### The method to build a hybrid Input-output table with detailed employment data for China

#### Key processes

1. Aggregate the sectoral employment data from 511 sector classification in the original population census to 139 sectors in IO table.
2. Adjust sectoral employment data from 2010 to 2012.
3. Split the labor input in IO table to 28 detailed labor types.

The first process is realized through a GAMS file. However, the introduction for sector mapping is in Chinese since there are too much sectors in the original data.

The second and third process are realized through a single GAMS file. Detailed introduction for the methodology is shown in the rest of this document.

#### Objectives

Building a hybrid Input-output (IO) table with detailed employment data means splitting the labor input in the original IO table to a labor input matrix that includes the compensation of different types of labors. The splitting should not break the row and column equilibrium of IO table. The process is shown as **Fig.1**.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Sector 1 |  | Sector *i* |
| Labor Compensation |  | … |  |
|  | | | | |
| Compensation for labor *l* |  | … |  |
| … | | | |

Fig.1 The process of labor splitting

#### Data

The required data are shown as followed:

，the total value of labor compensation for sector *i*. Source: 2012 IO table. Set *i* represents different sectors in IO table.

，the quantity of employment of labor *l* in sector *i*. Source: the 6th population census in 2010. Set *l* represents up to 28 labor types by gender (male/female), by region (urban/rural), and by educational level (unlettered, elementary school, middle school, high school, junior college, regular college, postgraduate). Those data were adjusted from 2010 to 2012 according to China’s population growth in the statistic yearbook.

，the average wage for labor *l*. Source: 2013 Chinese Household Income Project (CHIP), which is a household survey hold by Beijing Normal University and includes 26527 samples. The wage data is shown in Table 1.

Table 1. Wage data for different labor types

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Gender** | **Region** | **Education** | **Wage**  **(Yuan)** | **No.** | Gender | **Region** | **Education** | **Wage (Yuan)** |
| L1 | Male | Urban | unlettered | 23431 | L15 | Female | Urban | unlettered | 14356 |
| L2 | elementary school | 26275 | L16 | elementary school | 18451 |
| L3 | middle school | 34098 | L17 | middle school | 23097 |
| L4 | high school | 39976 | L18 | high school | 31570 |
| L5 | junior college | 47648 | L19 | junior college | 36160 |
| L6 | regular college | 57187 | L20 | regular college | 46625 |
| L7 | postgraduate | 93353 | L21 | postgraduate | 68316 |
| L8 | Rural | unlettered | 17891 | L22 | Rural | unlettered | 12910 |
| L9 | elementary school | 21849 | L23 | elementary school | 16950 |
| L10 | middle school | 28150 | L24 | middle school | 20751 |
| L11 | high school | 30022 | L25 | high school | 23483 |
| L12 | junior college | 35971 | L26 | junior college | 29295 |
| L13 | regular college | 38878 | L27 | regular college | 33715 |
| L14 | postgraduate | 47189 | L28 | postgraduate | 28733 |

#### Method

Theoretically, the relationship between labor compensation and quantity is shown as followed. represents the compensation for labor *l* in sector *i*; represents the average wage for labor *l* in sector *i*; represents the quantity of labor *l* in sector *i*.



The current database can well support the data required for and . However, we cannot find so detailed sectoral wage data, . As a result, we should make some assumption to deal with the data base.

The method used here follows *Peters J.C., Hertel T.W. Matrix balancing with unknown total costs: Preserving economic relationships in the electric power sector[J]. Economic Systems Research, 2016, 28 (1): 1–20.*

Here we define the targeted matrix as X={}, where represents the balanced compensation for labor *l* in sector *i*. In order to get X, we build an original matrix A={} based on the available data (、和). In this way, the splitting of labor input is converted to the optimal problem to minimize the distance from matrix A to matrix X under several constrains. The definitions of and are shown as followed.





In the definition of , the first item represents the relative wage rate of labor *l* to the average level; The second item represents the average wage level in sector *1*. **There is a key assumption that the relative wage rate of different labor types is the consistent among sectors.** We don’t use the absolute value of wage data from micro survey but use them to calculate the relative wage rate. In this way, the problems caused by the differentials in data scope of macro and micro data source can be largely solved.

The optimal problem can be represented as followed:



S.t. 

The constrain is to keep the consistency between the split labor compensation with the compensation in original IO table. **Through the above processes, we can get the targeted matrix X with split labor compensation as well as balanced sectoral employment and sectoral wage for each labor type.**