A Method to Concatenate Multi-regional Input-ouput Tables to Inter-regional Input-output Table with Processing Trade*

Ye Sun†
Kunfu Zhu‡

Abstract

This article introduces a quadratic-program-based method to concatenate multi-regional input-output table to an inter-regional input-output table with processing trade displayed separately. Additional information from the DPN table and customs data is used. This method is demonstrated by using China in 2012 as an example.

*This paper is originally the data manual of IOP table excluding the introduction section and the empirical study part. This extended version is specifically rearranged as writing sample. For data and R/MATLAB program inquiry please contact 201519040@uibe.edu.cn

†University of International Business and Economics, Beijing 100029, CHINA, 201519040@uibe.edu.cn

‡Research Institute on Global Value Chains, University of International Business and Economics, Beijing 100029, CHINA, kfzhu@uibe.edu.cn

1 Introduction

China's export showed drastic rise during the past two decades. According to the China Customs Statistics (CCS) Information Center, China's total export rises from 0.15 trillion US dollars in 1995 to 2.35 trillion US dollars in 2014. This *China Shock* phenomenon has been a heated topic for economists around the world. Some attributes a number of problems to *China Shock*, say domestic unemployment in United States (Autor et al., 2016). Some justify the development of China. One of the major disagreements is the accounting method: When using value-added to measure the status quo of China's trade relationship with other countries (Wang et al., 2017), arguments that are based on the trade deficit are invalidated (Wang et al., 2018). This argument is valid especially when global cooperation on production intensifies: according to CCS, out of China's total export, around 40% are intermediate goods. Also, intermediate goods take up as much as 70% China's import. China's active participation in Global Value Chains (GVC) makes it inadequate to merely view trade data without disaggregation. One vital tool to scrutinize trade pattern is the Input-Output table, which Leontief (1936) first uses to develop one of his major contributions: input-output (IO) analysis. Several complementary IO analysis is developed on this basis. IO table with finer sectors, or regional divisions, or both (say WIOT), are being used in computerized economic model like Computable General Equilibrium (CGE), GVC research, and IO analysis.

One reason that China have relatively low value-added ratio lies in the customs regime. For approximately half of its export, Chinese enterprises do not actually control the property of the foreign supplier. This portion of trade is referred to as processing trade. According to CCS, there are three types of trade regimes which are

PA *Process and assembling*. It refers to the type of inward processing in which foreign suppliers provide raw materials, parts or components under a contractual arrangement for the subsequent re-exportation of the processed products. Under this type of transaction, the imported inputs and the finished outputs remain the property of the foreign supplier.

PI *Process with imported materials*. It refers to the type of inward processing other than *Processing and Assembling*, in which raw materials or components are imported for the manufacture of the export-oriented products.

OT *Ordinary trade*. Any other trade other than inward processing. It is also frequently referred to as *non-processing trade*.

Processing trade tend to have low value-added ratio. Because the job entailed, usually assembling, does not involve high-skilled labor, which results in the high elasticity of demand curve in the labor market. Moreover, given its strong dependence on imports, processing exports induce less domestic economic activity. Known as the World Factory, China have almost half of the export as processing trade (See Figure 1).

The behavior of the processing export part differs a lot from the ordinary trade. It is meaningful and necessary to disaggregate gross trade figures into non-processing trade and processing trade. In previous researches, economists discovered relevance of processing trade to a number of adjacent indicators. For instance, the percentage of domestic content in China's exports rise from 50.7% to 73.7% for the year 2006 if processing trade is excluded (Koopman et al., 2008).

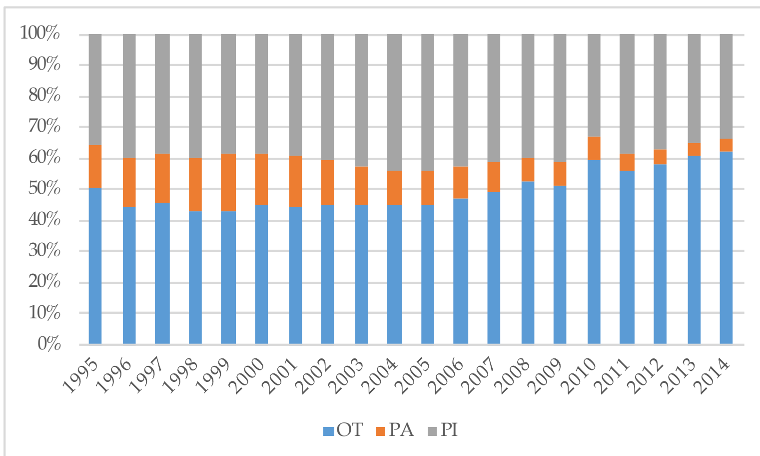

Figure 1: The Proportion of different Customs Regimes

Of all the variations of IO table, DPN table (Lau et al., 2007) is proposed to characterize both processing exports and non-processing exports of China. Unlike a classical IO table which scrutinizes the flow of value among sectors, DPN table adopts a similar layout while its focus is the flow of value among sector-regime pairs. The name, *DPN table*, is straightforwardly taken after the 3 regimes: production for domestic use(D), production for processing export(P), and production for non-processing exports and other production of foreign invested enterprises(N).

In the making of a classical IO table or IO-based extensive model, it is inevitable to reconcile and integrate data from different sources which normally would not agree with each other on the exact number. In order to maintain consistency and comparability, quadratic programing (QP) are introduced to reconcile the nuance among sources of data. The philosophy of this methodology is basically (i) constraining the endogenous variables to be consistent with the observable IO data, (ii) constraining the endogenous variables to be consistent within the system, and (iii) loosing proportionality assumption and exploiting the remaining degrees of freedom by allowing the researcher to incorporate additional information. In previous literature, this QP framework is conducted with China's intra-national regional IO table and customs data of every province. China's entries in the WIOT are disaggregated into 5 different regions (Meng et al., 2013), which enables us to see how GVC is extended within China on a regional level. The presumption that validates this methodology is the proportionality assumptions. Alonso de Gortari (2017) discussed in-depth about this QP method: Although this method cannot guarantee that the solution to the QP framework is exact the real situation, but it can still improve upon I-O analysis as the GVC estimates become more accurate.

In this paper, with the aid of the QP framework, we use the DPN table and customs data to expand China's intra-national regional IO table into an intra-national regional IOP table, which stands for *IO table with processing trade distinguished*.

2 Sources of Data

2.1 Multi-regional Input-output Table

This series of tables, hereafter referred to as MRIO for convenience, is compiled by Development Research Center of the State Council. The table consists of 31 provincial IO tables[1]. Each IO table comes with 2 inter-provincial matrices (import and export). The basic layout of the IO table of each province is shown below.

[1] Taiwan, Hong Kong and Macao are excluded.

<table>
<tr><th></th><th>S01</th><th>S02</th><th>$\cdots$</th><th>S42</th><th>FU101</th><th>$\cdots$</th><th>FU202</th><th>EX</th><th>PEX</th><th>IM</th><th>PIM</th><th>ERR</th><th>total</th></tr>
<tr><td>S01</td><td colspan="4" rowspan="4">$\mathbf{Z}$</td><td colspan="3" rowspan="4">$\mathbf{F}$</td><td rowspan="4">$\mathbf{E}$</td><td rowspan="4">$\mathbf{S}$</td><td rowspan="4">$\mathbf{M}$</td><td rowspan="4">$\mathbf{Q}$</td><td rowspan="4">$\mathbf{R}$</td><td rowspan="4">$\mathbf{X}$</td></tr>
<tr><td>S02</td></tr>
<tr><td>$\vdots$</td></tr>
<tr><td>S42</td></tr>
<tr><td>VA001</td><td colspan="4" rowspan="3">$\mathbf{V}$</td></tr>
<tr><td>$\vdots$</td></tr>
<tr><td>VA004</td></tr>
<tr><td>total</td><td colspan="4">$\mathbf{X}$</td></tr>
</table>

Table 1: Layout of MRIO 1

The table is reconciled. Let $\mathbf{u} = (1, 1, \cdots, 1)^T$ (42 ones), $\mathbf{v} = (1, 1, \cdots, 1)^T$ (31 ones), and $\mathbf{w} = (1, 1, 1, 1, 1)^T$. The reconciliation condition can be expressed as below

$$\mathbf{Z} \cdot \mathbf{u} + \mathbf{F} \cdot \mathbf{w} + \mathbf{E} + \mathbf{S} - \mathbf{M} - \mathbf{Q} + \mathbf{R} = \mathbf{X} \tag{1}$$

$$\mathbf{u}^T \mathbf{Z} + (1, 1, 1, 1) \cdot \mathbf{V} = \mathbf{X}^T \tag{2}$$

Attached to each provincial IO table, the inter-provincial import matrix specifies the amount of value that flows in from all 31 provinces. The basic layout is shown below:

<table>
<tr><th></th><th>S01</th><th>S02</th><th>$\cdots$</th><th>S42</th></tr>
<tr><td>Beijing</td><td colspan="4" rowspan="4">$\mathbf{T}$</td></tr>
<tr><td>Tianjin</td></tr>
<tr><td>$\vdots$</td></tr>
<tr><td>Xinjiang</td></tr>
</table>

Table 2: Layout of MRIO 2

By definition,

$$\mathbf{T}^T \mathbf{v} = \mathbf{Q} \tag{3}$$

2.2 DPN table

		INT.USE			FU	X
		D	P	N		
INT.INPUT	D	$\mathbf{Z}^{DD}$	$\mathbf{Z}^{DP}$	$\mathbf{Z}^{DN}$	$\mathbf{F}^{D}$	$\mathbf{X}^{D}$
	P	$\mathbf{Z}^{PD}$	$\mathbf{Z}^{PP}$	$\mathbf{Z}^{PN}$	$\mathbf{F}^{P}$	$\mathbf{X}^{P}$
	N	$\mathbf{Z}^{ND}$	$\mathbf{Z}^{NP}$	$\mathbf{Z}^{NN}$	$\mathbf{F}^{N}$	$\mathbf{X}^{N}$
M		$\mathbf{Z}^{MD}$	$\mathbf{Z}^{MP}$	$\mathbf{Z}^{MN}$	$\mathbf{F}^{M}$	$\mathbf{X}^{M}$
VA		$\mathbf{V}^{D}$	$\mathbf{V}^{P}$	$\mathbf{V}^{N}$		
X		$(\mathbf{X}^{D})^{T}$	$(\mathbf{X}^{P})^{T}$	$(\mathbf{X}^{N})^{T}$		

Table 3: Layout of DPN table

The basic layout of DPN table is shown above in Table 3, where $\mathbf{Z}^{AB}$ is a square matrix whose length is the total number of sectors.[2] In the example of China, it is a 139×139 square matrix. In order to put it into use, we convert the 139-sector table to the 42-sector table using the concordance that comes together with the MRIO table.

This table provides detailed information about how imported goods of all sectors is absorbed in different sector-regime pairs (highlighted in light yellow in Table 3). $\mathbf{Z}^{MP}$ is import-to-processing matrix. By adding up $\mathbf{Z}^{MD}$ and $\mathbf{Z}^{MN}$, we can derive the import-to-nonprocessing matrix, which together with import-to-processing matrix, is the crucial ingredient of the compilation of final table.

2.3 Data from China Customs Statistics (CCS), 2012

This very database is acquired through Hong Kong-based China Customs Statistics (CCS) Information Center. The raw data comes with 12 variables, ranging from customs districts to value in USD of the year. We concentrate mainly on 6 relevant variables in the raw data table:

1. Locations of Chinese traders[3]

2. Customs regime[4]

[2] $A \in \{D, P, N, M\}; B \in \{D, P, N\}$.

[3] Grouped by provincial administrative division.

[4] Customs regime totals 12 regimes. Mark 2 categories, *process and assembling* and *process with imported materials*, as *processing trade*(P), and other 10 categories as *normal trade*(N).

3. Countries of departure/destination

4. Countries of origin/consumption

5. HS codes (8-digit)

6. Values of the year (US$)

Using concordance table published by World Integrated Trade Solution (WITS), we map the 8-digit HS codes to BEC classification[5], which breaks down each transaction in raw data into 3 parts: household consumption goods, capital goods and intermediate goods (abbreviated to CONS, CAP and INT, respectively). Similarly, we map the 8-digit HS codes to 42 main sectors according to the concordance table attached to the MRIO table 2012. 4 new variables is therefore joined on to the data frame: sector (adding up to 42 sectors, all coded from 01 to 42), value of CONS, value of INT and value of CAP.

3 Initial Values for Endogenous Variables

We group the import data of 2012 by sector, locations of Chinese traders, and customs regime, aggregate respectively value of CONS, value of INT and value of CAP.

sector	L/ct	cr	v.CONS	v.INT	v.CAP
01	Beijing	N			
01	Beijing	P			
⋮	⋮	⋮			

Table 4: Layout 1 of import data frame

Aggregate the 3 variables: value of CONS, value of INT and value of CAP when the regime is *processing trade*, and transpose it alongside the original value of CONS, value of INT and value of CAP. Rename the new variable to P.

[5]see also in https://unstats.un.org/unsd/tradekb/Knowledgebase/50089/Classification-by-Broad-Economic-Categories-Rev4

sector	L/ct	v.CONS	v.INT	v.CAP	P
01	Beijing				
02	Beijing				
⋮	⋮				

Table 5: Layout 2 of import data frame

Use the proportion of these 4 variables, we can decompose **M** of each province into 4 vectors $CONS$, CAP, INT, P. Note that both intermediate import and processing import go completely into intermediate production and nothing else. Decompose P according to the import-to-processing matrix extracted from 42-sector DPN table, hence unfolding the vector into a square matrix **P**. Similarly, decompose vector INT according to the import-to-nonprocessing matrix, unfolding the vector into another square matrix **N**. By adding up **N** and **P**, we have a new matrix $\mathbf{Z}^M$ which later become one of the jigsaw pieces in the import matrix.

Exercise a similar method to vector $CONS$, however, according to matrix **F** of each province instead: decompose $CONS$ using the proportion of the first 3 column of **F**(denoted as **C**), and expand it into a 42×3 matrix, namely $\mathbf{C}^M$. By far we have decompose vector **M** into a 42×46 initial import matrix.

$$[\mathbf{Z}^M, \mathbf{C}^M, CAP]$$

Since each element in $[\mathbf{Z}, \mathbf{C}, FU201]$[6] is, theoretically speaking, aggregated from goods of import origin, inter-provincial origin and inner-provincial origin, and none of these elements should have a negative value, certain adjustment is needed to prevent violation of the non-negative constraint. We compare $[\mathbf{Z}^M, \mathbf{C}^M, CAP]$ and $[\mathbf{Z}, \mathbf{C}, FU201]$, then substitute element in the former with the counterpart in the latter matrix if the former is larger than the latter, and the discrepancy caused by this move is aggregated every row and attached to the matrix as a new column indicating change of inventories. By the end of this adjustment, the final import matrix $[\mathbf{Z}^M, \mathbf{C}^M, CAP, INV]$ is derived. To put it simply, $[\mathbf{Z}^M, \mathbf{F}^M]$ is derived, given that $\mathbf{F}^M = [\mathbf{C}^M, CAP, INV]$.

By deducting $[\mathbf{Z}^M, \mathbf{F}^M]$ from $[\mathbf{Z}, \mathbf{F}]$, we eliminate the use of imported goods in **Z** and **F** in the MRIO table. What remains purely results from domestic (both inner- and inter- provincial) production.

$$\mathbf{Z}^D = \mathbf{Z} - \mathbf{Z}^M \tag{4}$$

$$\mathbf{F}^D = \mathbf{F} - \mathbf{F}^M \tag{5}$$

[6] $FU201$ is the variable that denotes fixed capital formation, directly extracted from the MRIO table.

The domestic matrix $[\mathbf{Z}^D, \mathbf{C}^D, CAP^D]$ can be further decomposed into inner-provincial matrix and inter-provincial counterpart, of which the inner-provincial matrix $[\mathbf{Z}^\star, \mathbf{C}^\star, CAP^\star]$[7] should also be subject to non-negative constraint, which means the column sum of $[\mathbf{Z}^D, \mathbf{F}^D]$ should always be larger than $\mathbf{Q}$ for every element.

$$\mathbf{Z}^D \cdot \mathbf{u} + \mathbf{F}^D \cdot \mathbf{w} \geqslant \mathbf{Q} \tag{6}$$

If there is any violation of condition (6), deduct the difference from $INV^\star$, which can be derived later.

Assuming that domestically within sectors and between provinces, the usage of goods is not subject to substantial difference, we can thus decompose $\mathbf{T}$ using domestic matrix $[\mathbf{Z}^D, \mathbf{F}^D]$:

$$T_i^{rs} \frac{ZD_{ij}^s}{\sum_j ZD_{ij}^s + \sum_f FD_{if}^s} = \hat{Z}_{ij}^{rs} \tag{7}$$

$$T_i^{rs} \frac{FD_{if}^s}{\sum_j ZD_{ij}^s + \sum_f FD_{if}^s} = \hat{F}_{if}^{rs} \tag{8}$$

Once the decomposition is done, we can combine all these block matrices. The layout of the combination is shown in below.

	Beijing	Tianjin	$\cdots$	Xinjiang	Beijing	Tianjin	$\cdots$	Xinjiang
Beijing	$\mathbf{0}$	$\hat{\mathbf{Z}}^{1,2}$	$\cdots$	$\hat{\mathbf{Z}}^{1,31}$	$\mathbf{0}$	$\hat{\mathbf{F}}^{1,2}$	$\cdots$	$\hat{\mathbf{F}}^{1,31}$
Tianjin	$\hat{\mathbf{Z}}^{2,1}$	$\mathbf{0}$	$\cdots$	$\hat{\mathbf{Z}}^{2,31}$	$\hat{\mathbf{F}}^{2,1}$	$\mathbf{0}$	$\cdots$	$\hat{\mathbf{F}}^{2,31}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
Xinjiang	$\hat{\mathbf{Z}}^{31,1}$	$\hat{\mathbf{Z}}^{31,2}$	$\cdots$	$\mathbf{0}$	$\hat{\mathbf{F}}^{31,1}$	$\hat{\mathbf{F}}^{31,2}$	$\cdots$	$\mathbf{0}$

Given that the domestic matrix $[\mathbf{Z}^D, \mathbf{C}^D, CAP^D]$ can be further decomposed into inner-provincial matrix and inter-provincial counterpart, and that inter-provincial import matrix is derived. We can derive the inner-provincial matrix from deducting all inter-provincial matrices from domestic matrix.

$$\mathbf{Z}^{\star s} = \mathbf{Z}^D - \sum_r \hat{\mathbf{Z}}^{rs} \tag{9}$$

$$\mathbf{F}^{\star s} = \mathbf{F}^D - \sum_r \hat{\mathbf{F}}^{rs} \tag{10}$$

[7] Change of inventories is excluded now and will be added back on later.

If there is a violation of condition (6) for some province-sector pair, the elements of the corresponding row of $\mathbf{Z}^\star$ is mainly negative, which does not satisfy the non-negative constraints. We aggregate all the elements of the negative row and put it in $INV^\star$, and then zeros all the negative value in $\mathbf{Z}^\star$, $\mathbf{C}^\star$ and $CAP^\star$. Through this adjustment, a non-negative inner-provincial matrix [$\mathbf{Z}^\star$, $\mathbf{F}^\star$] is derived. Concatenate all these matrices as shown below.

	R1	R2	$\cdots$	R31	R1	R2	$\cdots$	R31	EX	ERR	X
R1	$\mathbf{Z}^{\star 1}$	$\hat{\mathbf{Z}}^{1,2}$	$\cdots$	$\hat{\mathbf{Z}}^{1,31}$	$\mathbf{F}^{\star 1}$	$\hat{\mathbf{F}}^{1,2}$	$\cdots$	$\hat{\mathbf{F}}^{1,31}$	$\mathbf{E}^1$	$\mathbf{R}^1$	$\mathbf{X}^1$
R2	$\hat{\mathbf{Z}}^{2,1}$	$\mathbf{Z}^{\star 2}$	$\cdots$	$\hat{\mathbf{Z}}^{2,31}$	$\hat{\mathbf{F}}^{2,1}$	$\mathbf{F}^{\star 2}$	$\cdots$	$\hat{\mathbf{F}}^{2,31}$	$\mathbf{E}^2$	$\mathbf{R}^2$	$\mathbf{X}^2$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
R31	$\hat{\mathbf{Z}}^{31,1}$	$\hat{\mathbf{Z}}^{31,2}$	$\cdots$	$\mathbf{Z}^{\star 31}$	$\hat{\mathbf{F}}^{31,1}$	$\hat{\mathbf{F}}^{31,2}$	$\cdots$	$\mathbf{F}^{\star 31}$	$\mathbf{E}^{31}$	$\mathbf{R}^{31}$	$\mathbf{X}^{31}$
IM	$\mathbf{Z}^{M,1}$	$\mathbf{Z}^{M,2}$	$\cdots$	$\mathbf{Z}^{M,31}$	$\mathbf{F}^{M,1}$	$\mathbf{F}^{M,2}$	$\cdots$	$\mathbf{F}^{M,31}$			
V	$\mathbf{V}^1$	$\mathbf{V}^2$	$\cdots$	$\mathbf{V}^{31}$							
X	$(\mathbf{X}^1)^T$	$(\mathbf{X}^2)^T$	$\cdots$	$(\mathbf{X}^{31})^T$							

Table 6: Layout of IO table

This table is automatically reconciled. Check if

$$\mathbf{u}^T \cdot \left(\mathbf{Z}^{\star s} + \sum_{r \neq s} \hat{\mathbf{Z}}^{rs} + \mathbf{Z}^{Ms} \right) + (1,1,1,1) \cdot \mathbf{V}^s = (\mathbf{X}^s)^T, \forall s \tag{11}$$

$$\left(\mathbf{Z}^{\star r} + \sum_{s \neq r} \hat{\mathbf{Z}}^{rs} \right) \cdot \mathbf{u} + \left(\mathbf{F}^{\star r} + \sum_{s \neq r} \hat{\mathbf{F}}^{rs} \right) \cdot \mathbf{w} + \mathbf{E}^r + \mathbf{R}^r = \mathbf{X}^r, \forall r \tag{12}$$

4 Reconciliation

Due to forcing element to be non-negative,

$$\mathbf{Z}^{\star s} + \sum_{r \neq s} \hat{\mathbf{Z}}^{rs} + \mathbf{Z}^{Ms} = \mathbf{Z}^s \tag{13}$$

$$\mathbf{F}^{\star s} + \sum_{r \neq s} \hat{\mathbf{F}}^{rs} + \mathbf{Z}^{Ms} = \mathbf{F}^s \tag{14}$$

equation (13) and equation (14) are invalidated. However, as a complete reconciled IO table, this should be satisfied. Therefore, a quadratic program is introduced to find a reconciled solution.

Let the orange part of Table 6 be $\hat{\mathbf{Z}}$, elements of the matrix being called $\hat{Z}_{ij}^{rs}$. Let the green section of Table 6 be $\hat{\mathbf{F}}$, elements of the matrix being called $\hat{F}_{ij}^{rs}$. Let the purple part of Table 6 be $\hat{\mathbf{Z}}^{M}$, elements of the matrix being called $\hat{Z}_{ij}^{Ms}$, and the pink part be $\hat{\mathbf{F}}^{M}$, element being called $\hat{F}_{ij}^{Ms}$.

$$\min_{{}^*Z_{ij}^{rs},{}^*Z_{ij}^{Ms},{}^*F_{if}^{rs},{}^*F_{if}^{Ms}} \quad \sum_i\sum_j\sum_r\sum_s \frac{({}^*Z_{ij}^{rs}-\hat{Z}_{ij}^{rs})^2}{\alpha_{ij}^{rs}} + \sum_i\sum_j\sum_r\sum_s \frac{({}^*Z_{ij}^{Ms}-\hat{Z}_{ij}^{Ms})^2}{\alpha_{ij}^{Ms}}$$
$$+\sum_i\sum_f\sum_r\sum_s \frac{({}^*F_{if}^{rs}-\hat{F}_{if}^{rs})^2}{\beta_{if}^{rs}} + \sum_i\sum_f\sum_r\sum_s \frac{({}^*F_{if}^{Ms}-\hat{F}_{if}^{Ms})^2}{\beta_{if}^{Ms}}$$

$$\text{s.t.} \quad \sum_r {}^*Z_{ij}^{rs} + {}^*Z_{ij}^{Ms} = Z_{ij}^{s}, \forall s, \forall i, \forall j \tag{15}$$
$$\sum_r {}^*F_{if}^{rs} + {}^*F_{if}^{Ms} = F_{if}^{s}, \forall s, \forall i, \forall f \tag{16}$$
$$\sum_j {}^*Z_{ij}^{rs} + \sum_f {}^*F_{if}^{rs} = T_i^{rs}, \forall s, \forall r \neq s, \forall i \tag{17}$$
$$\sum_j {}^*Z_{ij}^{Ms} + \sum_f {}^*F_{if}^{Ms} = M_i^{s}, \forall s, \forall i \tag{18}$$
$$\sum_r\sum_i {}^*Z_{ij}^{rs} + \sum_i {}^*Z_{ij}^{Ms} + \sum_v V_{vj}^{s} = X_j^{s}, \forall j, \forall s \tag{19}$$
$$\sum_s\sum_j {}^*Z_{ij}^{rs} + \sum_s\sum_f {}^*F_{if}^{rs} + E_i^r + R_i^r = X_i^{r}, \forall i, \forall r \tag{20}$$
$${}^*Z_{ij}^{rs}, {}^*Z_{ij}^{Ms} \geqslant 0, \forall i, \forall j, \forall r, \forall s \tag{21}$$
$${}^*F_{if}^{rs}, {}^*F_{if}^{Ms} \geqslant 0, \forall i, \forall f \leqslant 4, \forall r, \forall s \tag{22}$$

where $\alpha_{ij}^{rs}, \alpha_{ij}^{Ms}, \beta_{ij}^{rs}, \beta_{ij}^{Ms}$ are weights that determines the flexibility of variables in the program. Normally speaking,

$$\alpha_{ij}^{rs} = |\hat{Z}_{ij}^{rs}| \tag{23}$$
$$\alpha_{ij}^{Ms} = |\hat{Z}_{ij}^{Ms}| \tag{24}$$
$$\beta_{if}^{rs} = |\hat{F}_{if}^{rs}| \tag{25}$$
$$\beta_{if}^{Ms} = |\hat{F}_{if}^{Ms}| \tag{26}$$

when RHS of these equations do not equal to 0.

Situations when the RHS equals to 0 can result from various possibilities:

1. corresponding part in $\mathbf{Z}$ equals to 0
2. corresponding part in $\mathbf{Q}$ or $\mathbf{M}$ equals to 0
3. generated by imposing non-negative constraints during deducting matrices from matrices.

If 0 is generated due to the first 2 reasons, it should remain 0 through the QP. On the contrary, when 0 is generated by processing the matrices, we expect it to vary. So when the RHS is 0, we can set the weight to a relatively small number. In this way, all of the 0 values are open to variation during QP, but not too much. Since 0 generated by reason 1 and reason 2 is constrained by the constraints in QP, it manages to stay 0 somehow.

After the QP, replace

$$\begin{bmatrix} \hat{\mathbf{Z}} & \hat{\mathbf{F}} \\ \hat{\mathbf{Z}}^M & \hat{\mathbf{F}}^M \end{bmatrix}$$

with

$$\begin{bmatrix} {}^*\mathbf{Z} & {}^*\mathbf{F} \\ {}^*\mathbf{Z}^M & {}^*\mathbf{F}^M \end{bmatrix}$$

in the IO table. The compilation of the IO table is completed.

5 IOP Table

Following the compilation of the IO table, we now distinguish the processing trade part embedded within the the IO table. Our purpose is to create a IO table with processing trade distinguished (IOP) whose layout is shown in Table 7.

This IOP table (Table 7) is different from a normal interregional IO table (Table 6) in that the column name have a new dimension: N for non-processing trade and P for processing trade, indexing between the regional dimension and the sectorial dimension. In this way, the domestic intermediate input, import intermediate input, value added, gross output are all separated into 2 part: N section and P section. Moreover, we separate processing export and non-processing export apart, rather than display the aggregated export.

5.1 Initial Values for IOP's Endogenous Variables

To begin with, we need to extract the processing export out of the aggregated export data, using the export data from CCS Information Center. Apply the same

<table>
<tr><td colspan="2" rowspan="3"></td><td colspan="8">Region1</td><td>⋯</td><td colspan="8"></td></tr>
<tr><td colspan="4">N</td><td colspan="4">P</td><td>⋯</td><td colspan="3">Region1</td><td>⋯</td><td colspan="4"></td></tr>
<tr><td>S01</td><td>S02</td><td>⋯</td><td>S42</td><td>S01</td><td>S02</td><td>⋯</td><td>S42</td><td>⋯</td><td>F101</td><td>⋯</td><td>F202</td><td>⋯</td><td>EX_N</td><td>EX_P</td><td>ERR</td><td>X</td></tr>
<tr><td rowspan="4">Region1</td><td>S01</td><td colspan="9" rowspan="5">$^{N}Z_{ij}^{rs}$ and $^{P}Z_{ij}^{rs}$</td><td colspan="4" rowspan="5">F_{if}^{rs}</td><td rowspan="5">$^{N}E_{i}^{r}$</td><td rowspan="5">$^{P}E_{i}^{r}$</td><td rowspan="5">R_{i}^{r}</td><td rowspan="5">X_{i}^{r}</td></tr>
<tr><td>S02</td></tr>
<tr><td>⋮</td></tr>
<tr><td>S42</td></tr>
<tr><td>⋮</td><td>⋮</td></tr>
<tr><td rowspan="4">Import</td><td>S01</td><td colspan="9" rowspan="4">$^{N}Z_{ij}^{Ms}$ and $^{P}Z_{ij}^{Ms}$</td><td colspan="4" rowspan="4">F_{if}^{Ms}</td></tr>
<tr><td>S02</td></tr>
<tr><td>⋮</td></tr>
<tr><td>S42</td></tr>
<tr><td rowspan="4">Value Added</td><td>V001</td><td colspan="9" rowspan="4">$^{N}V_{vj}^{s}$ and $^{P}V_{vj}^{s}$</td></tr>
<tr><td>V002</td></tr>
<tr><td>V003</td></tr>
<tr><td>V004</td></tr>
<tr><td colspan="2">X</td><td colspan="9">$^{P}X_{j}^{s}$ and $^{N}X_{j}^{s}$</td></tr>
</table>

Table 7: Layout of IOP

method to export data as is applied to process import data (Table 5). We expect the exact same layout of the export data frame. Use the proportion of P, we can decompose $\mathbf{E}$ of each province ($\mathbf{E}^r$) into ${}^P\hat{\mathbf{E}}^r$ and ${}^N\hat{\mathbf{E}}^r$. Note that the processing input will not be contributing to any other sectors production, but goes directly to processing export. Therefore,

$$
\begin{aligned}
{}^P\hat{\mathbf{X}}^r &= {}^P\hat{\mathbf{E}}^r && (27)\\
{}^N\hat{\mathbf{X}}^r &= \mathbf{X}^r - {}^P\hat{\mathbf{X}}^r && (28)
\end{aligned}
$$

Suppose that the allocation of gross input is similar and proportionate across trade regime (N and P) within region-sector pair and within trade regime across region-sector pair. We give our estimation based on both trade regime data based on DPN table and region-sector pair data based on the IO table that we previously compiled.

However, the DPN table displays separately the production for domestic use(D) production for non-processing exports and other production of foreign invested enterprises(N), while we mainly focus on production for processing trade(P) and all other types of production(both D and N). Hence, we define that:

$$
\begin{aligned}
\mathbf{Z}^{NN} &:= \mathbf{Z}^{DD} + \mathbf{Z}^{DN} + \mathbf{Z}^{ND} + \mathbf{Z}^{NN} && (29)\\
\mathbf{Z}^{NP} &:= \mathbf{Z}^{DP} + \mathbf{Z}^{NP} && (30)\\
\mathbf{Z}^{PN} &:= \mathbf{Z}^{PD} + \mathbf{Z}^{PN} && (31)\\
\mathbf{V}^{N} &:= \mathbf{V}^{D} + \mathbf{V}^{N} && (32)\\
\mathbf{X}^{N} &:= \mathbf{X}^{D} + \mathbf{X}^{N} && (33)
\end{aligned}
$$

With rearranged DPN table, we embark on giving our estimation to the initial IOP table[8].

$$
\begin{aligned}
\mathbf{Z}^{all} &= \mathbf{Z}^{NP} + \mathbf{Z}^{NN} && (34)\\
\mathbf{V}^{all} &= \mathbf{V}^{N} + \mathbf{V}^{P} && (35)\\
\mathbf{X}^{all} &= \mathbf{X}^{N} + \mathbf{X}^{P} && (36)\\
{}^P\hat{\mathbf{Z}}^{rs} &= \left[{}^*\mathbf{Z}^{rs}\mathrm{diag}^{-1}(\mathbf{X}^s) \circ \mathbf{Z}^{NP}\mathrm{diag}^{-1}(\mathbf{X}^P) \circ \mathbf{Z}^{all}\mathrm{diag}^{-1}(\mathbf{X}^{all})^{\circ -1}\right] \mathrm{diag}({}^P\hat{\mathbf{X}}^s) && (37)\\
{}^N\hat{\mathbf{Z}}^{rs} &= \left[{}^*\mathbf{Z}^{rs}\mathrm{diag}^{-1}(\mathbf{X}^s) \circ \mathbf{Z}^{NN}\mathrm{diag}^{-1}(\mathbf{X}^N) \circ \mathbf{Z}^{all}\mathrm{diag}^{-1}(\mathbf{X}^{all})^{\circ -1}\right] \mathrm{diag}({}^N\hat{\mathbf{X}}^s) && (38)\\
{}^P\hat{\mathbf{V}}^{s} &= \left[{}^*\mathbf{V}^{s}\mathrm{diag}^{-1}(\mathbf{X}^s) \circ \mathbf{V}^{P}\mathrm{diag}^{-1}(\mathbf{X}^P) \circ \mathbf{V}^{all}\mathrm{diag}^{-1}(\mathbf{X}^{all})^{\circ -1}\right] \mathrm{diag}({}^P\hat{\mathbf{X}}^s) && (39)\\
{}^N\hat{\mathbf{V}}^{s} &= \left[{}^*\mathbf{V}^{s}\mathrm{diag}^{-1}(\mathbf{X}^s) \circ \mathbf{V}^{N}\mathrm{diag}^{-1}(\mathbf{X}^N) \circ \mathbf{V}^{all}\mathrm{diag}^{-1}(\mathbf{X}^{all})^{\circ -1}\right] \mathrm{diag}({}^N\hat{\mathbf{X}}^s) && (40)
\end{aligned}
$$

[8]The notation $\circ$ is to calculate *Hadamard product*, or *entrywise product*, of two matrices with the same size. In this article, *matrix product* is granted a higher precedence than *Hadamard product* and other analogous operations like *Hadamard power*.

Our estimation for ${}^{P}\mathbf{Z}^{Ms}$ and ${}^{N}\mathbf{Z}^{Ms}$ (${}^{P}\hat{\mathbf{Z}}^{Ms}$ and ${}^{N}\hat{\mathbf{Z}}^{Ms}$) is each province's matrix **N** and matrix **P** derived from earlier in the making of import matrix.

5.2 Re-reconcilation

The balance is broken through last operation. We again introduce a quadratic program to find a reconciled solution for the IOP table.

$$\begin{aligned}
\min_{{}^{\delta}\mathbf{Z}^{rs},{}^{\delta}\mathbf{Z}^{Ms},{}^{\delta}\mathbf{V}^{s},{}^{\delta}\mathbf{X}^{r},{}^{\delta}\mathbf{E}^{r},\delta\in\{N,P\}} \quad & \sum_{\delta}\sum_{r}\sum_{s}\|({}^{\delta}\mathbf{Z}^{rs}-{}^{\delta}\hat{\mathbf{Z}}^{rs})\circ{}^{\delta}\mathbf{\Gamma}^{rs}\|_F^2 \\
&+\sum_{\delta}\sum_{s}\|({}^{\delta}\mathbf{Z}^{Ms}-{}^{\delta}\hat{\mathbf{Z}}^{Ms})\circ{}^{\delta}\mathbf{\Delta}^{s}\|_F^2 \\
&+\sum_{\delta}\sum_{s}\|({}^{\delta}\mathbf{V}^{s}-{}^{\delta}\hat{\mathbf{V}}^{s})\circ{}^{\delta}\mathbf{\Theta}^{s}\|_F^2 \\
&+\sum_{\delta}\sum_{r}\|({}^{\delta}\mathbf{X}^{r}-{}^{\delta}\hat{\mathbf{X}}^{r})\circ{}^{\delta}\mathbf{\Lambda}^{r}\|_F^2 \\
&+\sum_{\delta}\sum_{r}\|({}^{\delta}\mathbf{E}^{r}-{}^{\delta}\hat{\mathbf{E}}^{r})\circ{}^{\delta}\mathbf{\Xi}^{r}\|_F^2
\end{aligned}$$

$$\begin{aligned}
\text{s.t.} \quad & {}^{P}\mathbf{Z}^{rs}+{}^{N}\mathbf{Z}^{rs}={}^{*}\mathbf{Z}^{rs},\forall r,\forall s & (41)\\
& {}^{P}\mathbf{Z}^{Ms}+{}^{N}\mathbf{Z}^{Ms}={}^{*}\mathbf{Z}^{Ms},\forall s & (42)\\
& {}^{P}\mathbf{V}^{s}+{}^{N}\mathbf{V}^{s}=\mathbf{V}^{s},\forall s & (43)\\
& {}^{P}\mathbf{X}^{r}+{}^{N}\mathbf{X}^{r}=\mathbf{X}^{r},\forall r & (44)\\
& {}^{P}\mathbf{E}^{r}+{}^{N}\mathbf{X}^{r}=\mathbf{E}^{r},\forall r & (45)\\
& {}^{P}\mathbf{X}^{r}={}^{P}\mathbf{E}^{r},\forall r & (46)\\
& \sum_{i}\sum_{r}{}^{P}Z_{ij}^{rs}+\sum_{i}{}^{P}Z_{ij}^{Ms}+\sum_{v}{}^{P}V_{vj}^{s}={}^{P}X_{j}^{s} & (47)\\
& \sum_{i}\sum_{r}{}^{N}Z_{ij}^{rs}+\sum_{i}{}^{N}Z_{ij}^{Ms}+\sum_{v}{}^{N}V_{vj}^{s}={}^{N}X_{j}^{s} & (48)\\
& \text{All cells must be no less than 0, except for V002 and V004.} & (49)
\end{aligned}$$

where ${}^{\delta}\mathbf{\Gamma}^{rs}$, ${}^{\delta}\mathbf{\Delta}^{rs}$, ${}^{\delta}\mathbf{\Theta}^{rs}$, ${}^{\delta}\mathbf{\Lambda}^{rs}$ and ${}^{\delta}\mathbf{\Xi}^{rs}$ are all weight matrix with the same size of the corresponding original matrix, respectively. It controls the range of varying of a variable in the QP process. In the making of our database, we generally adopt that for all r ,s and δ:

$$ {}^{\delta}\mathbf{\Gamma}^{rs} = (\Gamma\times{}^{\delta}\hat{\mathbf{Z}}^{rs})^{\circ\frac{1}{2}} \tag{50}$$

$$
\begin{aligned}
{}^{\delta}\boldsymbol{\Delta}^{s} &= (\Delta \times {}^{\delta}\hat{\mathbf{Z}}^{Ms})^{\circ\frac{1}{2}} && (51)\\
{}^{\delta}\boldsymbol{\Theta}^{s} &= (\Theta \times {}^{\delta}\hat{\mathbf{V}}^{s})^{\circ\frac{1}{2}} && (52)\\
{}^{\delta}\boldsymbol{\Lambda}^{r} &= (\Lambda \times {}^{\delta}\hat{\mathbf{X}}^{r})^{\circ\frac{1}{2}} && (53)\\
{}^{\delta}\boldsymbol{\Xi}^{r} &= (\Xi \times {}^{\delta}\hat{\mathbf{E}}^{r})^{\circ\frac{1}{2}} && (54)
\end{aligned}
$$

expect for cells of 0 value. We replace the 0 cells in the unbalanced initial IOP table with a very small number (1×10^{-10}). Due to different data reliability of each block, our principle is to let reliable block stick at around the original value. We achieve this principle by setting our five parameters:

$$
\begin{aligned}
\Gamma &= 1 \times 10^{5} && (55)\\
\Delta &= 1 \times 10^{1} && (56)\\
\Theta &= 1 && (57)\\
\Lambda &= 1 \times 10^{2} && (58)\\
\Xi &= 1 \times 10^{2} && (59)
\end{aligned}
$$

5.3 Additional Remarks

5.3.1 Negativity

When preparing the weight matrix, you may come across negative cells. Our practice is replacing the negative cells with the absolute value of the cell while the other parts of objective function remain unchanged. It ensures that the objective function is convex, and the range of varying of such cell remain the same.

5.3.2 Double-weighted variables

According to Equation 46, gross input of processing trade and export of processing trade should be the same, and these two variables are both in the objective function. That gives rise to the double-weighted problem: they should be granted the same weight as other variables that on the 1×10^2 level, but as a matter of fact they are granted twice the weight that they should be given. It may cause overestimation or underestimation of processing trade.

Our remedy is simple: we adjust the weights of those double-weighted variables to half of their original weight.

5.4 Concatenating

We fill our IOP table with the reconciled solution. Fill the remaining blocks with corresponding part of IO table, namely $^*\mathbf{F}$, $^*\mathbf{F}^M$ and $\mathbf{R}$.

5.5 Rearrangement

According to export data, some country-sector pair do not have processing trade, which is properly presented in the initial unbalanced IOP table. However, since we only set 0 cells to a very small value in the QP, we do not force these variable to 0. Even if they are not prone to vary away from 0, but they still have very small values. For precision purposes, we add these values to the corresponding N section, and zeros out the column with small values in P section.

6 Examples of Using IOP for Empirical Studies

6.1 Value-added Ratio of Processing Trade

Figure 2 is the scatter plot of sector-province pairs' value-added to the total output **X** where the two customs regimes is distinguished by colors. It is easy to tell that green dots that represent processing trade is more clustered at the bottom left (low Va, low X) whereas the read dots that represent non-processing trade is clustered at the top right (high Va, high X). The elasticity of X to Va for these two regimes are not significantly different, while the y-intercept for processing trade is lower, meaning that processing trade have lower value-added for a same total input X_0.

The differences between processing trade and non-processing trade are not limited to the Va-X ratio from the input perspective. From the output perspective, if we further disaggregate Va into several component according to different final uses, the weigh of processing trade in value added are further lessened. Based on a presumption of linearity, Leontief (1936) used a very simple formula to capture the influence on total output resulting from the variation of the final demand vector. Let **A**, **X**, **Y** be the direct coefficient matrix, total input/output vector, final use (aggregated column-wise hence a vector), respectively. In an autarky situation,

$$\mathbf{AX} + \mathbf{Y} = \mathbf{X} \tag{60}$$

$$(\mathbf{I} - \mathbf{A})\mathbf{X} = \mathbf{Y} \tag{61}$$

$$\mathbf{X} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y} \tag{62}$$

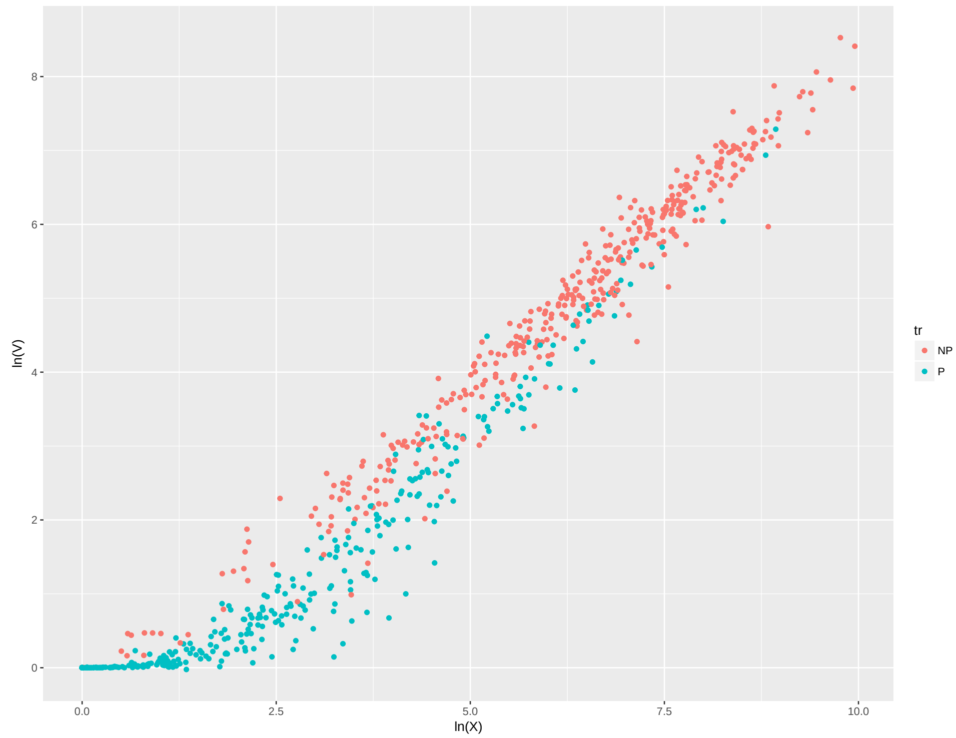

Figure 2: Value-Added **Va** changes along with Total Output **X**

Similarly to this methodology, Wang et al. (2017) made a step forward on the foundation of Leontief's work. Let **Va** be the value-added vector (aggregated row-wise and transposed into a column vector), and **V** be the value-added ratio to the total input ($\mathbf{V} = \text{diag}(\mathbf{X})^{-1}\mathbf{Va}$). Multiply both sides of Eq. (62) by diag(**V**), we have

$$\text{diag}(\mathbf{V})\mathbf{X} = \text{diag}(\mathbf{V})(\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y} \tag{63}$$

the LHS is the representation of value-added vector **Va**

$$\text{diag}(\mathbf{V})\mathbf{X} = \text{diag}(\mathbf{X})^{-1}\text{diag}(\mathbf{Va})\mathbf{X} = \text{diag}(\mathbf{Va})\text{diag}(\mathbf{X})^{-1}\mathbf{X} = \text{diag}(\mathbf{Va})\mathbf{1} = \mathbf{Va} \tag{64}$$

In Wang's work, they distinguish the domestic final use and foreign final use and disaggregated **Y** to further analyze the value-added flow to final use of different destination.

However, as for an open economy where there are import and export, some minor modifications are needed to maintain the validation of this model. Specifically considering the layout of the IOP table, let processing export be $\mathbf{E}_p$ and non-processing export be $\mathbf{E}_n$. Name the error term, if any, $\mathbf{R}$. Rewrite the Eq. (63),

$$\text{diag}(\mathbf{V})\mathbf{X} = \text{diag}(\mathbf{V})(\mathbf{I}-\mathbf{A})^{-1}(\mathbf{Y}+\mathbf{E}_p+\mathbf{E}_n+\mathbf{R}) \tag{65}$$

Hence, we divide the vector **Va** into 4 additive components,

$$\mathbf{Vd} = \text{diag}(\mathbf{V})(\mathbf{I}-\mathbf{A})^{-1}\mathbf{Y} \tag{66}$$
$$\mathbf{Vp} = \text{diag}(\mathbf{V})(\mathbf{I}-\mathbf{A})^{-1}\mathbf{E}_p \tag{67}$$
$$\mathbf{Vn} = \text{diag}(\mathbf{V})(\mathbf{I}-\mathbf{A})^{-1}\mathbf{E}_n \tag{68}$$
$$\mathbf{Vr} = \text{diag}(\mathbf{V})(\mathbf{I}-\mathbf{A})^{-1}\mathbf{R} \tag{69}$$
$$\mathbf{Va} = \mathbf{Vd}+\mathbf{Vp}+\mathbf{Vn}+\mathbf{Vr} \tag{70}$$

It is easy to examine the ratio of $\mathbf{Vn}$ to $\mathbf{Ve} = \mathbf{Vp} + \mathbf{Vn}$, just by running a simple OLS regression on this cross-sectional data excluding sectors that by nature do not contain processing trade.

$$Vp_{s,r} = \beta_1 Ve_{s,r} + \Sigma_s + \Gamma_r + \epsilon_{s,r} \tag{71}$$

Table 8: Vp to Ve

	Dependent variable:			
	Vp			
	(1)	(2)	(3)	(4)
Ve	0.295***	0.291***	0.285***	0.280***
	(0.005)	(0.005)	(0.005)	(0.005)
Sector FE	No	Yes	No	Yes
Province FE	No	No	Yes	Yes
Observations	2,604	2,604	2,604	2,604
R^2	0.568	0.605	0.583	0.621
Adjusted R^2	0.568	0.599	0.578	0.610
Note:		*p<0.1; **p<0.05; ***p<0.01		

It is not surprising to see that for processing trade in 2012 that makes up 43% of the total export according to CCS, it actually makes up 28% of the marginal value-added of export. That is to say, processing trade statistically tend to have low value-added ratio, which is a very intuitive result because the job entailed, usually assembling, does not involve high-skilled workers, which results in the high elasticity of demand curve in the labor market. Moreover, given the strong dependence on imports, processing exports induce less domestic economic activity.

6.2 Upstreamness

Apart from processing trade and ordinary trade, we can look further into the upstreamness that Antràs et al. (2012) proposed. According to his 2012 paper, the upstreamness U_i of sector i can be quantified by

$$\mathbf{U} = (\mathbf{I} - \mathbf{\Delta})^{-1}\mathbf{1} \tag{72}$$

where $\mathbf{\Delta}$ is the matrix with $Z_{ij}/(\sum_j Z_{ij} + \sum_f F_{if})$ [9] in entry (i, j) and $\mathbf{1}$ is a column vector of ones. The upstreamness actually means how many iterations are there for the value-added in the target sector to completely transfer to the final use. By aligning every two province, we can scrutinize the difference of upstreamness for a province pair. Take Anhui province and Jiangsu province as an example: these two provinces are geographically adjacent and share a lot of sectorial similarities (see Figure 3, right). This is specifically apparent when compared to Anhui-Guangzhou provincial pair. These two province have really different sectorial organization (see Figure 3, left). If two province have similar sectorial organization, the dot plot will lie adjacent to the red line ($x = y$). If it is not the case, the dot plot will fit the red line loosely. Similarly, another typical example would be Chongqing-Sichuan (see Figure 4).[10]

[9] i is the column index and j is the row index. f is the row index for Final Use matrices

[10] There are totally 961 (=31*31) figures for each province-province pair. To request the plot of a specific province-province pair, please contact me at 201519040@uibe.edu.cn

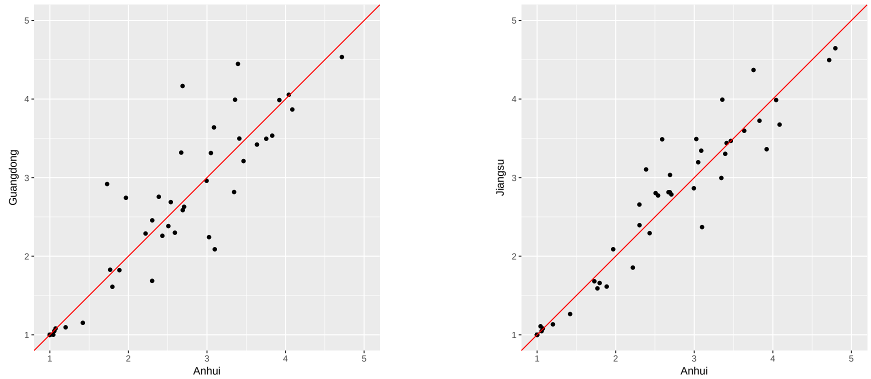

Figure 3: Anhui-Guangdong, Anhui-Jiangsu Upstreamness comparison

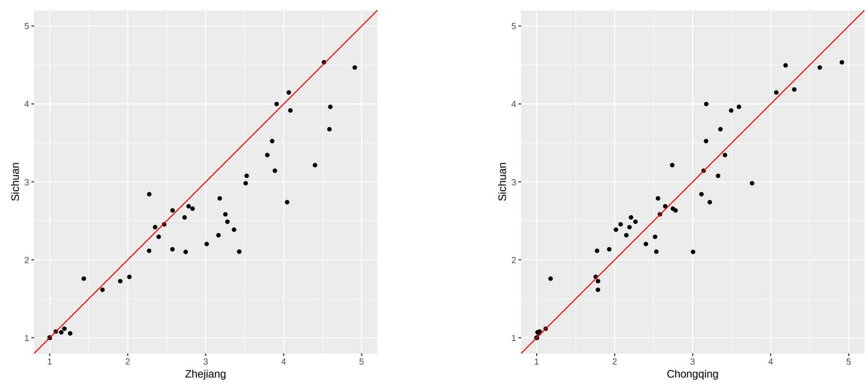

Figure 4: Sichuan-Zhejiang, Sichuan-Chongqing upstreamness comparison

To conclude, the sectorial organization is closely related to regional endowment, geographical similarities, to name a few. That is why adjacent regions (both geographically and culturally) are more likely to have similar upstreamness for each sector. More rigorous econometrical models can be developed to scrutinize the hypothesis. However, since we only have a set of cross-sectional data, without loss of generality, we can use cluster analysis (Maechler et al., 2017) to grasp the profile of the upstreamness similarities of China's 31 different regions.

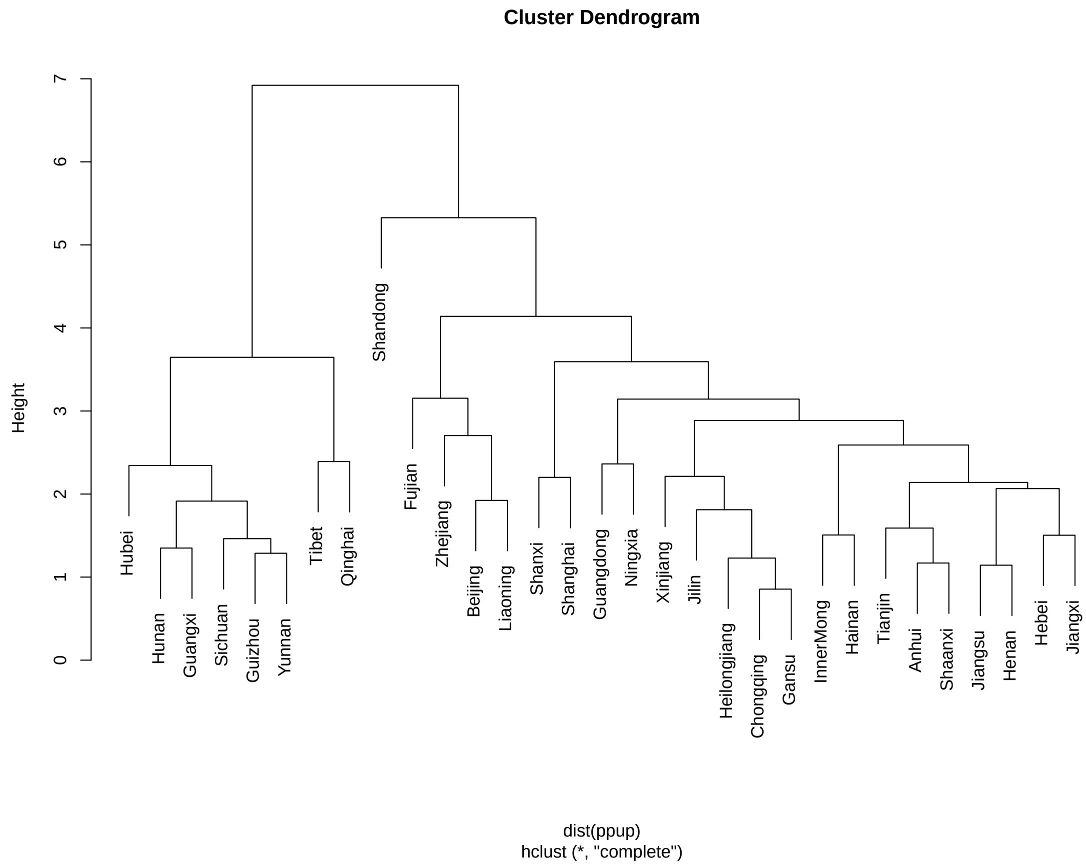

Figure 5: Cluster Analysis of Sectorial Upstreamness of 31 Regions

This dendrogram provides some proof of our hypothesis: Hunan, Guangxi, Sichuan, Guizhou, Yunnan formed a south/southwest cluster. For other parts of China, the pattern is not obvious enough to observe.

7 Conclusion

IOP table is a very effective tool for regional and sectorial studies in Economics. It, by nature, is a massive social accounting matrix (SAM) that contains more information than any conventional IO table or customs data. It has great potential in the field of international trade studies, especially when the international trade status quo have been challenged by several promising countries: Japan decades ago, China recently, and possibly more to come in the future.

We use multi-regional IO tables, DPN table, and customs data to create the IOP database for 2012 China. This quadratic-program-based method can be universally applied to database of all countries. One possible future development of IOP table is the time-series database which allows econometricians to extract penal data that they need for empirical models.

References

Pol Antràs, Davin Chor, Thibault Fally, and Russell Hillberry. Measuring the upstreamness of production and trade flows. *American Economic Review*, 102(3): 412–16, 2012.

David H Autor, David Dorn, and Gordon H Hanson. The china shock: Learning from labor-market adjustment to large changes in trade. *Annual Review of Economics*, 8:205–240, 2016.

Alonso de Gortari. Disentangling global value chains. *The Faculty of Arts and Sciences, Harvard, November. https://scholar. harvard. edu/files/alonsodegortari/files/degortari_dgvcs_october2017. pdf*, 2017.

Robert Koopman, Zhi Wang, and Shang-Jin Wei. How much of chinese exports is really made in china? assessing domestic value-added when processing trade is pervasive. Technical report, National Bureau of Economic Research, 2008.

Lawrence Lau, Xikang Chen, Leonard Cheng, K.C. Fung, Y Sung, Cuihong Yang, Kunfu Zhu, Jiansuo Pei, and Zhipeng Tang. Non-competitive input-output model and its application: an examination of the china-u.s. trade surplus. *Social Sciences in China*, 5:91–103, 01 2007.

Wassily W Leontief. Quantitative input and output relations in the economic systems of the united states. *The review of economic statistics*, pages 105–125, 1936.

Martin Maechler, Peter Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik. *cluster: Cluster Analysis Basics and Extensions*, 2017. R package version 2.0.6 — For new features, see the 'Changelog' file (in the package source).

Bo Meng, Zhi Wang, and Robert Koopman. How are global value chains fragmented and extended in china's domestic production networks? 2013.

Zhi Wang, Shang-Jin Wei, Xinding Yu, and Kunfu Zhu. Characterizing global value chains: Production length and upstreamness. Technical report, National Bureau of Economic Research, 2017.

Zhi Wang, Shang-Jin Wei, Xinding Yu, and Kunfu Zhu. Re-examining the effects of trading with china on local labor markets: A supply chain perspective. Technical report, National Bureau of Economic Research, 2018.