ISSN: 2688-9323

Estimation of Intermediate Input of Chinese Industrial Enterprises

Guimei Feng*

College of Economics, Jinan University, Guangzhou 510632, China

Abstract

The lack of intermediate input information in China's industrial enterprise database after 2008 makes the research on micro industrial organization face a bottleneck. Based on the property that the production function of enterprises has a certain stability in the short term, this paper takes the manufacturing enterprises from 1998 to 2013 in the Database of Chinese industrial enterprises as samples, and estimates the intermediate input of enterprises from 2008 to 2010 and 2012 to 2013 by using the total industrial output value and intermediate input information from 1998 to 2007 as well as the total industrial output value information from 2008 to 2010 and 2012 to 2013. In order to test whether the estimation method presented in this paper will produce large errors, we also use this method to estimate the intermediate inputs from 1998 to 2007 and compare them with the observed values of the intermediate inputs in corresponding years. It is found that the estimated intermediate inputs are distributed symmetrically on and close to the observed intermediate inputs line. Our study is helpful for scholars to estimate the production function of enterprises by using the data from 2008 to 2013, so as to carry out relevant researches.

Keywords

Chinese Industrial Enterprises; Intermediate Input; Production Function Estimation.

1. Introduction

The database of Chinese industrial enterprises is established by the National Bureau of Statistics of China. It contains all state-owned industrial enterprises and non-state-owned industrial enterprises above designated size. The database contains the vast majority of China's industrial enterprises. At present, besides the economic census database, the database of Chinese industrial enterprises is the largest enterprise database available in China. The advantages of this database are obvious, which are embodied in the following aspects: First, its sample is large, covering all state-owned industrial enterprises and non-state-owned industrial enterprises above designated size in China. Second, its indicators are rich, including the basic situation and financial data of enterprises, such as enterprises code, name, address, industry, ownership, intermediate input (missing in 2008 and subsequent years), industrial total output value, export delivery value and so on. Third, it covers a longer period of time. Starting from 1998, the database has been updated to 2013. Due to the unique advantages of this database, it has been used by a large number of economists in recent years. Research papers using this database cover subjects such as industrial organization theory, firm theory, corporate finance, transformation economics, international trade, labor economics and regional economics [1]. The database of Chinese industrial enterprises provides indispensable materials for the study of microeconomic, but it is not perfect, and one of its problems is the lack of indicators.

The information of manufacturing enterprises in the database from 1998 to 2007 includes total industrial production value, industrial added value, number of employees, total wage, total fixed assets, intermediate input, etc. Scholars can use this information to estimate the

ISSN: 2688-9323

production function of an enterprise, so as to carry out researches related to variables such as productivity and markup [2-3]. However, the data of industrial enterprises from 2008 to 2013 lacks the information of intermediate input and industrial added value, which makes it impossible for scholars to effectively estimate the production function of enterprises. Therefore, most scholars limit the data of industrial enterprises to the period from 1999 to 2007 when studying related problems. For example, [4] calculated the productivity of Chinese industrial enterprises from 1999 to 2007 by using the least square method, fixed effect method, OP, LP. [5] used the data of Chinese industrial enterprises from 1999 to 2007 to explain the mystery of low-price export of Chinese enterprises from the perspective of enterprise markups. [6] used the data of all manufacturing enterprises from 2001 to 2007 to study the impact of financial marketization and financing constraints on the cost plus of enterprises. If the information of intermediate input of Chinese industrial enterprises from 2008 to 2013 can be estimated effectively, then the relevant research related to intermediate input, enterprise productivity and markup can be extended to 2013, which will be more helpful for the analysis of relevant problems after 2008.

Based on the fact that the production function of enterprises is stable in the short term, this paper takes the manufacturing industry of Chinese industrial enterprises database from 1998 to 2013 as the sample, and uses the information of industrial total output value and intermediate input from 1998 to 2007 as well as the information of industrial total output value from 2008 to 2010 and from 2012 to 2013 to estimate the intermediate input of enterprises from 2008 to 2010 and from 2012 to 2013.In order to test whether the estimation method in this paper will produce large errors, we also use this method to estimate the intermediate input from 1998 to 2007 and compare it with the observed intermediate input in the corresponding years. It is found that the estimated intermediate input is symmetrically distributed on the observed intermediate input line and is close to the observed intermediate input. The following arrangement of this paper is as follows: the second part is the relevant literature review, the third part is the estimation methods and results of enterprise intermediate input from 2008 to 2013, and the fourth part is the conclusion.

2. Literature Review

Related to this study are the following categories of literature: The first category is about the use of the databases of Chinese industrial enterprises. The second category is about the relationship between enterprise productivity estimation and intermediate input information. The third category is the relationship between enterprise markup estimation and intermediate input. The fourth category is the research on the topic of intermediate input. We will elaborate from these four aspects respectively.

The usage of the database of Chinese industrial enterprises. Scholars have obtained many outstanding achievements in economics by using the data of Chinese industrial enterprise database. For example, in terms of enterprise innovation research and development, [7] used the panel data of industrial enterprises above the specified size in China from 2001 to 2005 to investigate the factors influencing the innovation activities of Chinese enterprises by using the Tobit model, and found that there was an inverted U-shaped relationship between enterprise innovation, scale and market competition. In terms of international trade, [8] used the data of manufacturing enterprises from 1999 to 2003 to find that export is conducive to enterprises' TFP improvement. That is, export has a "learning effect". [9] investigated the influence of factor market distortion on the domestic value-added ratio of Chinese export enterprises by using the data of Chinese industrial enterprise database and customs trade database from 2000 to 2006, and found that factor market distortion significantly increased the domestic value-added ratio of Chinese export enterprises. In terms of industrial agglomeration, [10] investigated the impact of industrial agglomeration on enterprises and found that industrial agglomeration has a significant positive impact on enterprise scale. In terms of markup and productivity, [11] used the data of all state-owned enterprises and non-state-owned enterprises above designated size during 1998-2007 to study the impact of China's accession to the WTO on markup and productivity of manufacturing enterprises. It is found that cutting production tariff reduces the markup and improves the productivity of enterprises. The reduction of input tariffs also increased the markup and productivity of enterprises.

The total factor productivity (TFP) and intermediate input. The total factor productivity of an enterprise is generally interpreted as the "residual" part of total output after deducting the contribution of factor inputs. For example, output growth due to non-factor inputs such as technological progress or system improvement. Total factor productivity is a micro concept at the enterprise level. Therefore, theoretically, it should be estimated from the fitting production function of enterprises. that is, the coefficient of each factor should be estimated using enterprises output and input information, so as to get the remaining part after deducting the contribution of factors ---Total factor productivity. However, in the early stage, due to the lack of micro data, the estimation of total factor productivity was mainly carried out from the macro level of the country or industry. For example, [12-13], studied the situation and changes of China's total factor productivity from the perspective of economic growth. [14] studied the growth of total factor productivity in 37 double-digit industries in China from 1995 to 2002. The method to estimate total factor productivity at the macro level only uses macro factor inputs such as total fixed asset investment and employed population, which can still reveal national differences in economic performance in the absence of micro enterprise statistical data, but cannot answer micro questions such as which domestic enterprises are more efficient in production and operation. With micro data acquisition, there are more and more scholars start from the micro level to estimate the total factor productivity of the enterprise. For example, [15] used the data of industrial enterprises above designated size in China during 1998-2005 to estimate the TFP of enterprises by parametric method; [8] used the industrial data of all state-owned and non-state-owned enterprises above designated size during 1999-2003 to estimate the TFP of enterprises by non-parametric OP method. From micro level to estimate the total factor productivity of enterprises should use enterprise microscopic information such as output and inputs. For example, the fixed effect estimation method uses information such as industrial added value, fixed assets and employee size of an enterprise. OP method further uses current investment as the proxy variable of unobserved productivity shocks to solve the simultaneity problem, while LP uses intermediate inputs instead of investment as the proxy variable of unobserved productivity shocks. The generalized moment method (GMM) also uses the information of enterprise output, fixed assets, employees, intermediate input and so on. Thus, it can be seen that enterprise output and input information, such as industrial added value and intermediate input, are necessary to estimate enterprise total factor productivity.

Enterprise markup and intermediate input. Markup rate is defined as the deviation of price from marginal cost, which is usually used to describe the market structure and measure the market power of an enterprise. There are two main methods to measure the enterprise markup: accounting method and production function method. In the accounting method, the added value, wage expenditure and intermediate input of an enterprise are used to calculate the markup [16]. Due to the short life span of data samples in China, estimation results using accounting methods are not affected by the economic cycle and external shocks [5]. However, [17]believed that due to the different accounting methods used to deal with fixed costs, there was no stable time series relationship between markup and profit. The production function method estimates the production function of enterprises by using the information of enterprises output and input, and then calculates the markup of enterprises according to the formula derived from the cost minimization condition. [18] used this method to estimate the markup of enterprises. As can be

ISSN: 2688-9323

seen from above, both the accounting method and the production function method use the information of enterprises industrial added value or intermediate input in the estimation of enterprises markup, so the lack of information of enterprises intermediate input poses a challenge to the estimation of enterprises markup.

Research on intermediate input of enterprises. In addition to assisting scholars in estimating total factor productivity and markup, there are many researches directly related to firm intermediary input. For example, many scholars have studied the relationship between intermediate goods import and the productivity and performance of enterprises from different perspectives [19-24]. Recently, [25] discussed the relationship between intermediate tradable goods liberalization, market structure, and firms' markup. [26] explained the "quality change puzzle" by combining the complementarity of labor skill input and intermediate input quality of enterprises, that is, the increase in the proportion of college students employed by Chinese enterprises did not bring about significant improvement in the quality of export products.

As can be seen from the literature review above, the database of Chinese industrial enterprises has been widely used in the field of economics and has made a great contribution to the analysis of China's economic problems. The index of enterprise intermediate input information in China Industry Database is necessary for scholars who want to use the database for productivity estimation, markup estimation or research directly related to intermediate input. However, the lack of intermediate input information from 2008 to 2013 has made it difficult for scholars to analyze relevant issues in these years. Therefore, putting forward an effective method to estimate the input of intermediate factors in these years will be helpful for scholars to study and analyze related problems.

3. Methods and Results of Intermediate Input Estimation

Because the output of an enterprise is determined by the input and the production function of the enterprise, and the production function of enterprise is stable in time. Therefore, we can use the information of intermediate inputs of enterprises from 1998 to 2007 to estimate the intermediate inputs of enterprises from 2008 to 2013.The estimation method only requires the premise that enterprises seek to minimize costs.

Specifically, assumed that the production function of firm i in period t is $Y_{it} = F_{it}(L_{it}, K_{it}, M_{it})$, where L_{it}、K_{it}、M_{it} are labor, capital and intermediate input respectively. Enterprises pursue cost minimization. Considering the associated cost minimization problem, the Lagrange function is:

$$\Phi\left(L_{it}, K_{it}, M_{it}\right) = w_{it}L_{it} + r_{it}K_{it} + t_{it}M_{it} + v_{it}\left(\overline{Y_{it}} - F_{it}(L_{it}, K_{it}, M_{it})\right) \tag{1}$$

Where, w_{it} , r_{it} and t_{it} are the prices of labor, capital and intermediate goods respectively. The Lagrange multiplier $v_{it} = \partial\Phi/\partial Y_{it}$ reflects the marginal cost of the firm. According to the first-order condition of cost minimization, enterprises' demands for intermediate goods meet the following requirements:

$$\partial\Phi/\partial M_{it} = t_{it} - v_{it}\left(\partial F_{it}/\partial M_{it}\right) = 0 \tag{2}$$

According to Equation (2), the intermediate input of an enterprise can be obtained as follows:

$$t_{it}M_{it} = \left(\left(\frac{\partial F_{it}}{\partial M_{it}} \frac{M_{it}}{Y_{it}} \right) \Big/ \left(\frac{P_{it}}{\nu_{it}} \right) \right) P_{it}Y_{it} \quad (3)$$

Where P_{it} is the price of enterprise products. Denote $\theta_{it} = \left(\left(\partial F_{it} / \partial M_{it} \right) \left(M_{it} / Y_{it} \right) \right) / \left(P_{it} / \nu_{it} \right) = t_{it}M_{it} / P_{it}Y_{it}$, that is, θ_{it} is the ratio of the intermediate input to the total output value of the enterprise, and it changes at the level of the enterprise-year. In particular, we want to know whether θ_{it} varies significantly from year to year for any given enterprise, and if so, what is the trend?

In order to solve this problem, we use the data of industrial enterprises from 1998 to 2007, take the variable "total intermediate input" as the intermediate goods expenditure of the enterprise, the variable "total industrial output value" as the total output value of the enterprise, and set θ_{it} as the ratio of the two variables. In the calculation process, due to the measurement error of the data and the unexpected random shocks, the logarithm of the observed industrial gross output value is $y_{it} = \ln P_{it}Y_{it} + \varepsilon_{it}$, where ε_{it} is the independent identical distribution (*i.i.d*) random shocks. In order to eliminate the influence of non-observable errors on total output value, we refer to the practice of [18], and assume that the specific production function of an enterprise is in the form of Translog (hereinafter referred to as Translog), that is:

$$\begin{aligned} \mathrm{y}_{it} &= \beta_l l_{it} + \beta_k k_{it} + \beta_m m_{it} + \beta_{ll} l_{it}^2 + \beta_{kk} k_{it}^2 + \beta_{mm} m_{it}^2 \\ &+ \beta_{lk} l_{it} k_{it} + \beta_{km} k_{it} m_{it} + \beta_{lm} l_{it} m_{it} + \beta_{lkm} l_{it} k_{it} m_{it} + \omega_{it} + \varepsilon_{it} \end{aligned} \quad (4)$$

Where lowercase letters are the natural logarithm of the corresponding variable, such as $m_{it} = \ln t_{it}M_{it}$. ω_{it} is the non-observable heterogeneity productivity of the enterprise. Referring to [27], we selected intermediate material input (m) and capital (k) to control unobservable productivity, i.e $\omega_{it} = h_{it}\left(m_{it}, k_{it}\right)$. $h_{it}(\bullet)$ is still in Translog form ([28] used firm investment as a proxy for productivity, while [27] used intermediate input as a proxy for productivity.). Therefore, in order to obtain ε_{it}, we took the logarithm of industrial gross output value y_{it} as the dependent variable and $(l_{it}, k_{it}, m_{it}, l_{it}^2, k_{it}^2, m_{it}^2, l_{it}k_{it}, k_{it}m_{it}, l_{it}m_{it}, l_{it}k_{it}m_{it})$ as the independent variable to make OLS regression for each 2-bit code industry, and took the regression residual as the estimated value of ε_{it}. After $\widehat{\varepsilon_{it}}$ is estimated, then $\theta_{it} = t_{it}M_{it} \Big/ \left(P_{it}Y_{it} \Big/ \exp\left(\widehat{\varepsilon_{it}}\right) \right)$.Table 1 shows the distribution of θ_{it} in each industry, from which it can be seen that 90% of enterprises' value of θ_{it} is between 0.55 and 0.9.

In order to observe whether θ_{it} has a great difference in different years, for each enterprise, we use the difference between the maximum and minimum value of θ_{it} , $\vartheta_i = \max\left(\theta_{it}\right) - \min\left(\theta_{it}\right)$, to measure the variation of A from 1998 to 2007. As can be seen from Table 2, during the period 1998-2007, although the value of θ_{it} of enterprises has certain changes, the change of θ_{it} of more than half of enterprises is less than 0.1.

ISSN: 2688-9323

Table 1. Distribution of θ_{it} (the ratio of intermediate inputs to gross industrial output value)

Industry	p5	p25	p50	p75	p95
13	.595	.719	.754	.801	.867
14	.61	.712	.739	.781	.836
15	.563	.67	.711	.759	.815
16	.389	.501	.589	.675	.78
17	.648	.741	.773	.802	.852
18	.557	.698	.75	.789	.847
19	.609	.711	.754	.793	.852
20	.615	.715	.746	.782	.85
21	.615	.723	.757	.789	.837
22	.652	.742	.772	.797	.832
23	.55	.676	.73	.768	.816
24	.591	.718	.763	.795	.845
25	.608	.719	.764	.8	.852
26	.605	.725	.762	.797	.854
27	.54	.666	.707	.746	.8
28	.703	.767	.8	.826	.857
29	.608	.721	.754	.788	.841
30	.623	.739	.776	.807	.857
31	.596	.701	.738	.769	.815
32	.647	.741	.777	.808	.859
33	.619	.738	.79	.828	.888
34	.614	.732	.772	.803	.863
35	.614	.722	.754	.783	.84
36	.585	.702	.734	.773	.838
37	.584	.721	.756	.789	.854
39	.62	.738	.777	.805	.86
40	.524	.695	.759	.805	.894
41	.54	.689	.735	.78	.865
42	.588	.694	.745	.785	.844
43	.508	.699	.796	.857	.916

Note: P5, P25, P50, P75, and P95 represent the 5th, 25th, 75th, and 95th percentiles, respectively.

ISSN: 2688-9323

Table 2. Distribution of ϑ_i (the difference of the ratio of intermediate input to industrial gross output value from 1998 to 2007)

Industry	p5	p25	p50	p75	p95
13	0	.03	.065	.121	.251
14	0	.024	.06	.111	.199
15	0	.033	.08	.135	.251
16	0	.062	.114	.175	.272
17	0	.023	.052	.095	.204
18	0	.032	.076	.136	.307
19	0	.03	.065	.114	.242
20	0	.015	.048	.095	.21
21	0	.024	.053	.096	.204
22	0	.024	.049	.083	.187
23	0	.029	.063	.106	.221
24	0	.03	.068	.12	.262
25	0	.031	.065	.113	.229
26	0	.03	.065	.115	.238
27	0	.042	.088	.138	.28
28	0	.025	.046	.077	.161
29	0	.029	.067	.114	.231
30	0	.026	.06	.105	.233
31	0	.026	.059	.102	.221
32	0	.027	.057	.104	.208
33	0	.028	.066	.116	.239
34	0	.025	.059	.109	.234
35	0	.022	.053	.098	.21
36	0	.024	.06	.113	.228
37	0	.025	.06	.11	.232
39	0	.026	.063	.114	.243
40	0	.036	.091	.163	.332
41	0	.034	.085	.152	.291
42	0	.028	.064	.112	.235
43	0	.017	.074	.139	.275

Note: P5, P25, P50, P75, and P95 represent the 5th, 25th, 75th, and 95th percentiles, respectively.

In order to intuitively observe the change of the value of θ_{it} with years, we take the mean of θ_{it} by industry-year. Figure 1 shows the change of the mean of θ_{it} in each sector from 1998 to 2007. As can be seen from Figure 1, θ_{it} shows a downward trend in other industries except industry 23.

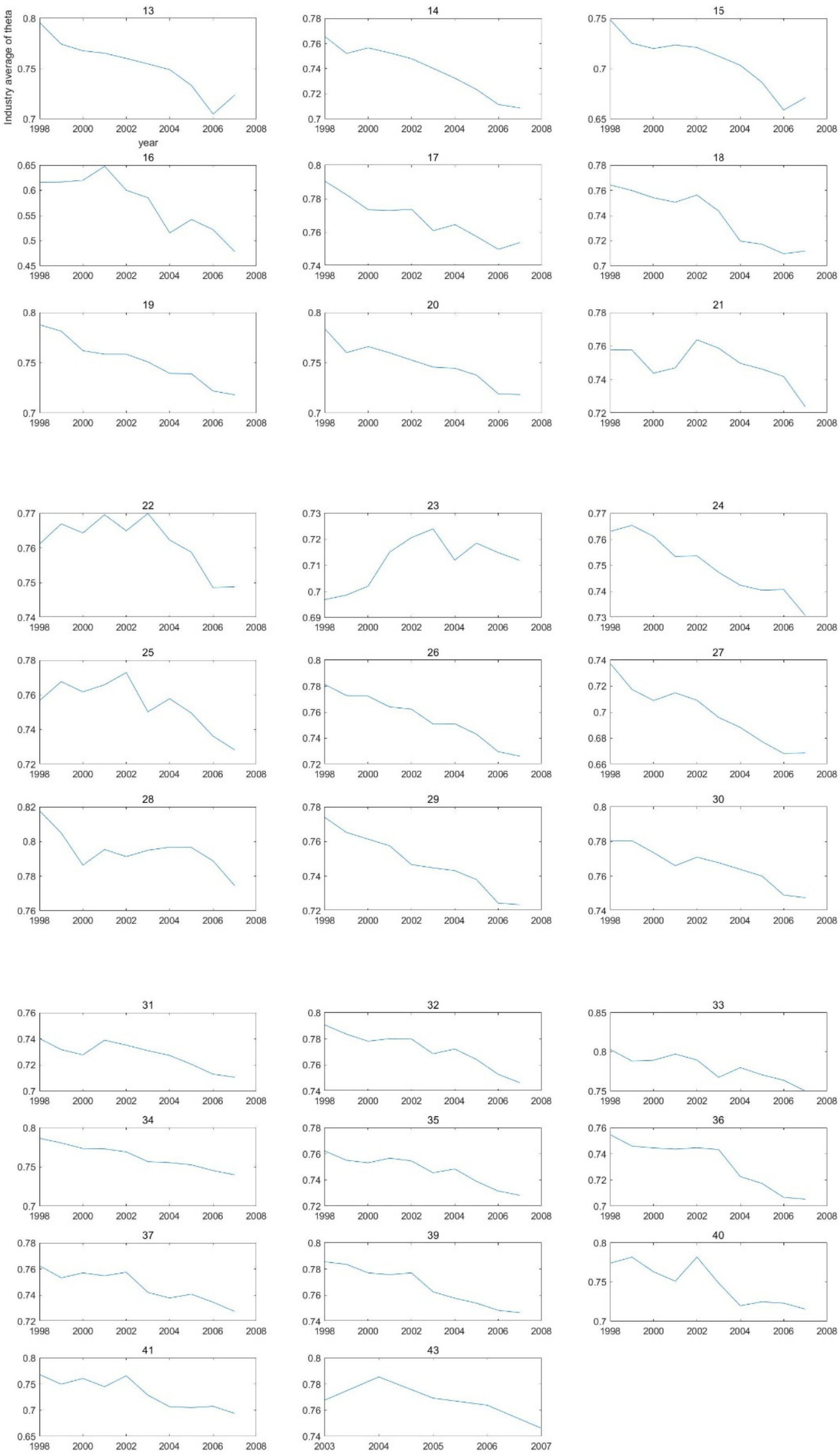

Figure 1. The trend of the industry average of the ratio of intermediate input to industrial gross output value from 1998 to 2007

ISSN: 2688-9323

According to Table 2 and Figure 1, we believe that θ_{it} has a downward trend. In order to depict the change of θ_{it} with time more specifically, we took 1998 as the base year to explore the influence of the distance between the observation year and the base year on θ_{it}. Specifically, we defined Δt to be equal to the observed year minus 1998, and performed the following OLS regression for each department:

$$\theta_{it} = \beta_0 + \beta_1 \Delta t + \phi_i + \varepsilon_{it} \tag{5}$$

Where ϕ_i is the firm fixed effect, and ε_{it} is the random error. Table 3 presents the coefficients and their t statistics for each industry. As can be seen from the table, β_1 is significantly negative and between -0.005 and -0.015, which indicates that θ_{it} decreases about 0.005 to 0.015 per year.

Table 3. Influence of distance between observation year and base year on θ_{it} (the ratio of intermediate input to gross industrial output value)

Industry	The coefficient of Δt	The value of statistic t	Industry	The coefficient of Δt	The value of statistic t
13	-0.014***	-126.30	28	-0.005***	-20.00
14	-0.013***	-113.01	29	-0.010***	-44.33
15	-0.015***	-76.55	30	-0.008***	-60.98
16	-0.011***	-19.26	31	-0.007***	-88.21
17	-0.008***	-94.92	32	-0.008***	-46.74
18	-0.009***	-55.38	33	-0.009***	-37.04
19	-0.009***	-55.83	34	-0.009***	-68.48
20	-0.011***	-48.34	35	-0.009***	-97.75
21	-0.006***	-24.29	36	-0.012***	-97.19
22	-0.006***	-50.19	37	0.010***	-84.06
23	-0.007***	-50.81	39	-0.009***	-84.06
24	-0.007***	-28.38	40	-0.012***	-61.02
25	-0.006***	-19.75	41	-0.015***	-62.06
26	-0.010***	-108.48	42	-0.007***	-32.14
27	-0.011***	-66.55	43	-0.013***	-6.13

By doing the regression of Equation (5), we prove that θ_{it} has a slight downward trend. Next, in order to purify the time trend of θ_{it}, we make the following regression for each industry:

$$\theta_{it} = \beta_0 + \beta_1 \Delta t + \beta_2 \Delta t^2 + \varepsilon_{it} \tag{6}$$

Where $\beta_1 \Delta t + \beta_2 \Delta t^2$ is the influence of time trend on θ_{it}. we subtracted this item from θ_{it} to get the ratio of intermediate input expenditure to industrial gross output value after removing time trend and denoted it as $\widetilde{\theta_{it}}$, that is $\widetilde{\theta_{it}} = \theta_{it} - \beta_1 \Delta t - \beta_2 \Delta t^2 = \beta_0 + \varepsilon_{it}$. We then take the mean of $\widetilde{\theta_{it}}$

ISSN: 2688-9323

from 1998 to 2007 for each enterprise, $\overline{\theta_i}$, as the ratio of intermediate input to gross industrial output value of each enterprise after removing the time trend. To get an estimate of the ratio of intermediate inputs to gross industrial output from 2008 to 2013, including the time trend, we add $\overline{\theta_i}$ to the corresponding time trend for each year, i.e. $\widehat{\theta_{it}} = \overline{\theta_i} + \beta_1 \Delta t + \beta_2 \Delta t^2$ .With this estimate, we can use the total industrial output for 2008-2011 and 2012-2013 to estimate the intermediate input for 2008-2013.

Before estimating intermediate inputs from 2008 to 2013, we first examine the difference between intermediate inputs estimated using this method and observed intermediate inputs using data from 1998 to 2007. It is worth pointing out that for every true intermediate input observed from the data, there is an estimated intermediate input corresponding to it. Now, we arrange the data for each year in ascending order of intermediate input observations, noting that after sorting, the true value of each intermediate input still corresponds to its estimated value. In Figure 2, we set the sort number as the abscissa, the observations and estimates of intermediate inputs as the ordinate. Due to large sample, the intermediate input values in the two-dimensional coordinate characterized by a smooth curve on the graph (solid curve on the graph), and because of the estimation error, the estimate of the corresponding intermediate input fall above or below the observed value (scattered points on the graph). Therefore, the vertical distance between the true value of each intermediate input and the corresponding estimated value measures the effectiveness of the estimation method in this paper. If the vertical distance between the true value of each intermediate input and the estimated value is small, then the estimation method in this paper is effective. If not, the error is large. It can be seen intuitively from this figure that the estimated intermediate input is symmetrically distributed on the observed intermediate input line, and the estimated value is close to the observed value. Therefore, the estimation method in this paper is effective.

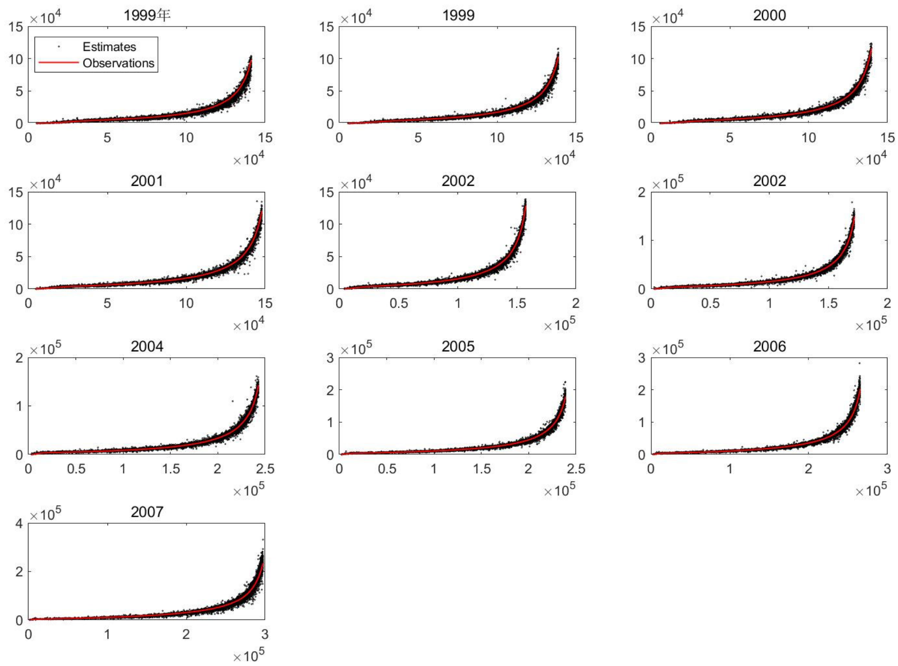

Figure 2. Observations and estimates of intermediate inputs from 1998 to 2007

ISSN: 2688-9323

To get an accurate measure of the difference between the two, we calculate the variable $\Delta\left(t_{it}M_{it}\right)=\left(\widehat{t_{it}M_{it}}-t_{it}M_{it}\right)\Big/t_{it}M_{it}$. Table 4 shows $\Delta\left(t_{it}M_{it}\right)$ for each industry. From this table, we know that approximately 50% of the estimates differ from the observed value by less than 2.5%, and approximately 90% of the estimates differ from the observed value by less than 10%. Therefore, the intermediate input estimated by this method does not produce large error.

Table 4. Differences between estimated and observed intermediate inputs from 1998 to 2007

Industry	p5	p25	p50	p75	p95
13	-.079	-.023	0	.019	.095
14	-.064	-.018	0	.017	.07
15	-.082	-.023	0	.022	.099
16	-.106	-.04	0	.044	.135
17	-.064	-.019	0	.016	.077
18	-.099	-.029	0	.023	.135
19	-.073	-.022	0	.018	.089
20	-.073	-.02	0	.015	.081
21	-.068	-.018	0	.019	.085
22	-.061	-.017	0	.014	.07
23	-.083	-.024	0	.026	.104
24	-.083	-.024	0	.02	.104
25	-.077	-.023	0	.021	.103
26	-.075	-.022	0	.017	.094
27	-.09	-.026	0	.024	.123
28	-.049	-.015	0	.015	.06
29	-.074	-.023	0	.018	.092
30	-.076	-.022	0	.017	.091
31	-.071	-.021	0	.019	.089
32	-.068	-.02	0	.016	.086
33	-.08	-.025	0	.02	.103
34	-.078	-.024	-.001	.016	.095
35	-.067	-.019	0	.018	.083
36	-.074	-.021	0	.019	.085
37	-.076	-.021	0	.019	.087
39	-.075	-.023	0	.018	.094
40	-.111	-.035	0	.027	.151
41	-.094	-.03	0	.024	.118
42	-.077	-.023	0	.023	.104
43	-.088	-.017	.003	.064	.24

Note: P5, P25, P50, P75, and P95 represent the 5th, 25th, 75th, and 95th percentiles, respectively.

ISSN: 2688-9323

Next, we estimate intermediate inputs for 2008-2013. The intermediate input estimation formula is, $\widehat{t_{it}M_{it}} = \widehat{\theta_{it}}\left(P_{it}Y_{it}/\exp(\varepsilon_{it})\right)$ but for 2008-2013, we do not have $\exp(\varepsilon_{it})$, so we estimate it in two steps. The first step is to get $\exp(\varepsilon_{it})$.

We set $\widetilde{t_{it}M_{it}} = \widehat{\theta_{it}}P_{it}Y_{it}$, and make the logarithm of industrial gross output value y_{it} perform OLS regression on $(l_{it}, k_{it}, \widetilde{m_{it}}, l_{it}^2, k_{it}^2, \widetilde{m_{it}^2}, l_{it}k_{it}, k_{it}\widetilde{m_{it}}, l_{it}\widetilde{m_{it}}, l_{it}k_{it}\widetilde{m_{it}})$ to obtain the residual $\widehat{\exp(\varepsilon_{it})}$ as the estimated value of $\exp(\varepsilon_{it})$, where $\widetilde{m_{it}}$ is the logarithm of $\widetilde{t_{it}M_{it}}$. In the second step we set $\widehat{t_{it}M_{it}} = \widehat{\theta_{it}}\left(P_{it}Y_{it}/\widehat{\exp(\varepsilon_{it})}\right)$ as an estimate of the intermediate input.

We have now estimated intermediate inputs for 2008-2013. In Table 5, we present complete summary statistics of intermediate inputs from 1998 to 2013.Among them, the data from 1998 to 2007 were directly observed from the database, and the intermediate input from 2008 to 2013 was obtained by the estimation method in this paper. However, the estimated intermediate input was missing due to the lack of industrial gross output value in 2011.

Table 5. Summary statistics of intermediate input after supplement from 1998 to 2013

Year	Sample size	Mean	Standard error	Min	Max
1998	148686	31147.5	194000	0	2.10e+07
1999	146094	33694.56	217000	-216	2.24e+07
2000	147243	39292.96	284000	-51806	2.51e+07
2001	155722	41647.05	302000	0	2.92e+07
2002	165853	45291.03	337000	0	3.89e+07
2003	181174	53734	430000	0	5.24e+07
2004	255768	52455.24	456000	-23	6.96e+07
2005	251482	66250.91	601000	-4911	1.19e+08
2006	279257	75055.47	702000	-85421	1.48e+08
2007	313017	86127.01	803000	-972	1.73e+08
2008	275287	97991.65	778000	.566	1.26e+08
2009	209728	99570.64	804000	.26	1.11e+08
2010	285694	114000	957000	.08	1.42e+08
2012	147392	239000	1630000	46.104	3.04e+08
2013	78986	259000	1880000	4.212	2.50e+08

Note: The intermediate input from 1998 to 2007 is directly observed from the database, while the intermediate input from 2008 to 2013 is estimated using the method in this paper. However, the estimated intermediate input is missing due to the lack of industrial gross output value in 2011.

4. Conclusion

This paper takes the manufacturing industry of China's industrial enterprise database from 1998 to 2013 as the sample, and uses the total industrial output value and intermediate input of enterprises from 1998 to 2007 as well as the information of total industrial output value from 2008 to 2010 and from 2012 to 2013 to estimate the intermediate input of enterprises from 2008 to 2010 and from 2012 to 2013.In order to test whether the estimation method in this paper will produce large errors, we also use this method to estimate the intermediate input from 1998 to 2007 and compare it with the observed intermediate input in the corresponding

ISSN: 2688-9323

years. It is found that the estimated intermediate input is symmetrically distributed on the observed intermediate input line and is close to the observed intermediate input. Therefore, this paper effectively estimated the information of intermediate input of enterprises during 2008-2010 and 2012-2013, which is helpful for scholars to use the data from 2008-2013 to conduct related research.

Acknowledgments

This work was financially supported by Chinese National Funding of Social Sciences (No. 17BJL110).

References

[1] Nie, H., Jing, T., & Yang, N. (2012). [The current situation and potential problems of Chinese industrial enterprise database]. Shi Jie Jing Ji, (05),142-158. (in Chinese).

[2] Yu, M. (2010). [Trade liberalization and productivity: evidence from Chinese firms]. Jing Ji Yan Jiu, (12), 97-110. (in Chinese).

[3] Liu, Q., & Huang, J. (2015). [Export probability, learning by exporting and low markup trap]. Jing Ji Yan Jiu, (12), 143-157. (in Chinese).

[4] Lu, X., & Lian, Y. (2012). [Estimation of total factor productivity of industrial enterprises in China: 1999-2007]. Jing Ji Xue, (02),541-558. (in Chinese).

[5] Sheng, D., & Wang, Y. (2012). [The puzzle of low-price export of Chinese enterprises -- based on the perspective of enterprise markup]. Guang Li Shi Jie, (05),8-23. (in Chinese).

[6] Li, X., & Xu, B. (2018). [Financial marketization, financing constraints and firm markups: evidence from Chinese manufacturing firms]. Guo Ji Mao Yi Wen Ti, (02), 164-174. (in Chinese).

[7] Nie, H., Tan, S., & Wang, Y. (2008). [Innovation, firm size, and market competition: based on panel data analysis at the Chinese firm level]. Shi Jie Jing Ji, (07),57-66. (in Chinese).

[8] Zhang, J., Li, Y., & Liu, Z. (2009). [Do exports boost productivity in Chinese firms? Empirical evidence from local manufacturing firms in China :1999 ~ 2003]. Guang Li Shi Jie, (12), 11-26. (in Chinese).

[9] Gao, X., Liu, Q., & Huang, J. (2018). [Factor market distortion and domestic value added ratio of Chinese firms' export: facts and mechanisms]. Shi Jie Jing Ji, (10), 26-50. (in Chinese).

[10] Li, D., Lu, Y., & Wu, M. (2012). Industrial agglomeration and firm size: Evidence from China. Regional Science and Urban Economics, 42(1-2), 135-143.

[11] changes in export product quality"]. Guan Li Shi Jie, (03), 1-23. (in Chinese).

[12] Brandt, L., Van Biesebroeck, J., Wang, L., & Zhang, Y. (2017). WTO accession and performance of Chinese manufacturing firms. American Economic Review, 107(9), 2784-2820.

[13] Zhang J. (2002). [Capital formation, industrialization and economic growth: understanding China's economic reform]. Jing Ji Yan Jiu, (06),3-13+93. (in Chinese).

[14] Guo, Q., & Jia, J. (2005). [Estimating total factor productivity in China]. Jing Ji Yan Jiu, (06),51-60. (in Chinese).

[15] Tu, Z., & Xiao, G. (2007). [Cost-efficiency analysis of China's large and medium-sized industries :1995 ~ 2002]. Shi Jie Jing Ji, (07),47-55. (in Chinese).

[16] Xie, Q., Luo, S., & Zhang, Y. (2008). [Productivity growth and convergence across China's industrial ezconomy]. Ji Jing Xue, (03), 809-826. (in Chinese).

[17] Domowitz, I., Hubbard, R. G., & Petersen, B. C. (1986). Business cycles and the relationship between concentration and price-cost margins. The RAND Journal of Economics, 1-17.

[18] Bresnahan, T. F. (1987). Competition and collusion in the American automobile industry: The 1955 price war. The Journal of Industrial Economics, 457-482.

[19] De Loecker, J., & Warzynski, F. (2012). Markups and firm-level export status. American economic review, 102(6), 2437-71.

[20] Zhang, J., Zheng, W., & Chen, Z. (2015). [Imports and firms' productivity: evidence from China]. Jing Ji Xue, (03), 1029-1052. (in Chinese).

[21] Yu, M., & Li, J. (2015). [Imports, heterogeneous industry and improvement of firm productivity]. Jing Ji Yan Jiu, (08), 85-97+113. (in Chinese).

[22] Xiang, X., Chen, T., & Chen, F. (2016). [Imported inputs, productivity and RMB exchange rate pass-through: an empirical study based on micro firm-level data]. Jing Rong Yan Jiu, (09), 33-49. (in Chinese).

[23] Qian, X., & Wang, B. (2017). [Intermediate inputs import, product switching and structure upgrading of firms' factor endowment]. Jing Ji Yan Jiu, (01),58-71. (in Chinese).

[24] Huang, X., Gao, Y., & Chan, S. (2018). [Imported intermediate inputs and firm productivity: short run effect and long run eqffect]. Guo Ji Mao Yi Wen Ti, (05), 54-67. (in Chinese).

[25] Gu, K., Li, X., & Cui, X. (2020). [Productive input imports and total factor productivity: horizontal effect and vertical spillover]. Guo Ji Mao Yi Wen Ti, (10), 1-16. (in Chinese).

[26] Liu, Z., & Ma, H. (2019). [Input trade liberalization, market structure and firm markup]. Jing Ji Ping Lun, (06), 109-133. (in Chinese) .

[27] Liu, Q., & Tie, Y. (2020). [Employment structure, intermediate inputs and China's "mystery of Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. The review of economic studies, 70(2), 317-341.

[28] Olley, G. S., & Pakes, A. (1992). The dynamics of productivity in the telecommunications equipment industry (No. w3977). National Bureau of Economic Research.