp. 127–147.

Kymn, K. O. (1990) 'Aggregation in Input–Output Models: a Comprehensive Review, 1946–71', *Economic Systems Research*, *2*(1), pp. 65–93.

Lahr, M. (2001) 'A Strategy For Producing Hybrid Regional Input-Output Tables', In M. Lahr and E. Dietzenbacher (Eds.), *Input-Output Analysis: Frontier and Extensions* (1st ed.). Palgrave MacMillan, UK.

Lenzen, M. (2011) 'Aggregation and Disaggregation in Input-Output Analysis of the Environment', *Economic Systems Research*, *23*(1), pp. 73–89.

Lenzen, M., Gallego, B., and Wood, R. (2009) 'Matrix Balancing Under Conflicting Information', *Economic Systems Research*, *21*(1), pp. 23–44.

Lenzen, M., Geschke, A., Abd Rahman, M. D., Xiao, Y., Fry, J., Reyes, R., Dietzenbacher, E., Inomata, S., Kanemoto, K., Los, B., Moran, D., Bäumen, H. S. in den, Tukker, A., Walmsley, T., Wiedmann, T., Wood, R., and Yamano, N. (2017) 'The Global MRIO Lab - Charting the World Economy', *Economic Systems Research*, *29*(2), pp. 158–186.

Lenzen, M., Moran, D., Kanemoto, K., Foran, B., Lobefaro, L., and Geschke, A. (2012) 'International Trade Drives Biodiversity Threats in Developing Nations', *Nature*, *486*, pp. 109–112.

Lenzen, M., Moran, D., Kanemoto, K., and Geschke, A. (2013) 'Building Eora: A Global Multi-Region Input-Output Database at High Country and Sector Resolution', *Economic Systems Research*, *25*(1), pp. 20–49.

Lenzen, M., Pade, L.-L., and Munksgaard, J. (2004) 'CO2 Multipliers in Multi-region Input-Output Models', *Economic Systems Research*, *16*(4), pp. 391–412.

Leontief, W. W. (1936) 'Quantitative Input-Output Relations in the Economic System of the United States', *Review of Economics and Statistics*, *18*(3), pp. 105–125.

Los, B., Timmer, M., and de Vries, G. (2015) 'How Important are Exports for Job Growth in China? A Demand Side Analysis', *Journal of Comparative Economics*, *43*, pp. 19–32.

Maji, I. K., Saari, M. Y., Habibullah, M. S., and Utit, C. (2017) 'Measuring the Economic Impacts of Recent Oil Price Shocks on Oil-dependent Economy: Evidence from Malaysia', *Policy Studies*, *38*(4), pp. 375–391. https://doi.org/10.1080/01442872.2017.1320706

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

Mohamad Akhir, N. H., Ismail, N. W., and Utit, C. (2018) 'Malaysian Batik Industry Contribution Analysis Using Direct and Indirect Effects of Input-Output Techniques', *International Journal of Business and Society*, *19*(1), pp. 181–194.

Oita, A., Malik, A., Kanemoto, K., Geschke, A., Nishijima, S., and Lenzen, M. (2016) 'Substantial Nitrogen Pollution Embedded in International Trade', *Nature Geoscience*, *9*, pp. 111–115.

Owen, A. (2017) *Techniques for Evaluating the Differences in Multiregional Input-Output Databases: A Comparative Evaluation of CO2 Consumption-Based Accounts Calculated Using Eora, GTAP and WIOD*, Switzerland: Springer.

Saari, M. Y., Abdul Rahman, M. A., Hassan, A., and Habibullah, M. S. (2016) 'Estimating the Impact of Minimum Wages on Poverty Across Ethnic Groups in Malaysia', *Economic Modelling*, *54*, pp. 490–502.

Saari, M. Y., Dietzenbacher, E., and Los, B. (2015) 'Sources of Income Growth and Inequality Across Ethnic Groups in Malaysia, 1970-2000', *World Development*, *76*, pp. 311–328.

Saari, M. Y., Hassan, A., Abd Rahman, M. D., and Mohamed, A. (2014) 'Evaluation of the Relative Performance of RAS and Cross-Entropy Techniques for Updating Input-Output Tables of Malaysia', *Malaysian Journal of Economic Studies*, *51*(2), pp. 217–229.

Saari, M. Y., and Pei, J. (2013) 'Skilled-Biased Export Growth: Accounting for the Changes in Labor Contents Across Ethnic Groups in Malaysia (1991-2000)', *The Singapore Economic Review*, *58*(2), pp. 1–25.

Steen-Olsen, K., Owen, A., Barret, J., Guan, D., Hertwich, E. G., Lenzen, M., and Wiedmann, T. (2016) 'Accounting for Value-Added Embodied in Trade and Consumption: An Intercomparison of Global Multiregioal Input-Output Databases', *Economic Systems Research*, *28*(1), pp. 78–94.

Steen-Olsen, K., Owen, A., Hertwich, E. G., and Lenzen, M. (2014) 'Effects of Sector Aggregation on CO2 Multipliers in Multiregional Input-Output Analyses', *Economic Systems Research*, *26*(3), pp. 284–302.

Su, B., Huang, H. C., Ang, B. W., and Zhou, P. (2010) 'Input–output Analysis of CO2 Emissions Embodied in Trade: The Effects of Sector Aggregation', *Energy Economics*, *32*(1), pp. 166–175.

Wiebe, K. S., and Lenzen, M. (2016) 'To RAS or Not to RAS? What is the Difference in Outcomes in Multi-Regional Input–Output Models?', *Economic Systems Research*, *28*(3), pp. 383–402.

Yen, S. H., Ong, W. L., and Ooi, K. P. (2015) 'Income and Employment Multiplier Effects of the Malaysian Higher Education Sector', *Margin—The Journal of Applied Economic Research*, *9*(1), pp. 61–91.

Zhang, J., Yu, B., Cai, J., and Wei, Y.-M. (2017) 'Impacts of Household Income Change on CO2 Emissions: An Empirical Analysis of China', *Journal of Cleaner Production*, *157*(Supplement C), pp. 190–200. https://doi.org/https://doi.org/10.1016/j.jclepro.2017.04.126

Zhang, W., Peng, S., and Sun, C. (2015) 'CO2 Emissions in the Global Supply Chains of Services: An Analysis Based on a Multi-Regional Input–Output Model', *Energy Policy*, *86*(Supplement C), pp. 93–103. https://doi.org/https://doi.org/10.1016/j.enpol.2015.06.029

APPENDIX

Table A1: Aggregation factors used in the disaggregation steps described in the main text

Broad categories (4 sectors)	MSIC 2000 (17 sectors)	ISIC rev.3 code (35 sectors)
Primary industries	Agriculture Hunting and Forestry Fishing	Agriculture, Hunting, Forestry and Fishing
	Mining and Quarrying	Mining and Quarrying
Manufacturing	Manufacturing	Food, Beverages and Tobacco Textiles and Textile Products Leather, Leather and Footwear Wood and Products of Wood and Cork Pulp, Paper, Paper, Printing and Publishing Coke, Refined Petroleum and Nuclear Fuel Chemicals and Chemical Products Rubber and Plastics Other Non-Metallic Mineral Basic Metals and Fabricated Metal Machinery, not elsewhere classified Electrical and Optical Equipment Transport Equipment Manufacturing, not elsewhere classified; Recycling
Construction and utilities	Electricity Gas and Water Supply	Electricity, Gas and Water Supply
	Construction	Construction
Services	Wholesale and Retail Trade; Repair of Motor Vehicles, Motorcycles and Personal and Household Goods	Sale, Maintenance and Repair of Motor Vehicles and Motorcycles; Retail Sale of Fuel Wholesale Trade and Commission Trade, Except of Motor Vehicles and Motorcycles Retail Trade, Except of Motor Vehicles and Motorcycles; Repair of Household Goods
	Hotel and Restaurant	Hotels and Restaurants
	Transport, Storage and Communications	Inland Transport Water Transport Air Transport Other Supporting and Auxiliary Transport Activities; Activities of Travel Agencies Post and Telecommunications
	Financial Intermediation	Financial Intermediation
	Real Estate, Renting and Business Activities	Real Estate Activities Renting of Machinery & Equipment and Other Business Activities
	Public Administration and Defence; Compulsory Social Security	Public Admin and Defence; Compulsory Social Security
	Education	Education
	Health and Social Work	Health and Social Work
	Other Community, Social and Personal Service Activities	Other Community, Social and Personal Services
	Private Households with Employed Persons Extra-territorial Organisations and Bodies	Private Households with Employed Persons

Note: ISIC rev. 3 code of 35 sectors is following to sectoral classification in WIOD database (see (Dietzenbacher et al., 2013)

Table A2: Matrix **G**

.	A	B	C	Final Demand	Total Output
A	9	134	9	79	231
B	19	181	107	589	896
C	21	118	264	544	947
Value-added	182	463	567		
Total Input	231	896	947		
Employment	19	52	190		

Table A3: Matrix **G***

	A1	**A2**	**A3**	**A4**	**B1**	**B2**	**B3**	**B4**	**C1**	**C2**	**C3**	**C4**	**Final Demand**	**Total Output**
A1	0	0	0	0	59	0	3	0	0	2	1	0	26	91
A2	0	7	0	0	1	9	0	1	0	1	2	0	10	33
A3	0	0	1	0	1	0	55	1	0	0	0	0	42	100
A4	0	0	0	0	0	0	4	0	3	0	0	0	1	7
B1	3	1	0	0	49	0	5	0	0	6	9	1	109	182
B2	0	0	0	0	0	9	1	2	4	4	0	6	35	62
B3	5	4	5	0	3	4	66	24	28	14	1	20	157	330
B4	0	1	0	0	0	1	1	14	1	7	0	6	289	321
C1	1	1	1	0	2	2	8	6	15	1	3	19	82	140
C2	2	2	1	0	17	4	19	16	7	7	2	6	118	202
C3	0	0	0	0	0	0	0	0	0	2	3	6	35	47
C4	8	1	3	0	8	5	15	16	8	18	3	163	308	557
Value-Added	72	15	89	6	41	28	153	241	75	139	24	329		
Total Input	91	33	100	7	182	62	330	321	140	202	47	557		
Employment	12	2	4	0	7	5	15	25	22	31	8	129		

Table A4: Total employment multiplier matrix for **G**

	A	**B**	**C**
A	0.09	0.02	0.00
B	0.01	0.08	0.01
C	0.03	0.05	0.29
Total	0.13	0.15	0.30

Note: Summation may not equal due to rounding

Table A5: Total employment multiplier matrix for **G***

	A1	**A2**	**A3**	**A4**	**B1**	**B2**	**B3**	**B4**	**C1**	**C2**	**C3**	**C4**
A1	0.14	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.01	0.00
A2	0.00	0.07	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
A3	0.00	0.00	0.05	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
A4	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B1	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.01	0.00
B2	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00
B3	0.00	0.01	0.00	0.00	0.00	0.01	0.06	0.01	0.01	0.01	0.00	0.00
B4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00
C1	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.00	0.18	0.00	0.01	0.01
C2	0.01	0.02	0.00	0.01	0.02	0.02	0.01	0.01	0.01	0.16	0.01	0.00
C3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00
C4	0.03	0.03	0.01	0.02	0.04	0.04	0.03	0.02	0.03	0.04	0.04	0.33
Total	0.19	0.15	0.07	0.10	0.19	0.18	0.12	0.13	0.25	0.22	0.28	0.36

Note: Summation may not equal due to rounding