RIETI Discussion Paper Series 12-E-065

Accounting for Labor Input in Chinese Industry, 1949-2009

Harry X. WU
Hitotsubashi University

Ximing YUE
Renmin University of China

RIETI
Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
http://www.rieti.go.jp/en/

RIETI Discussion Paper Series12-E-065
October 2012

ACCOUNTING FOR LABOR INPUT IN CHINESE INDUSTRY, 1949-2009*

Harry X. WU
Institute of Economic Research, Hitotsubashi University, Tokyo
harry.wu@ier.hit-u.ac.jp

Ximing YUE
School of Public Finance, Renmin University of China, Beijing

ABSTRACT

Following the user cost theory on measuring labor input, after a careful scrutiny of available information, we construct employment and compensation matrices for China's industrial workforce over the period 1949-2009. Our measures are able to capture both individual and interactive effects of changes in gender, age, education, industry and ownership types of China's industrial workforce, and decompose the growth of labor input in Chinese industry into quantity and composition ("quality") effects. We find that the annual growth of the labor input in Chinese industry experienced a substantial decline from 6.9% per annum in the pre-reform period to 3.8% per annum in the post-reform period. Change of labor composition accounted for about 12% in the planning period (or 0.8% growth per annum), but it made little contribution during the reform period. We also find that the changes in industrial structure and age structure (reflecting the effects of seniority and experience) almost explained for the entire (positive) change in labor composition in the planning period. However, the change of education turned into negative after 1965 which made the average contribution of education negative in the planning period. Following the reform, education showed the most important contribution in the 1990s when the reform deepened, but the effect turned into negative again alongside China's entry into the World Trade Organization (WTO).

Keywords: Quantity of labor employment; Labor compensation; Translog labor input indexing; Iterative proportional fitting (IPF); Labor in economic transition

JEL Classification: E24, C82, O47

* We are indebted to Mary O'Mahony for her critical comments and constructive suggestions, to Barbara Fraumeni, Hiroki Kawai, Mun S. Ho and Marcel Timmer for their helpful advice on methodological issues, and to Li Huimin, Liu Fujiang, Qi Shuchang, Song Changqing, Wang Mingshan, Wang Xiaohui, Xu Xianchun, Zhang Xin and Zhou Xuewen from China's National Bureau of Statistics (NBS) for their kind help on data issues for many years. We are especially grateful for Zhang Guang's excellent research assistance in handling data problems, programming and carrying out rounds of estimation. Helpful suggestions received from participants at the WEAI in Taipei, the 31st IARIW Conference in Switzerland and seminars at GGDC, University of Groningen, Japan Research Institute of Economy, Trade and Industry (RIETI), IER at Hitotsubashi University, and The Conference Board China Centre. The financial support to the original data work by the Hong Kong Polytechnic University Research Grant (P-A93) and to the recent revision and update by the RIETI/CIP Project are gratefully acknowledged. The usual claims apply.

1. INTRODUCTION

Economists have long been debating about whether and to what extent China's market-oriented, gradualist reform has improved China's productivity performance. However, serious data problems encountered in measuring input and output variables in growth accounting exercise or production function analysis for the Chinese economy have made the debate remain inconclusive. Attempting to seriously tackle major measurement problems for all input and output variables in one project is difficult if not impossible. This is because it requires researchers to work at industry or disaggregate level of the economy dealing with the differences between the Chinese statistical practices and the international norms or standards, reconciling historical inconsistencies in statistical concept, coverage and classification, as well as looking for useful information to help fill important gaps. In this study, we aim to solve one of the key measurement problems for Chinese industry, that is, the problem of measuring labor input that is determined by both the quantity and composition of the industrial workforce.

The core issue in measuring labor input is how to hold the quality of hours worked constant when there are actually changes in the quality of workforce due to changes in the composition of the age, gender, education, occupation and industry of the workforce, or in other words, how to convert heterogeneous hours worked into homogenous volume of labor input. If failed to do so, for example, in the case of an increase of labor quality, the growth of total labor input will be understated and hence the growth of total factor productivity (TFP) will be exaggerated. The measure of natural numbers employed or hours worked is essential in that it provides a natural quantity base for the key task in the measurement. But, it alone does not conform to the theory of homogenous production function because if any input is not homogeneous in its components, the production function cannot be homothetically separable (Jorgenson, 1990, p.33). This is a core issue that has been made theoretically sound with empirical evidence thanks to the studies by, for example, Denison (1962, 1974), Jorgenson and Griliches (1967), Kendrick (1961, 1973), and later contributions by Chinloy (1980), Gollop and Jorgenson (1980, 1983), and Jorgenson, Gollop and Fraumeni (1987).[1]

[1] See a comprehensive review of these studies by Jorgenson (1990, pp. 32-41).

Labor input in the Chinese economy has never been properly measured, which is a major obstacle to an accurate understanding of the sources of growth in the economy. Most studies on the growth of the Chinese economy or its individual sectors (e.g. Borensztein and Ostry, 1996; Chan et al., 1988; Chow, 1993; Hu and Khan, 1997) have simply used the numbers employed as a proxy for labor input. Even if the measure of numbers employed is accurate, which is unfortunately untrue, we still face two major serious problems. Firstly, when there are changes in institutional working hours, the quantity base of labor input becomes inconsistent overtime. Secondly, when there are changes in the quality of labor, they will be counted as changes in the residual of the production function, hence making the estimated TFP performance of the Chinese economy or its individual sectors ambiguous.

There have been few studies attempting to measure the labor input in the Chinese economy according to the standard concept, especially for the long run. Li et al (1993) made the first ever effort to construct labor input indices for 34 sectors of the economy for a short period 1981-87 using the Jorgenson approach (Jorgenson, 1990). However, they did not attempt to tackle any conceptual and inconsistency problems in the official labor statistics. By contrast, Young (2003) devoted a significant part of his study on China's post-reform productivity growth to identifying and reconciling inconsistencies in the employment data from the official surveys and censuses. However, he did not attempt to work at disaggregate level to solve the problems. There is certainly an important knowledge gap in the measuring of labor input and hence the understanding of the main factors that have determined the changes of labor quality over both the central planning and the reform periods of the Chinese economy.

In this study, we first make a pioneer attempt to construct labor employment and compensation matrices for Chinese industry, cross-classified by demographic, educational, occupational and sectoral attributes of the workforce for selected benchmark years over the period 1949-2009. The most challenging task to us is how to apply the standard methodology to the available data that are not only rather limited in terms of what the methodology requires, but also suffer from serious conceptual and inconsistency problems. Obviously, we have to make various assumptions in order to reconcile the inconsistencies and fill the data gaps. In order to support the assumption making, we have to conduct thorough search for any relevant information, direct or indirect, through various historical official documents on labor regulation

and administration, employment planning, and policy studies on various labor issues, many of which were kept in the state archives until very recently. With the so-constructed employment and compensation matrices, following Chinloy (1980) and Jorgenson, Gollop and Fraumeni (1987), we then are able to express changes in labor quality as the sum of main effects associated with these demographic, educational, occupational, sectoral and institutional factors and their interactive effects in various orders yielding a growth accounting equation for the labor input in Chinese industry at different stage of economic development associated with important shifts of the policy regime.

This paper is structured as follows. In the next section, we review the main problems in measuring labor input in the Chinese economy. In Section 3, we introduce the methodology of labor input indexing and the decomposition of the contribution of individual human capital attributes to the change of labor quality. We devote Section 4 to discuss the key data problems in the construction of the marginal labor employment and compensation matrices. Section 5 describes the procedures of constructing full-dimension labor employment and compensation matrices. Section 6 reports the results of our labor input index and discuss the changes of the labor input due to the changes of the main and interactive effects of different attributes of workforce against the background of economic development and policy regime shifts in China. Finally, Section 7 concludes this study.

2. Problems in Measuring Quantity and Quality of the Chinese Workforce

The basic problem in Chinese labor statistics is not only that the Chinese system was substantially influenced by the Soviet system since the early 1950s which is fundamentally different from what commonly adopted as today's international standard, but more importantly, it cannot reflect significant changes in China's labor employment system as the consequences of policy or institutional changes while maintaining historical consistency. Like other factors of production, the absence of the market system in the allocation of labor under central planning implies that data on prices (wage rates) are scant. As for data on the quantity of employment, frequent policy shifts have affected the official indicators with significant changes in the definition of employment, the standard of industrial classification, and the statistical coverage of ownership type, size and accounting status of establishment (Wu, 2002)

but there is no sufficient information or an effective system for researchers to reconcile these inconsistencies by themselves.

The number one question in measuring sectoral or industry-level labor input is how to get the basic numbers right, that is, the number of workers employed and furthermore, the number of hours worked, which should be used as the control totals in the construction of the labor employment matrix. To answer this question, one encounters the following problems in the Chinese labor statistics.

Firstly, inconsistency in industrial classification has been a big hurdle to a proper measure of numbers employed at sector level over time. After China's implementation of the Soviet-style industrial classification standard to serve the administration of central planning, there were major changes in 1972, 1985, 1994 and 2002. These changes were to shift the standard of classification from one mainly facilitating the administrative and planning controls over individual sectors to one reflecting more about the technological nature of individual sectors in line with the international standard industrial classification (ISIC). However, there has been no official adjustment to the statistics of individual industries compiled under different standards. Available information is far from sufficient for the adjustment. In its largest ever data compilation entitled *Fifty Years of Chinese Industrial Development*, the Department of Industrial and Transportation Statistics (DITS, NBS), could only present discontinuous industry-level indicators in three separate tables for the periods 1949-84, 1985-92 and 1993-99, respectively. This means that at sectoral/industry level there are serious inconsistency problems in both input and output indicators that obstructed the construction of a conceptually compatible and integrated time series.

Secondly, as a long tradition in the central planning era, employees of an industrial establishment who provide services in education units, medical clinics, child care centres, commercial outlets, and social and political organisations, as long as they do not have independent accounting status, which is true in most state-owned enterprises, are included in the industrial employment statistics. [2] Systematic accounting for these "misallocated" employees simply does not exist. The only useful information at sector level can be found in two industrial censuses for 1985 and 1995

[2] For details about the categories of inappropriately included non-industrial employees in industrial labour statistics, see publications by statistical authorities, for example, NBS and MOL (1994, pp. 19-20) and DITS (1999, p. 52).

and recent economic census for 2004. However, any attempt for interpolation or extrapolation based on the census data has to first deal with the problem of inconsistent industrial classification. Chen et al. (1988) made a very crude correction for this factor in their productivity study on Chinese industry as a whole by assuming that the proportion of non-industrial employment in Chinese industry was equal to the ratio of residential housing stock to total fixed assets possessed by firms. Their approach is unlikely to produce a close proxy because it assumes that the non-industrial employment has the same fixed assets-labor ratio as that of the industrial employment and it ignores non-industrial employment engaged in services that are not related to residential housing. Besides, as this is largely a state sector phenomenon, the post-reform rapid development of the non-state sector implies that ignoring the difference between the state and non-state sectors is an inappropriate treatment to the problem.

Thirdly, in the Chinese labor statistics the quantity of employment has never been measured in its natural unit, that is, hours worked. Systematic data on hours worked simply do not exist. Almost all studies directly adopt the official indicator of numbers employed with little adjustment (e.g. Borensztein and Ostry, 1996; Chan et al., 1988; Chow, 1993; Hu and Khan, 1997), which implicitly assumes that there was no change in weekly working hour standard over time. However, as will be shown in the data section later, there have been several important reductions in weekly working hour standard, which means that numbers employed would have overstated the actual hours worked. Besides, since the institutional working hours are never the same across industries since the 1950s (Zhu, 1999) and the practice in regular working time is different between the state and non-state sectors since the reform, changes in industrial and ownership structure have definitely affected the actual hours worked by an average industrial worker. Besides, retaining off-post workers in the payroll and hence employment statistics is a different but related problem (DPSSTS, 1998, pp.62-63). This practice has been abandoned since 1998, but there has been no consistency adjustment in the official statistics (Holz and Lin, 2001, p.48). This means that even if one can convert numbers employed to hours worked, the actual hours worked would have still been inflated due to the improper inclusion of off-post workers.

Our number two question is how to measure the quality of China's workforce. From the growth accounting perspective, the proper measure of the quality change of

labor input is the difference between the user-cost-weighted index of labor input and the un-weighted or hours worked index (Chinloy, 1980; Denison, 1961; Gollop and Jorgenson, 1980 and 1983). This requires constructing both labor employment and compensation matrices at least for the benchmark years of the period under study. The construction of the employment matrix requires hours worked cross-classified by detailed demographic, educational and occupational attributes of China's workforce, whereas the construction of the compensation matrix requires exact element matching of the employment matrix with the compensation paid for per hour worked.

Unfortunately, compared with the information on the quantity of China's workforce, there is even scarcer information on the demographic, educational and occupational characteristics of China's workforce. Consequently, instead of searching alternative measure of labor quality, many growth accounting exercises or productivity studies on the Chinese economy have simply used numbers employed as a proxy for labor input, regardless the aforementioned problems in the official statistics on numbers employed. This implicitly assumes that workers embodied with different human capital stock are paid the same marginal product.

The only time series data source for measuring human capital contribution is the official statistics on the number of annual graduates with different levels of education attainment. But this time series is national aggregate only and comes without matching information on any characteristic of workforce. Following Barro and Lee (1997; 2000), some growth accounting studies (e.g. Wang and Yao, 2002) apply the perpetual inventory method (PIM) to such data to measure the stock of human capital in the Chinese economy. The so-estimated human capital stock cannot be a reliable proxy for the actual human capital service in the Chinese economy because education in China has been heavily controlled by the state regulations and national plans which have little concern about the (underlying) market needs. In such a context, it is also difficult to justify the (underlying) function of the depreciation of human capital. Even if this approach can be used as an useful alternative, it is inappropriate for the current study as there is no such data available at sectoral or industry level.

There are also some relevant data from occasional censuses and surveys. But, of China's five population censuses and three industrial censuses,[3] only the 1990 and

[3] The five population censuses were conducted for 1953, 1964, 1982, 1990 and 2000 and the three industrial censuses were conducted for 1951, 1985 and 1995.

2000 population censuses and the 1985 and 1995 industrial censuses are somewhat useful. However, the design of both types of the censuses does not allow full cross-classification of different demographic, educational and occupational attributes of the workforce. Besides, the two censuses are incompatible not only with each other, but also with the official labor statistics based on the regular annual reporting system. The quality of the census data has been seriously questioned by researchers. Young has empirically shown that the age-education profiles of the population censuses to some extent exaggerate the actual education attainment due to the improper inclusion of the data on adult education (2003, pp. 1240-44).

Compared with labor employment data, there are even much less information on labor compensation. The only available time series data are annual wage bills and average wage per employee by (broader) sector without any cross-classification by human capital attributes. To obtain relative wage estimates for weighting the changing composition of labor force, Li et al had to rely on their own labor compensation survey of less than 50,000 effective samples for the estimation of 34 sectors, but they give no detailed information about the time and the location of the survey (1993, p.163). Young (2003, pp. 1245-46) was able to access to personal income data from the NBS household surveys in 1986-92,[4] supplemented by the CASS (Chinese Academy of Social Sciences) household surveys in 1988 and 1995, arriving at a final sample size of 222,281. He relied on regression approach to capture the effects of human capital attributes of individuals. He found out that these household surveys have been heavily biased towards better-educated households (p. 1245). However, Young's data do not allow similar work at disaggregate level.

Another difficulty in constructing the compensation matrix is how to estimate the non-wage/salary income or income in kind paid to employees as part of labor compensation. For example, employees of the state sector enjoyed heavily subsidized housing (up to the end of 1990s) and other welfare payments in kind, which may vary greatly across industries and regions. NBS has made some efforts to improve its measure on labor compensation along with its development of SNA-type of input-

[4] The NBS conducts annual rural and urban household surveys that include some income data of household members. However, there has been no public access to the full survey data except for regularly published national and regional averages based on the survey results. In the early 1990s, part of the original survey data became commercially available in Hong Kong.

output tables (Xu, 2000). Nevertheless, these efforts have not been taken into account in the previous studies on labor input.

3. LABOR INPUT INDEXING

Labor input indexing should be discussed coherently with a production function aggregating the services provided by different types of labor and capital. The essential idea of constructing labor input index roots in the heterogeneity of labor in the sense that different types of labor have different marginal products in a given period. For example, an increase in the share of hours worked by skilled labor or by labor with better human capital will increase labor input even if the total hours worked remain unchanged (Denison, 1962; Jorgenson and Griliches, 1967). Directly using the numbers employed as a proxy for labor input in a production function implicitly assumes that labor is homogenous and the same hours worked by different types of labor will provide identical volume of services, which will certainly affect the reliability of the estimated residual.

Suppose that we have the following production function at time *t*, separable between labor and capital inputs, with a Hick's neutral shift parameter *A*:

$$Y_t = A_t f(L_t, K_{1t}, ..., K_{jt}) \tag{1}$$

where Y_t represents output, L_t labor input, and $K_{1t}, ..., K_{jt}$ the services of different capital inputs. Now let us define the labor aggregate as a function of hours worked by different types of labor:

$$L_t = \phi(H_{1t}, ..., H_{nt}) \tag{2}$$

where H_{it}, $i = 1, \ldots, n$, represents hours worked by type *i* labor. Following Chinloy (1980) and Jorgenson, Gollop and Fraumeni (1987), if assume efficient labor market and linear homogeneity of ϕ, then we have:

$$\frac{\partial \ln L_t}{\partial t} = \sum_{i=1}^{n} s_{it} \frac{\partial \ln H_{it}}{\partial t} \tag{3}$$

where s_{it} is the share of the *i*th type of labor in total labor compensation, which is equal to its logarithmic marginal output under the efficiency assumption:

(4) $$s_{it} = \frac{w_{it} H_{it}}{\sum_{i=1}^{n} w_{it} H_{it}} = \frac{\partial \ln \phi}{\partial \ln H_{it}}$$

In Equation (4), the hourly wage of the *i*th type of labor is w_{it} and its compensation is $w_{it} H_{it}$. The growth rate of labor input is a convex combination of growth rates of total hours for each type of labor, with compensation shares as weights. Equation (4) also indicates that the necessary condition for producer equilibrium is given by equality between the share of the *i*th type of labor in the labor aggregate and the elasticity of the aggregate with respect to the *i*th type of labor.

Let total hours worked by all types of labor be $H_t = \sum_{i=1}^{n} H_{it}$. Then, the growth rate of H_t is the sum of the weighted growth rates of hours worked by each type of labor:

(5) $$\frac{\partial \ln H_t}{\partial t} = \sum_{i=1}^{n} b_{it} \frac{\partial \ln H_{it}}{\partial t}$$

with $b_{it} = H_{it} / \sum_{i=1}^{n} H_{it}$ the weight of the *i*th labor type. Therefore, average labor quality per hour can be defined as labor input divided by hours worked:

(6) $$Q_t = L_t / H_t$$

and its growth rate is:

(7) $$\frac{\partial \ln Q_t}{\partial t} = \sum_{i=1}^{n} (s_{it} - b_{it}) \frac{\partial \ln H_{it}}{\partial t}$$

which is the sum of growth rates of hours worked by each type of labor, weighted by the *difference* between the shares in labor compensation and hours worked.

Now following Christensen, Jorgenson and Lau (1973), we specify the labor aggregate in the translog form:

(8) $$\ln L_t = \alpha_0 + \sum_{i=1}^{n} \alpha_i \ln H_{it} + 1/2 \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} \ln H_{it} \ln H_{jt},$$

where α_0, α_i, i = 1, …, n, and β_{ij}, i, j = 1, …, n, are parameters and where $\beta_{ij} = \beta_{ji}$ to satisfy the required symmetry conditions. Under linear homogeneity, we have $\sum_{i=1}^{n} \alpha_i = 1$ and $\sum_{j=1}^{n} \beta_{ij} = 0, \quad i = 1, \ldots, n$.

With the efficiency assumption, the relative share of the ith type of labor equals its logarithmic marginal product:

$$s_{it} = \alpha_i + \sum_{j=1}^{n} \beta_{ij} \ln H_{jt} \tag{9}$$

where s_{it} is defined the same as in equation (4).

Equations (8) and (9) as well as the symmetry conditions $\beta_{ij} = \beta_{ji}$ imply that the growth rate of the translog index of labor input l_t is:

$$l_t \equiv \Delta \ln L_t = \sum_{i=1}^{n} v_{it} \, \Delta \ln H_{it} \tag{10}$$

where $v_{it} = (s_{it} + s_{it-1})/2$ and Δ demotes the first difference operator.

From equation (6), the growth rate of quality q_t is defined as:

$$q_t \equiv \Delta \ln Q_t = \Delta \ln L_t - \Delta \ln H_t = l_t - h_t \tag{11}$$

where h_t represents the growth rate of total hours worked by all types of labor. Clearly, the growth rate of quality will be positive if hours worked by relatively high wage labor increase more rapidly than total hours worked.

Next, the contribution of each attribute of labor to quality change can be decomposed into two types of effects: the main effect of the attribute and the interactive effects of the attribute with each of the rest attributes. The main effect of the ith attribute is defined as the difference between the growth rates of labor input due to the ith attribute and total hours worked, regardless the time subscript:

$$q_i = l_i - h \tag{12}$$

where h is exactly the same as h_t defined in equation (11) and l_i is growth rate of labor input due to the ith attribute or the factor i. In the case of $q_i > 0$, as noted in

Jorgenson and Griliches (1967), labor input measured as total hours worked is biased downward and hence TFP is biased upward.

Suppose that there are two attributes of labor, *i* and *k*, as a proper subset from *n* factors, a first-order interactive effect is derived from the partial index growth rate l_{ik} for the two factors and the single factor indices l_i and l_k:

$$q_{ik} = (l_{ik} - h) - (l_i - h) - (l_k - h) = l_{ik} - h - q_i - q_k \tag{13}$$

that is, the joint effect of *i* and *k* or $(l_{ik} - h)$ less the main effect of each. If there are only two factors, *i* and *k*, the growth rate of labor quality is defined as the summation of the main effects of two factors and their first-order interactive effect:

$$q = l_{ik} - h = q_i + q_k + q_{ik}\,. \tag{14}$$

As for labor input with *j* factors, $l_{1,\ldots,j}$, interactive effects up to $(j - 1)$th order are obtainable following the same principle (Chinloy, 1980, p.111).

4. Construction of the Basic Data for Marginal matrices

The procedures explained in the previous section are a very data-demanding exercise that can encounter many problems even in the US case where better labor survey and census data are available than in many other countries (see Jorgenson, Gollop and Fraumeni, 1987). Given the data problems in the Chinese official labor statistics as discussed in Section 2, the challenges that we face are difficult to exaggerate. Only the size of the labor employment and compensation matrices that are to be constructed implies a big challenge. The number of factors that is considered affecting labor composition determines the number of the dimensions of the matrices. As listed in Table 1, in this study we aim to construct compatible labor employment and compensation matrices cross-classified by four primary aspects, each consisting of heterogeneous levels or qualifications, that is, gender (g=2), age (a=7 groups), education attainment (e=5 levels) and sector (s=24 manufacturing, mining and utility industries). Therefore, each of the full-dimension matrices is a four-dimension matrix with 1680 cells for each time point of the period 1949-2009.

TABLE 1
CLASSIFICATION FOR SECTORAL AND HUMAN CAPITAL ATTRIBUTES OF LABOR INPUTS

Industry	Labor Composition
1. Coal mining	**Primary:**
2. Oil and gas extraction	*Gender*: (*g*)
3. Metallic mineral mining	1. Male
4. Non-metallic minerals mining	2. Female
5. Food & kindred products	*Age Group*: (*a*)
6. Tobacco products	1. 15-19
7. Textiles	2. 20-24
8. Apparel	3. 25-29
9. Leather & leather products	4. 30-39
10. Saw mill products & furniture	5. 40-49
11. Paper products, printing & publishing	6. 50-54
12. Petroleum & coal products	7. >54
13. Chemicals & allied products	*Education Attainment*: (*e*)
14. Rubber & plastics products	1. Illiteracy or semi-illiteracy
15. Stone, clay & glass products	2. Primary school
16. Metals smelting, pressing & rolling	3. Junior high school
17. Metal products	4. Senior high school
18. Industrial machinery & equipment	5. Tertiary education
19. Transportation equipment	
20. Electrical equipment	**Additional:**
21. Electronic & telecommunication equip.	*Occupation Type*: (*j*)
22. Instruments and office equipment	1. Production worker
23. Miscellaneous manufacturing	2. Technician
24. Power, steam, gas and tap water supply	3. Administrator including office staff
	4. Auxiliary service staff
	Ownership Type: (*o*)
	1. State-owned enterprises (SOEs)
	2. Foreign-invested enterprises (FIEs)
	3. Domestic non-SOEs (NSEs)

In the exercise, as shown in Table 1, we also consider two additional aspects in the composition of labor, that is occupation (j=4 types) and ownership status (o=3 types). Ownership type is not considered in the standard framework. The literature also suggests that occupation could be ignored because it is insignificant due to its high correlation with education and age. They are considered in this study because of some advantages in data construction and problem identification. Firstly, more systematic time series data on labor employment and wage payment are available by ownership type. Compared with other ownership types, data for the state firms are generally more comprehensive and systematic. Secondly, occupation and ownership-based employment data occasionally include working hours, especially the state firms of different occupation types, which are helpful in our estimation of hours worked though they are still limited and unsystematic. Thirdly, occupation-based data can help identify auxiliary service workers who should be excluded from the industrial

workforce in theory. Lastly, considering labor compensation, although ownership type itself does not determine the quality of labor but it does affect the compensation of labor. This is especially the case following the market-oriented reforms because of the state control and exclusive power, the state firms especially those engaged in highly state-monopolized industries tend to pay more for the same type of labor than their non-state or foreign counterparts.

In what follows, based on sorting, cleansing and processing the available official data, we will first construct the total numbers of employment by sector, cross-classified by occupation and ownership types for the period 1949-2009 (i.e. $s \times j \times o$). Our major tasks at this stage include the clarification of officially used concepts of industrial employment, the reconciliation of the official data under different industrial classification standards, and the adjustment for coverage problems. The results of these exercises will be converted to hours worked. Using information from surveys and censuses we will then build up benchmark year matrices with more dimensions than what are constructed in the time series. The benchmark years are selected based on the best available data. But they are still marginal because no available data for any time point can satisfy full-dimensioned cross-classification. Data problems in constructing the benchmark matrices and techniques adopted to fix the problems will be discussed in detail.

4.1 Data Construction for Employment (Marginal Matrices)

Official data on numbers employed

To prepare for the basic data work we first have to understand the available official data on numbers employed. Let us begin with a discussion of conceptual problems in the official data. However, when dealing with conceptual problems, it is important to separate state firms from non-state firms because compared with the available data for non-state firms, the available data for state firms tend to be more detailed and reliable, which can serve as the "hard core" in our data construction. Such a separation can also help deal with ownership-specific data problems such as converting numbers employed to hours worked, and identifying and removing the employment by social or residential service units operated by state industrial firms for their employees.

There are two key concepts in the Chinese official labor statistics, namely "staff and workers" (*zhigong*) and "persons engaged in employment" (*congye renyuan*).

The latter is also known as "social laborers" (*shehui laodongzhe*) that ceased using in 1993. By definition, as a long tradition developed under central planning, the former refers to the employees who were administered by the state labor employment system, although they did not necessarily work in the state-owned enterprises or units, while the latter covers all wage earners including those who are not classified as "staff and workers". "Staff and workers" are employed by enterprises that are registered as legal entities (legal persons) with independent accounting status (officially defined as independent accounting units or IAUs which are required to maintain accounting books and make regular financial reports), whereas those who are not classified as "staff and workers" usually work in small (largely rural) factories attached to IAUs (i.e. making sideline products in addition to the main business of IAUs), in household or joint household-run (largely seasonal) business, or simply as self-employed. However, the official employment indicators based on these concepts are by no means clear.

Under China's National Bureau of Statistics (NBS) there are two departments that regularly publish employment statistics, namely, the Department of Industrial and Transportation Statistics (DITS) and the Department of Population and Employment Statistics (DPES), [5] both reporting data for "staff and workers" and "persons employed" and their sub-categories. However, users often find that for the same indicator the two official sources may report different data. This requires a good understanding of the definition used by the different authorities. Of all available indicators, the most compatible indicator from the two sources is the *state* "staff and workers". From 1952 to 1997, DITS and DPES reported almost identical data for the state "staff and workers", except for the period 1980-84 and 1994-97 in which there was a slight difference between the two sources. But, there were two major changes in 1998 which created a break in the indicator.

The first change relates to how to statistically treat "staff and workers" in the state units who were supposed to be permanently employed by the state under central planning. Prior to 1998, off-post staff and workers were still kept in the payroll and

[5] DPES was previously named as Department of Social Statistics (DSS) and renamed recently as Department of Population, Social and Science and Technology Statistics (DPSSTS). However, these changes do not affect DPES indicators. Since conceptually using DPES can avoid confusion due to the name change, we will use DPES throughout this study when it refers to any DPES employment concept. References will still be handled in the standard way.

hence in the employment list of state firms, which was a phenomenon that emerged along with the industrial reform in mid 1980s, especially since the 1990s, when the marketisation became intensified. In 1998 the statistical authorities decided to remove the off-post staff and workers, which have created a break in the existing series that is difficult to adjust because there are not available data on off-post staff and workers either prior to or after the change. Both DPES and DITS adopted this change. However, at the same time DITS renamed its indicator "staff and workers" to "persons engaged".[6] This confuses users because there is already a DPES indicator with the same name. For convenience, we will still use the original name of "staff and workers" for the DITS data since 1998.

The second change relates to how to define the "state economy". Traditionally, the "state economy" means completely state-owned. While DPES has followed this definition, DITS changed it to "state-dominated in share holdings" in 1998 (see DITS, 1999, p.57). Mainly because of this change, in 1998 the number of state industrial employment reported by DITS is 9.6% bigger than the number reported by DPES. In 2000, the difference jumped to 42.9%.

Now, let us enlarge the scope of our investigation from *state* to *total* "staff and workers". The biggest discrepancy between the two official sources is found with the indicator of *total* "staff and workers" since 1978 when the DITS series began to diverge from the DPES series surpassing the latter by 11.2% in 1978 and 35.5% in 2000, which deserves a closer examination of the definition used by the two authorities. We find a significant definitional incompatibility between the two sources. The DITS definition refers to the employment at or above the rural township level, while the DPES definition refers to the employment located in cities including the so-called "industrial areas" that are administratively treated as urban areas. Obviously, both definitions are administrative level-based, but the DITS definition has a wider coverage than that of the DPES because it includes employment outside cities. However, there was also a change in 1998 when DITS redefined its total "staff and workers" as all state enterprises and non-state enterprises with at least 5 million yuan of annual sales (DITS, 2000, p.16, footnote), shifting from administrative level-based

[6] Strictly speaking, DITS did not simply "rename" its indicator. Following the change in 1998 DITS decided to include "others" in its series. This category includes those who are either re-employed retirees or foreigners (DITS, 2000, p.296). We assume that this inclusion has no significant effect on the DITS series.

to ownership and enterprise size-mixed criterion. This again created a break that has to be tackled in our exercise.

As for the indicator of "persons engaged", i.e. the indicator that includes both "staff and workers" and those not defined as "staff and workers", DPES provides a long time series back to 1949 that largely maintains conceptual consistency. By contrast, DITS has never published the same indicator with the same definition (bear in mind that DITS confusingly renamed its "staff and workers" to "persons engaged" in 1998, but we have decided to stick to its original name). In this study, we use the DPES series of "persons engaged" as the control totals.

Our exercise requires these indicators with industry breakdown at least at the two digit level of the Chinese standard industrial classification (CSIC) that is consistent over time. For the state "staff and workers" both DPES and DITS provide industry level data based on their own definitions but there are significant gaps in some years and inconsistencies in industrial classifications over time. In addition, there have been two DITS industrial censuses in 1985 and 1995 that provide much more detailed industry breakdowns as well as other information that can help us check and adjust the historical series.

Table 2 summarises the features of the basic available data and required tasks in our data construction. Our exercise aims to construct a data set with three layers that are conceptually and CSIC-consistent: 1) the state employment as the "hard core", 2) the employment in the "township-layer" that consists of all the non-state employment qualified as at or above the rural township level (mixed with the size criterion since 1998 as previously discussed), including employment in collective-owned enterprises, foreign invested enterprises and private enterprises, and 3) the "outer layer" that contains all other people who also engaged in industrial activities. The first two layers cover all the "staff and workers" according to the DITS (wider) definition including the "staff and workers" as defined by the DPES narrower definition, plus employment in the category of "others" as in the practice of DITS.

TABLE 2
AVAILABLE OFFICIAL DATA ON INDUSTRIAL EMPLOYMENT AND THE REQUIRED DATA WORK IN THIS STUDY

Available Data	Required Data Work
DPES "Persons Engaged": • Industrial aggregate, 1949-2000 (CSIC inconsistency problem; coverage problem) • Mining, manufacturing, utilities, 1978-2009	• Used as the overall control totals that help derive persons engaged in industries outside the DITS "S&W", i.e. employment within the "outer layer" • Used as broad sectoral control totals that also help the work above
DPES "Staff & Workers": • State, 2-digit level industries, 1952-92, 1998-2009 (CSIC-inconsistent) • Total, state, urban collectives, industrial aggregate, 1949-77 • Total, urban collectives, other ownership types (including FDI); mining, manufacturing, utilities, 1978-2009	• Used as the "hard core"; correct for CSIC inconsistency; crosscheck with the DITS state "staff and workers", and fill gaps • Together with DITS data, estimate DITS non-state "staff and workers" at the 2-digit level • Same as the previous point; in addition, information on employment of other ownership types helps work on estimating labor compensation
DITS "Staff & Workers":[a] • State, 2-digit level industries, 1952-79, 1995-2009 (CSIC-inconsistent) • Total, 2-digit level industries, 1985-2009 (CSIC-inconsistent)	• Adjust to definitional changes of state "staff and workers"; reconcile with DPES data; correct for CSIC inconsistency • Used for deriving the "township layer" at the 2-digit level; reconcile data using different CSICs, and fill gaps
Other Sources: • The 1985 Industrial Census,[b] ownership types, 3/4-digit level, some indicators back to 1980 • The 1995 Industrial Census,[b] 3/4-digit level industries • The 2004 and 2008 National Economic Censuses,[b] 3/4-digit level industries • MoA[c] rural township and village enterprises, 2-digit level, 1987-97	• Both censuses can be used as much more detailed benchmark that help reconcile different CSICs and decompose non-state aggregates (including aggregates of the "outer layer"), as well as check and adjust annual data • Used for decomposing the "township layer" and "outer layer" aggregates to the 2-digit level

Notes:
a) Renamed to "Persons engaged" in 1998, see text for explanation.
b) DITS is the main authority of these censuses.
c) Data are published by the Bureau of Township and Village Enterprises, Ministry of Agriculture (MoA).

Reconciliation of different standards of industrial classification

As indicated in Table 2, the available sectoral-level data are inconsistent overtime because of changes in China's standard industrial classification (CSIC) system. China implemented its first CSIC in 1972, which in fact only institutionalised its practice in industrial classification following the Soviet classification system adopted in early 1950s. Subsequently it has made two major changes, that is, a shift from the 1972 CSIC to the 1985 CSIC and then from the 1985 to the 1994 CSIC, aiming to change

the industrial classification from one facilitating administrative and planning controls over industries to one that more reflects the production or technological nature of individual industries.

The Soviet industrial classification system was designed to serve the need of central planning. It intends to facilitate administrative controls by different ministries over resource allocation for the production of the key capital and consumer goods. Individual industries were therefore classified into "groups" according to their vertical links in input-output chains rather than their technological natures. Such "groups" are typically reflected by the classification of two-digit level industries in the 1972 CSIC. For example, since all metal ore mining, smelting and processing activities were administered by the Ministry of Metallurgy, they were grouped together as the metal industry (01, referring to the first two digits under the 1972 CSIC in Table 3). The economic reform began in the end of the 1970s induced a need for a significant change in the standard industrial classification, which was reflected in the 1985 CSIC. The 1985 CSIC, implemented in the DITS 1985 Industrial Census, is considered a major effort to move towards the international standard industrial classification (ISIC). In the current study, we use the 1994 CSIC for industrial classification, which is the revision of the 1985 CSIC and is in principle equivalent to the ISIC 1988 Revision (NBS, 1998, pp. 25-26).[7]

However, there has been no official adjustment to convert the data under the different versions of CSIC to the prevailing 1994 standards, which is a major problem for us. As the examples demonstrated in Table 3, since the change in CSIC involved separating or merging existing two-digit level industries, any reconciliation of the historical data using different versions of CSIC requires more detailed (higher-digit level) statistics. However, the published information is at most the aggregates of two-digit level industries if not broader industrial groups.

For example, in the 1972 CSIC, metallic, coal and petroleum are all two-digit-level industries, identified by the *first two digits* "01", "03" and "04", respectively. In the central planning period, these industries were accordingly administered by three ministries with the same names. As the examples show, at the two-digit level, mining or extraction activities are mixed with manufacturing activities, which do not comfort

[7] Note that the 1994 CSIC, coded as GB/T4754-1994 in China's national standard system, was further revised in 2003 (NBS, 2003, pp. 23-25).

to the homogeneity principle underlying the standard of industrial classification. However, if data on higher-digit level or subordinate industries were available, our reconciliation job would have not been too difficult. For instance, if the time series data for the 0310, 0321 and 0322 industries are available, employment in coal mining, coking and coal gas production could be easily separated and then converted to the two-digit industries in the 1994 CSIC. But, such data simply do not exist in the official publication.

TABLE 3
EXAMPLES ON RECONCILIATION OF DIFFERENT CHINESE STANDARDS OF INDUSTRIAL CLASSIFICATION

Wu-Yue Code[a]	1994 CISC		1985 CSIC		1972 CSIC	
	Code	Industry	Code	Industry	Code	Industry
02	⇦07	Oil and natural gas extraction	⇦0900	Oil and natural gas extraction	⇦0401 ⇦0402	Oil extraction Gas extraction
12	⇦25	Petroleum refinery and coking	⇦3400	Petroleum refinery	⇦0403	Petroleum refinery
			⇦3510 ⇦3520	Coking Coal gas	⇦0321 ⇦0322	Coking Coal gas
24[b]	⇦45	Coal gas			⇦0310	Coal mining
01	⇦06	Coal mining	⇦0800	Coal mining		
…	…	…	…	…	…	
03	⇦08	Ferrous metal ore mining	⇦1000	Ferrous metal ore mining	⇦0111	Ferrous metal ore mining
	⇦09	Non-ferrous metal ore mining	⇦1100	Non-ferrous metal ore mining	⇦0121	Non-ferrous metal ore mining
16	⇦32	Ferrous metal smelting and pressing	⇦4800	Ferrous metal smelting and pressing	⇦0112	Ferrous metal smelting and pressing
	⇦33	Non-ferrous metal smelting and pressing	⇦4900	Non-ferrous metal smelting and pressing	⇦0122	Non-ferrous metal smelting and pressing

a) See Table 1 for the name of the industry in the coding system of this study.
b) In this study, the coal gas industry is included in "utilities" (24).

To solve this problem, we mainly rely on the NBS annual bulletins (*nian bao*) published by various NBS departments for internal use.[8] Data for individual industries reported in these bulletins are national aggregates that are compiled through intermediate aggregations by NBS local offices based on regular enterprise-level statistical reports. Prior to the 1980s, the bulletins mainly focus on state enterprises. Subsequently, they have included non-state firms at or above the rural

[8] We are very much indebted to our NBS colleagues who made the achieved historical data available to the authors.

township level. For higher-digit level industries, the annual bulletins are by no means complete. For example, for some two-digit level industries, if the two-digit level aggregation was conducted in local offices, information on the industry's subordinate sectors would not be sent to the NBS headquarters and hence not be included in the bulletins. Furthermore, many annual bulletins were unfortunately lost in the Cultural Revolution (1966-1976) during which NBS was abolished for five years (1968-72). Lastly, most of the information on intermediate aggregations kept in local offices was also lost and the rest, given various constraints, cannot be easily retrieved.

We attempt to exhaust all available employment data for individual industries, as detailed as possible. However, for some periods we have to deal with those "industry groups" without any breakdown of subordinate industries. In case of a partial breakdown of a group, the targeted unknown industry may be derived as a residual. For example, as shown in Table 3, the unavailable employment data for crude oil and gas extraction (0401 and 0402) can be obtained by subtracting the number of employment in petroleum refinery (0403) from the number of employment of the petroleum industry (04). In the case of no breakdown at all for an industry in a period, we use the intra-industry weights of the nearest periods by the mid-point interpolation approach.[9] We try to avoid using output weights when filling the gaps in employment because it unrealistically assumes constant labor-output relationship.

At the end of this stage of our data work, we were able to construct the 1994 CSIC-consistent, two-digit level time series for the numbers employed by the state enterprises (the "hard core") for the period 1952-92 and 1995-2000, leaving gaps in the periods 1949-51 and 1993-94. We were also able to construct the 1994 CSIC-consistent numbers employed by all enterprises at or above the township level (the "township layer", i.e. the DITS concept of non-state "staff and workers") for the period 1980-2000, leaving the pre-1980 period uncovered. Strictly speaking, we need to adjust the 1998-2000 part of the series for the shift from "township" to "designated size" to maintain consistency. In what follows, we explain how all these gaps are filled and how the "outer layer" with the same industry breakdown is constructed.

Filling the gaps in the basic data

[9] We believe that the mid-point interpolation for structure has some advantage over the straight line interpolation because it does not affect the control totals. Besides, structural changes reflect not only an industry's own growth but also its relative growth to other industries within a group.

Our first task is to complete the state "hard core" series by filling the gaps in 1949-51 and 1993-94. Since it is unreasonable to expect any significant structural change over such short periods, if data on the state total are available, we can simply allocate the total into individual industries according the industrial structure at the time right before or/and after the gaps. With the readily available state totals for 1993 and 1994, we can fill this gap by using the average of the industrial structures in 1991 and 1995 (to decompose the state total of 1993) and the average of the industrial structures in 1993 (estimated) and 1995 (to decompose the state total of 1994). However, the state totals for 1949-51 deserve more work. Some historical data from a NBS publication (DSS, 1987, p. 83) suggest that they are incompatible with the series of the state "staff and workers" since 1952. In fact, what make up the difference are those non-state firms that were transformed into state-private joint ownership in the campaign of "socialist transformation" during 1956-57. Since these firms, together with other cooperatives, were further converted into complete state ownership after 1957, they are counted as the state "staff and workers" in all NBS series since 1952. We follow the same approach by adding the employment in such a case to the state sector, which gives a plausible result for the period 1949-51.[10] The so-estimated state totals are distributed to individual industries based on the industrial structure in 1952.

Our next task is to extend the 1980-2009 township-layer "staff and workers" or the non-state part of the DITS "staff and workers" back to 1949. For this purpose, we need information that is sufficient for constructing annual aggregates at this level and estimating industrial structure of employment at least for some benchmark years anchoring the required industry level interpolations. The township layer consists of two components: 1) "staff and workers" in urban non-state enterprises such as collective firms, cooperatives, and foreign invested enterprises, except for self-employed and people working in private firms, and 2) "staff and workers" in rural township enterprises. As given by the definition, the first component is in line with the DPES concept of "staff and workers" for which data on annual aggregates are already available (see Table 2). Our exercise then mainly focuses on the second component, that is, constructing annual aggregates of rural industrial employment for

[10] This approach is actually the same as what used in DSS, which can be seen in other two tables in DSS (1987, pp. 13 and 26) that perhaps give the only estimates for state total and industrial employment back to 1949, though they are seldom shown in other NBS publications.

the period 1949-79 that are compatible with the DITS concept of township "staff and workers".

In official statistics, rural enterprises mainly consist of township and village enterprises (TVEs) which were transformed from factories run by people's communes or their production brigades during the decollectivization in early 1980s which abandoned people's communes. In terms of the administrative hierarchy in rural China, post-reform townships are fully compatible with pre-reform communes. Back in history, the commune-run factories originated from rural industrial or handicraft cooperatives emerged in the cooperative movement in 1954-57. During the Maoist feverish Great Leap Forward (GLF) campaign in 1958-60, people's communes replaced all types of farmers' cooperatives as a more radical form of collective farming. Meanwhile, rural industrial and handicraft cooperatives were transformed into bigger commune factories. However, the collapse of the GLF was a dead blow to most of the commune factories which were closed down as required under the government retreat policy and did not recover until mid to late 1970s.

Historical data on rural enterprises are extremely scant. Thanks to NBS, we fortunately obtained some unpublished data including NBS surveys on China's traditional and handicraft industries in 1954-55, internal statistical report of commune factories in 1959, DITS *Annual Bulletin* on commune factories in 1978-83 (1979 missing) and DITS *Annual Bulletin* on township enterprises in 1984-85. After carefully processing these data and checking through total numbers employed and their industrial distribution, together with other scattered information from the government, we can set up three compatible benchmarks: 1956 (cooperatives), 1962 (early commune factory period) and 1978 (late commune factory period), which can be used to extend the non-state component of the DITS 1985-2000 township level series.

The 1954-55 survey on handicraft industries covered both urban and rural areas and distinguished laborers engaged in different industries from those who could not be identified by type of industrial activity. We assume that the majority of the former located in cities already worked in factories and, as we have pointed out, they should have been already included in the urban part of the non-state component of the series (covered by DPES, see Table 2). We assume that before the cooperative movement in 1955, those located in rural areas were largely self-employed or worked in family-

based workshops which are incompatible with the DITS series. A NBS publication has lent some support to our hypothesis. It reported that the number of employment of handicraft-making cooperatives increased from 0.60 million in 1954 to 0.98 million in 1955, and further jumped to 4.85 million in 1956 when "advanced cooperatives" were promoted.[11] Subtracting the urban part of 3.34 million in 1956, we obtain 1.51 million that is used as the starting point of the rural part of the series.

A large number of "advanced cooperatives" were transformed into commune factories during the Great Leap Forward campaign in 1958-60. It is not clear about the size of employment during that period. The only official statistical publication in the central planning period, *The Great Ten Years*, reported there were 700,000 commune factories by the mid-1959 (NBS, 1959, p.37), but did not give any data on the employment of these factories. Driven by GLF, 740,000 rural cooperatives were merged into 26,000 much larger communes (NBS, 1959, p.27). It is therefore not exaggerating to assume that on average each factory hired at least 20 workers. This means that there would be 14 million or more industrial workers at the commune (township) level or equivalent to nearly 60 percent of the urban employment in industry (24 million, as given by the DPES concept "staff and workers"), which may be unrealistic. This suggests that either the number of commune factories or the actual factory size was exaggerated. Other sources suggest either of these could be true.[12] One of the government circulars for the post-GLF policy retreat in 1962 disclosed that by the end of 1961 there were only 1.26 millions worked in the rural commune factories.[13] We take this information more seriously than other data in the context of the harsh policy retreat. Using the trend-deviation approach, we interpolate the commune factory employment between 1957 and 1961 with the growth of commune industrial output as an "indicator variable" (which gives the deviations over this period). The result is plausible showing that the number of commune factory

[11] See a table reporting changes of status of persons engaged in handicraft industries by DSS (1987, p.86).

[12] One source is newly discovered NBS internal report, which gives the number of employment in commune factories in 1959 as 5.8 million (NBS, 1960). Another source is one of the DITS annual publications, which reports the number of commune factories in 1960 as 117,000 (DITS, 1989, p.21). Since the government only began its retreat policy in the industrial sector in the mid-1961 (Wang and Dong, 1995, pp.98-102), this suggests that the number of commune factories, 700,000, by the mid-1959 as given in *The Great Ten Years* (NBS, 1959, p.37) is very implausible.

[13] This circular was jointly issued by the CCP Central Committee and the State Council on 27 May, 1962, aiming to support agricultural production and reduce demand for foodstuff after the collapse of GLF (SC/DRC, 2000, V.4 (II), pp. 537-540).

employment reached the record high peak in 1960 during this period (3.38 million). If following the change of the industrial output of communes, the commune factory employment would drop to about 500,000 in 1962 and 270,000 in 1963 or the lowest since 1956 that is consistent with the situation in the overall economy. Also based on the industrial output of communes, we assume that the employment of commune factories recovered to the 1962 level by 1965.

The next benchmark is 1978. The published DITS "staff and workers" data cover the period 1985 to 2000 (DITS, 2000, pp.84-93 and 111-117). Based on its *Annual Bulletin* and the 1985 industrial census, DITS statisticians helped us extend the series back to 1980. Instead of fully disclosing the approach that they used to construct the data for 1980-84, they provided us with the *Annual Bulletin* for 1978 and 1980-85. Crosschecking DITS estimates with these internal data, we have discovered that DITS estimates were smaller than what were directly available from the *Annual Bulletins*. This means that DITS must have conducted a screening exercise to make the numbers employed qualified for their concept of "staff and workers". For 1985, the number of township employment is 10.48 million, which is 79 percent of the *Annual Bulletin*-reported figure of 13.28 million. For 1980, as estimated by DITS upon our request, the estimated number is 5 million or 55 percent of the *Annual Bulletin*-reported 9.09 million. It seems that DITS considered the reported number of employment for commune factories being less "qualified" than that of township enterprises. We apply this "qualification assumption" to 1978 and obtain an estimate of 4.22 million for the number of commune factory employment for that year instead of accepting 7.67 million in the 1978 *Annual Bulletin*. The gap between 1965 and 1978 is filled with the trend-deviation interpolation approach. We should not forget that in 1998 the DITS definition for non-state "staff and workers" changed from "township" to "designated size", which created a significant break (down from 18.04 million in 1997 to 11.10 million in 1998 and further to 3.86 million in 1999). Thus, to maintain consistency of the whole series, the DITS data for 1998-2000 are adjusted back to the township definition based on the existing 1997 employment in the DITS series and the growth rate of rural enterprises derived from the DPES system.[14]

[14] DPES reports national employment data with ownership type break down for 16 (large) sectors (e.g. industrial data are available for mining, manufacturing and utilities), which can be used to derive rural collective enterprises by subtracting persons engaged in urban enterprises, private firms and the

To complete the data work for the "township layer", we need to distribute annual aggregates to the industrial sectors as defined in Table 1. Conceptually, we can derive employment for two-digit industries by subtracting the state "staff and workers" from the DITS "staff and workers". The difference is the non-state "staff and workers" located in both urban (collectives, joint ownerships, foreign invested firms) and rural (commune/township) enterprises. With the available data, after correcting for CSIC inconsistency, we can obtain such results for the period 1985-2000.[15] For the period 1949-84, we rely on four industrial structure benchmarks, namely 1955, 1959, 1978 and 1985, to anchor the two-digit level interpolations. The 1955 benchmark of industrial structure is obtained from the 1955 handicraft survey. We calculate the industrial structure for 1995 using the number of employment engaged in urban and rural industrial cooperatives with identifiable industries. For the 1959 and 1978 benchmarks, we directly used the industrial structure of commune factories, assuming urban non-state firms to have the same structure, as there is no other information available. The 1985 benchmark is set up with data from the 1985 *Annual Bulletin* for township enterprises and the 1985 industrial census.

Our last task is to construct annual aggregates of the "outer layer" and then distribute them to the same industrial sectors. The annual aggregates can be derived by subtracting the sum of the state and "township layer" employment from the DPES control totals (a concept that has been discussed).[16] For the pre-reform period, we rely on two benchmarks to control the sectoral distribution of the aggregates, 1955 and 1985. The 1955 benchmark is constructed based on the number of people engaged in the rural handicrafts attached to agriculture with identifiable industrial classification as found in the 1955 handicraft survey. The 1985 benchmark is constructed using industrial structure of the employment in village enterprises from the 1985 industrial census. We use the mid-point interpolation approach to estimate the employment structure between the two benchmarks. For the period 1949-54, we use the 1955 benchmark assuming the employment structure at this level did not change over that

---

self employed. However, since the so-derived contains rural village enterprises, we assume that the growth rate of township is the same as that of village enterprises.

[15] Since we have adjusted the annual aggregates of the rural component for 1998-2000, we use the existing industrial structure of this period to distribute the adjusted aggregates.

[16] However, China's population census data have shown that the DPES control totals may be wrong (Yue, 2005). Since any adjustment will affect the whole system, we do not attempt to do so in the present study.

period. Considering the level of economic development and technology in rural China during that period, this assumption is not too difficult to accept.

For the period after 1985, there are more data available. Over the period 1987-97, the Ministry of Agriculture (MoA) regularly published employment data on both township and village level of enterprises by industry (Table 2). Although the data provide more industry details than the NBS statistics but they cannot be reconciled with the latter. The MoA source suggests much larger size of employment by rural enterprises. It is needless to mention that NBS is more specialized than MoA in data collection and compilation. On the other hand, the MoA system (its regional offices and local agencies) is very likely to exaggerate the employment of rural enterprise because promoting rural industrialization was one of the main tasks of MoA assigned by the central government. Considering this, we use the employment structure derived from the data rather than accepting its numbers employed, and for 1995 we replace the MoA data by the 1995 industrial census data on the employment structure of village enterprises. For the period 1998-2000, we simply accept the structure of 1997. Finally, for people who were not engaged in village enterprises but worked in family business or simply as self-employed, we assume their industrial distribution to be the same as that of village enterprises.

After this gap-filling exercise, we have constructed the basic data for the entire period in 24 industrial sectors with the consistent standard of industrial classification (grouped based on the 1994 CSIC), and distinguished by three groups of ownership types or three layers, namely, the state layer ("hard core"), the "township layer" covering all non-state employment at the township or above level, and the "outer layer" containing all other ownership types. The so-constructed data are adjusted to mid-year (average) and reported in Appendix Table A1, which are ready for the adjustment for the coverage problems.

Adjustment for coverage problems

As discussed in Section 2, China's industrial employment statistics are exaggerated because they inappropriately includes employees engaged in non-industrial activities, classified in industrial labor statistics as "employees engaged in services" and "other employees". The first category includes people engaged in services provided by enterprise-run education units, medical clinics, child care centres and commercial

outlets, as well as social or political organisations attached to enterprises. The second category includes various types of staff and workers who are attached to but not working for the factor/enterprise that hire them.[17] We label all these people as "non-industrial workers" who should be removed from the current industrial employment statistics. Since these problems are typically a state-enterprise phenomenon, and to less extent observed in urban collective enterprises and some township government-involved or controlled enterprises, our adjustment for the coverage problem should be ownership-specific.

For this purpose, we need data on employment by occupation at the CSIC two-digit level of industrial details, cross-classified by ownership type (referring to each of the three layers). Strictly speaking, only comprehensive labor survey or census can satisfy such a data need. Fortunately, time series is not a basic data requirement. This is because institutional factors, which cause the Chinese industrial enterprises to run community, social and personal services, to accommodate the organs of political organizations, and to keep off-post employees in payrolls, do not change in the short term. Given the institutional settings, in any industry production technology should be the major factor determining the occupation structure of employment because it affects factor intensity, firm size and hence managerial structure. Since major technological change does not take place in the short term, one should not expect frequent changes of occupation structure in any industry. This means that we only need a few benchmarks that can reflect the change of occupation structure in individual industries over the entire period, and hence help anchor the interpolations between the benchmarks.

China's 1985 and 1995 industrial censuses can approximately meet our data requirement. The two benchmarks should be sufficient for the reform period. For the pre-reform period, we rely on some newly found information from the DPES *Annual Bulletins* for 1955 and 1959-64 that were survived from the Cultural Revolution (1966-76). Constrained to the level of industrial details of the available occupation data, we can only use four broad occupation categories, namely, "workers",

[17] As explained in various NBS documents, the category of "other employees" includes workers and staff who are on a factory's payroll but have stopped working for the enterprise, including those engaged in farming activities (attached to a factory), in long-term study leave, in government assignment (outside the routine work of the factory), and on industrial injury or long sick leave, and those off-post workers (*de facto* unemployed) (see NBS and MoL, 1994, p.20; DPSSTS, 1998, p.66).

"technicians", "managerial and administrative staff", and "non-industrial employees". Note that the last category includes the "employees engaged in services" and the employees defined as "others" by DPES.

We set up five benchmarks for the state sector (the "hard core"), that is, 1955, 1963, 1985, 1995 and 2000. The two census-based benchmarks, 1985 and 1995, are reconciled for classification consistency. In fact, the only approximately compatible benchmark for the period prior to 1985 is 1963. We find that compared with earlier statistics (data for 1955 as an example), the available 1959-64 issues of the DPES *Annual Bulletin* report increasingly more details of occupation. This is largely because during the policy retreat after the collapse of the GLF campaign, the government wanted to identify "non-productive workers" in cities and repatriate them to the countryside (Wang and Dong, 1995, pp.113-116). Of the 1959-64 issues, the 1963 issue (DPES, 1963, pp.38-49) gives the most industry details than other issues and then serves as a compatible benchmark with 1985. The gap between 1963 and 1985 should not be a big problem because there was no radical industrial policy change before the market oriented industrial reform in 1984 and, especially, during the ten years of Cultural Revolution.

However, the 1955 issue of the *Annual Bulletin* (DPES, 1955, pp.133-135) gives much less industry details of occupation than what this study requires (10 versus 24 as given in Table 1). We estimate the occupation structure for 1955 with the information available from the 1963 benchmark as the reference. In the estimation, we first re-group the 1963 data into the ten (larger) sectors to match the 1955 data and calculate the ratio of sub-sector to sector total for each sector. We then use the ratio to estimate the sub-sector occupation structure for 1955 assuming that the "relationship" between sector and its sub-sectors in occupation structure in 1955 was the same as that in 1963.

As for the 2000 benchmark, there is even less information available. The official labor statistics only report the number of technicians in three broad sectors, i.e. mining, manufacturing and utilities (DPSSTS, 2001, p.160). However, there is slightly more information available for 1997 when all four occupation categories of employment were reported for the same three sectors (DPSSTS, 1998, pp.211-219). To estimate the occupation structure for 2000 we use the same approach as we did for 1955 but have to work on much broader sectors. More precisely, our reference for the

ratio of "technicians" to each of the other occupation categories is based on the 1997 DPSSTS data (3 sectors) and our reference for the ratio of sub-sector to sector total is based on data from the 1995 industrial census (24 sectors), assuming all these ratios were held true for 2000.

TABLE 4 (A)
OCCUPATION STRUCTURE IN STATE INDUSTRIAL EMPLOYMENT (THE "HARD CORE"), 1963, 1985 AND 1995, BY INDUSTRIAL SECTOR
(Total employment = 100)

	1963				1985				1995			
	W	T	A	N	W	T	A	N	W	T	A	N
	A: The State "Staff and Workers" (the "Hard Core")											
All	75.3	4.4	9.6	10.6	70.3	4.1	10.7	14.9	62.0	6.0	10.2	21.8
1	76.2	3.1	7.8	12.8	69.1	1.6	7.5	21.8	56.1	2.4	6.8	34.8
2	68.7	8.0	12.1	11.2	61.5	4.1	12.9	21.5	52.0	8.5	11.8	27.8
3	70.2	4.3	8.8	16.8	67.3	3.1	10.2	19.3	59.6	4.7	10.1	25.6
4	72.6	3.1	10.1	14.2	70.2	1.9	11.9	15.9	64.9	4.4	11.3	19.5
5	79.0	1.3	11.8	7.9	79.8	1.9	9.9	8.4	66.2	5.5	11.2	17.1
6	82.7	1.4	7.6	8.3	80.8	1.2	9.2	8.8	68.5	6.2	11.1	14.2
7	81.3	2.6	6.6	9.4	80.6	1.7	7.4	10.3	68.7	3.1	6.5	21.7
8	82.4	1.7	8.1	7.8	78.8	1.6	7.7	11.9	69.6	3.3	9.4	17.7
9	83.4	1.6	8.5	6.5	78.8	1.5	10.0	9.7	66.7	3.0	8.6	21.7
10	79.3	1.7	10.1	8.9	74.3	2.3	11.1	12.3	62.1	4.1	11.2	22.6
11	78.5	3.0	9.0	9.6	76.4	2.2	10.4	11.0	68.2	4.3	10.6	16.9
12	70.4	6.1	8.9	14.5	61.5	6.1	12.1	20.3	54.9	11.1	12.5	21.6
13	73.8	5.0	10.4	10.8	69.2	5.0	11.9	14.0	63.3	7.9	11.1	17.7
14	78.4	3.5	9.4	8.7	74.6	3.2	11.6	10.6	62.5	5.0	9.9	22.6
15	78.4	2.0	9.9	9.6	74.6	2.7	10.6	12.0	68.0	4.9	10.8	16.3
16	71.8	5.1	8.3	14.8	69.4	3.8	10.4	16.4	64.3	5.6	9.6	20.5
17	77.3	4.1	10.7	7.9	71.8	3.8	13.0	11.4	59.7	5.3	11.8	23.2
18	70.8	9.1	10.8	9.2	68.8	5.8	12.2	13.2	58.5	7.1	11.7	22.6
19	74.2	8.0	9.1	8.7	65.4	6.9	12.9	14.8	58.4	8.8	11.8	21.1
20	70.8	9.1	10.8	9.2	69.4	6.0	12.8	11.8	60.6	7.6	11.8	20.0
21	70.8	9.1	10.8	9.2	65.3	10.1	12.8	11.8	51.7	12.5	10.8	25.1
22	70.7	9.8	11.8	7.7	64.0	10.0	13.6	12.4	50.9	9.5	11.7	27.9
23	81.4	1.3	10.2	7.0	76.2	2.6	12.0	9.2	63.3	6.8	10.6	19.3
24	72.2	6.4	12.3	9.2	67.7	5.3	13.0	14.0	64.7	10.0	12.4	12.9

Sources: Authors' calculation based on data from DPES (1963, pp. 38-49), NICLG (1988, V.3, pp. 546-561) and NICO (1997, Ownership Volume, pp. 168-203; Sector Volume, pp. 201-233). See the text for the details of the calculation.

Notes: Refer to Table 1 for the code of industrial sectors. W: workers, T: technicians, A: administrative staff, N: Non-industrial employees. See text for the details of N.

There is additional information that can be used for the interpolation between the benchmarks, that is, the number of employment for each occupation (as defined in this study) of the state industry as a whole in 1952-87 (DSS, 1987, p. 39; 1989, p. 49). This information is important because it gives the actual annual changes in different occupations over this period, even if it has no sectoral details. To incorporate it in the interpolation, we first interpolate the share of each occupation category by sector between the benchmarks, based on which we can calculate the ratio of sector to the industrial total for each occupation. Then, assuming the ratio is held for the actual industrial total as reported by DSS, we replace the interpolated share of each

occupation category in the total industry by the same share obtained from the DSS annual aggregates to estimate the actual total-adjusted occupation structure for each sector.[18] Note that to fully use the DSS annual aggregates, we extend the 1955 benchmark back to 1952 by assuming that the share of "workers" was 80 percent rather than 73 percent in 1955 and adjusting other occupation shares accordingly. For the period 1949-51, we simply adopt the estimated occupation shares for 1952. For the period 1988-2000, we conduct interpolation between the benchmarks and then adjust the results for 1988-91 because there is additional information available on the number of technicians by sector.

For the township layer that includes all non-state "staff and workers" (the DITS concept), only the 1985 and 1995 industrial censuses can provide the required data. To construct the time series for the entire period, we need at least one benchmark at the early time and one at the end of the period. For the earlier benchmark, we also choose 1963. We estimate the 1963 benchmark by assuming that the share of "workers" in the total employment increased to 90 percent from 80 percent in 1985 and adjusting the other occupation shares in the total and at sector level accordingly. This assumption is based on two observations: first, there was a rising trend in the share of service staff between by the two censuses, and second, the authorities launched a campaign in 1962-63 to cut "non-productive" staff and workers (see for example (SC/DRC, 2000, V.4 (II), pp. 537-540).

[18] We denote the share of the *j*th occupation in the total employment of a benchmark year *T* (in our case, T = 1955, 63, 85) that is obtained from the census/survey data as φ_j^T and the same share but obtained from other source without sector details as $\tilde{\varphi}_j^T$ (in our case the DSS 1952-87 series of the total state industry). Assuming that the relationship $\varphi_{jk}^T / \varphi_j^T = \tilde{\varphi}_{jk}^T / \tilde{\varphi}_j^T$ is held for the *j*th occupation of the *k*th industrial sector, then $\tilde{\varphi}_{jk}^T = \tilde{\varphi}_j^T (\varphi_{jk}^T / \varphi_j^T)$ must also be held. To estimate the share of the *j*th occupation in the *k*th sector between the benchmark years, we conduct interpolation between the benchmarks using the available census/survey data and obtain φ_{jk}^τ, where the superscript τ stands for any time point between the benchmark *T*. Then we adjust the results by the share of the *j*th occupation in the total employment from the other source (the DSS series) by $\tilde{\varphi}_{jk}^\tau = \tilde{\varphi}_j^\tau (\varphi_{jk}^\tau / \varphi_j^\tau)$.

TABLE 4 (B)

OCCUPATION STRUCTURE IN NON-STATE INDUSTRIAL EMPLOYMENT (THE "TOWNSHIP LAYER"), 1963, 1985 AND 1995, BY INDUSTRIAL SECTOR

(Total employment = 100)

	1963				1985				1995			
	W	T	A	N	W	T	A	N	W	T	A	N
	B: The Non-state "Staff and Workers" (the "Township Layer")											
All	90.0	0.9	5.2	3.9	79.6	1.6	10.8	8.0	73.4	5.4	10.0	11.2
1	89.4	0.6	4.1	6.0	79.1	1.2	8.0	11.7	77.8	3.6	7.2	11.4
2	77.0	2.0	8.6	12.4	68.1	2.8	11.9	17.2	56.9	17.9	15.2	10.0
3	89.7	0.9	4.6	4.9	79.3	1.8	9.2	9.7	77.0	4.8	8.6	9.6
4	93.8	0.2	3.2	2.9	83.0	0.5	8.7	7.8	79.6	4.3	8.7	7.4
5	89.3	0.5	6.3	3.9	79.0	1.0	12.4	7.6	71.5	6.4	12.3	9.8
6	93.6	0.3	3.7	2.4	82.8	0.8	9.9	6.5	72.1	5.4	12.5	10.0
7	93.3	0.4	3.3	3.0	82.6	1.0	8.6	7.7	77.3	4.0	7.8	10.9
8	96.6	0.1	2.1	1.2	85.5	0.4	8.9	5.1	79.8	3.4	8.2	8.5
9	94.2	0.2	3.5	2.1	83.3	0.7	10.1	5.9	78.8	3.3	8.7	9.2
10	91.5	0.3	4.8	3.4	81.0	0.7	10.8	7.6	74.1	4.8	9.9	11.2
11	91.9	0.4	5.0	2.7	81.3	0.9	11.5	6.3	73.8	4.9	10.8	10.5
12	86.7	1.4	6.4	5.5	76.7	2.4	11.3	9.6	70.2	6.7	12.0	11.2
13	85.4	1.7	7.3	5.6	75.6	2.8	12.3	9.3	68.3	7.3	12.2	12.2
14	90.8	0.6	5.4	3.3	80.3	1.3	11.5	7.0	72.1	4.6	10.6	12.7
15	94.9	0.3	3.0	1.8	84.0	0.9	9.5	5.6	77.6	5.3	8.9	8.1
16	87.3	1.3	5.8	5.6	77.2	2.3	10.4	10.1	73.5	5.2	9.9	11.4
17	86.8	0.7	6.4	6.1	76.8	1.3	11.2	10.7	70.6	5.6	10.8	13.1
18	86.1	1.7	7.2	5.0	76.2	2.9	12.4	8.5	68.1	6.8	11.5	13.5
19	85.9	1.5	7.4	5.2	76.0	2.6	12.6	8.8	68.0	6.8	11.8	13.4
20	86.9	1.5	7.1	4.5	76.9	2.6	12.5	8.0	67.2	6.7	11.6	14.5
21	81.5	3.8	8.4	6.3	72.1	5.7	12.6	9.5	67.9	9.2	10.1	12.7
22	79.6	3.9	9.6	7.0	70.4	5.6	13.8	10.1	62.3	7.7	11.5	18.5
23	95.3	0.2	2.9	1.5	84.3	0.8	9.7	5.1	76.0	3.8	8.3	11.9
24	81.1	2.0	10.1	6.8	71.7	3.0	15.0	10.2	66.4	11.1	13.6	8.8

Sources and *Notes*: See Table 4 (A).

For the period prior to 1963, firstly we assume that the occupation structure in the period 1953-62 were the same as that in 1963. There was no fundamental institutional change during this period. In technical terms, we however implicitly assume that the underlying capital-labor ratio remained unchanged over the period. This is not a very strong assumption given that industries at this level did not undergo significant technical advance. Secondly, for the period 1949-52, we assume that industrial firms did not hire "non-industrial employees" because there was no state control at this level and China had not yet adopted the central planning system. We then adjust the share of other occupations accordingly.

For the period between the 1985 and 1995 census-based benchmarks, we conduct simple interpolation to fill the gap. As for the period after 1995, we first estimate the 2000 benchmark using the same approach as we did for the state sector because the same sources also provide occupation information for collective firms (DPSSTS, 2001, p.160; DPSSTS, 1998, pp.211-219), and then interpolate the data between the so-estimated 2000 benchmark and the 1995 census-based benchmark.

So far, with the available information and assumptions, we have constructed the occupation structure by sector for both the state industry and the non-state industry at or above the township level for the entire period 1949-2000. This result is used to remove the inappropriately included "non-industrial employees" in both the "hard core" and the "township layer".

As we pointed out earlier, since the coverage problem discussed in this section is typically a state-enterprise phenomenon and to less extent observed in some urban collective enterprises and township enterprises, there is no need to adjust the rest of the employment (within the "outer layer") for this problem. We feel justified to assume that industries in this layer are highly labor intensive. We thus assume, though arbitrarily, that all sectors in this layer have the same occupation structure as that of the sector in the township layer which has the highest share of "workers", adjusted by dropping the "non-industrial employees" in that sector. To match the other layers, we also set up three benchmarks, i.e. 1963, 1985, 1995 and 2004.[19] The occupation structures between 1985 and 1995 are interpolated, and the occupational structures for the period 1949-62 and 1996-2009 are assumed the same as that of 1963 and 1995, respectively. Based on the so-constructed occupation structure for each industrial sector of the state enterprises and the non-state enterprise at the township layer, we can remove the "non-industrial employees" from the employment data.

If denoting employment matrix as **N**, we can now have a time series of marginal employment matrix $\mathbf{N}^{\mathrm{M}}$ with numbers employed cross-classified by 3 occupations (*j*), 3 ownership types (*o*) and 24 industrial sectors (*s*), which can be defined as $\mathbf{N}^{M}_{s+++jot} = \sum_{g=1}^{2}\sum_{a=1}^{7}\sum_{e=1}^{5}\mathbf{N}_{sgaejot}$, where the symbol + in the subscript stands for the missing dimension that is implicitly summed up. Therefore, the total number of cells for the period 1949-2009 is $\mathbf{N}^{M}_{s+++jot} = 24\times 3\times 3\times 61 = 13176$, that is 216 cells for each year. However, if removing the occupation dimension, the matrix becomes $\mathbf{N}^{M}_{s+++ot} = \sum_{g=1}^{2}\sum_{a=1}^{7}\sum_{e=1}^{5}\mathbf{N}_{sgaeot}$ and hence the total number of cells for the same

[19] Our assumption is based on observations of the most labour intensive industries, typically the apparel industry, in the "township layer", assuming without "non-industrial employees". We set the shares of "workers", "technicians" and "administrative staff" as 92.5, 0.5 and 7.0 percent, respectively, for 1963, 91.0, 0.5 and 8.5 percent for 1985, and 90.0, 2.5 and 7.5 percent for 1995.

period is $\mathbf{N}^{M}_{s+++ot} = 24 \times 3 \times 61 = 4392$, with 72 for each year. It becomes more manageable and may be subject to less calculation errors caused by unrealistic assumptions. Furthermore, to show the effect of the state ownership on labor input, we also consider a matrix without the ownership type dimension for a comparison, that is, $\mathbf{N}^{M}_{s+++t} = \sum_{g=1}^{2} \sum_{a=1}^{7} \sum_{e=1}^{5} \mathbf{N}_{sgaet}$.

The so constructed employment data are now ready for the last step, that is, converting the numbers employed to the hours worked.

Numbers employed converted to hours worked

In the absence of systematic records of any kind on standard working hours across industries and over time, one may think that the exercise of converting numbers employed into hours worked is inevitably arbitrary. However, official documents and studies do suggest that not only has the standard of working hours by law changed overtime, but also different standards were adopted by industries at the same time. Besides, instead of following the official standard of working hours, non-state firms and self-employed people tend to adjust their working hours in response to current demand and supply conditions. In this section, our work will focus on how to set up different standards in the hour-number conversion exercise for different industries, ownership types and periods.

It is important to treat state and non-state industries differently in this exercise. It is a long tradition since the central planning period that state enterprises and government sectors are integrated in human resource management. Institutionally, enterprise managers, administrative personnel, technicians and staff who represent political and social organizations receive equivalent ranks to those working in government offices. Regulations on labor compensation and welfare including working hours are strictly implemented in state enterprises as in the government sector. In fact, information on working hours disclosed in studies and official documents only refers to state enterprises.

Let’s begin with the state sector. To set up the standard number of working weeks per year for state enterprises, we need to subtract the number of public holidays from a calendar year. China maintained a seven-day public holiday system from 1949 to September 1999 and afterwards increased the number of public holidays

to ten. If using 365 calendar days as a standard, the non-public holiday calendar contains 51.1 weeks for the period from 1949 up to September 1999 and 50.7 weeks afterwards. In this study, we adopt a standard of 51 working weeks per calendar year.

It is commonly believed that the People's Republic implemented the 8-hour working day system from the right beginning of the new regime. This is true in principle. But, in reality, this was not strictly followed until China's first constitution passed in 1954. Prior to 1954, as required by Article 32 of the *Common Guiding Principles of the Chinese People's Political Consultative Conference* (CPPCC), which was a provisional constitution and inaugurated by the CPPCC's First Plenary Meeting in September 1949, all state and private enterprises should in general limit their working time to 8-10 hours per day (Zhu, 1999, p. 391). We therefore assume that the number of average working hours per day was 9 and the number of working hours per week was 54 in 1949-53. The 8-hour working day or 48-hour working week system was legally implemented in 1954 and had since been maintained till May 1994 when the number of weekly working hours was reduced to 44, equivalent to 5.5 working days. One year later, in May 1995 a new standard of 40 hours per week was adopted, equivalent to 5 working days (DPSSTS, 1998, p. 332; Bai, 2002, p. 409). This standard defines China's institutional working hours and will be used as our *baseline* (Table 5) for measuring working hours across different industrial sectors as explained below.

However, individual industries are allowed to adopt different working-hour standards that are less than what defined by the baseline. Technically, the production process of some industries such as steel, chemical, and power generation must not be interrupted. Therefore, their workers have to work shifts to ensure the continuation of production. But some industries adopt the shift system for non-technical reason. For example, when resources are in short supply mining industries work shifts to meet the high and increasing demand by manufacturing industries. In the central planning period pursuing heavy industrialization, industries engaged in the manufacturing of consumer goods such as textiles suffered from insufficient investment. To meet the demand, they had to adopt the shift system so that the existing equipment could be fully used.

In the absence of (sufficient) compensation for the negative physical impact of the shift system on workers, especially in industries with unfavourable working

conditions, the government adopted variations of the shift system in different industries to reduce the number of working hours while maintaining the full utilization of equipment. Typically, in industries such as textiles, the old system of "three shifts" was replaced by a system of "three shifts by four groups" in the early 1960s, under which workers work 6 hours less per week than the baseline standard. In industries with harsher conditions such as mining and oil extraction, a "four shifts" system was implemented under which workers only worked for 6 hours per day or 36 hours per week (Zhu, 1999, pp. 438-444). Consequently, as given in Table 5 there are three types of industries adopting different standards of weekly working hours, namely, the A type following the baseline, the B type with 6 hours less than the baseline and the C type with 12 hours less than the baseline. Whenever there was an institutional change in the baseline, we adjust the standard of the B and C types accordingly. Besides, given the rising pressure on resource industries (the B* type), we impose some *ad hoc* upward adjustment by 4 hours a week to these industries for the period since 1993. However, we only apply the standards for the B-type (and B*-type) and C-type industries to "workers" and "technicians". In other words, we assume that "administrative staff" and "non-industrial employees" in these industries follow the baseline standard.

TABLE 5
"STANDARDS" OF WORKING HOURS PER WEEK FOR WORKERS AND TECHNICIANS BY OWNERSHIP TYPE

	1949-1953	1954-1957	1958-1993	1994	1995-2009
State: A (the baseline)	54	48	48	44	40
B	54	48	42	38	34
B*	54	48	42 (-1992)	46 (1993)	46 (1994-)
C	54	48	36 (-1992)	40 (1993)	40 (1994-)

	1949-1953	1954-1984	1985-1992	1993-2009
"Township Layer" (all industrial sectors)	54	48	52	56

	1949-1957	1958-1959	1960-1984	1985-1992	1993-2009
"Outer Layer" (all industrial sectors)	31.2 (65% of 48)	48	31.2 (65% of 48)	38.4 (80% of 48)	48

Sources: See the text for detailed information and adjustment.
Notes: Type of industries: A=5, 8, 9, 10, 17-23; B=6, 7, 11, 13, 14; B*=12, 15, 16, 24; C=1-4 (see Table 1 for the sector codes). See text for the details of the standards and adjustments.

Let us now turn to the case of the non-state employment in the "township layer". We have treated all the employment in this layer indifferently although they include various types of non-state enterprises such as urban collectives, joint ownerships,

foreign invested firms and rural commune/township enterprises. A common feature of these enterprises is their heavy engagement in labor intensive industries. However, taking into account the changes in the policy regime and the composition of enterprises, we should treat the number-hour conversion in the pre-reform period, specifically, before the full-scale industrial reform began in 1985, and the post-reform period differently. During the central planning period, urban collective enterprises were integrated with the state industrial system. Therefore, it is reasonable to consider them as semi-state enterprises that followed the government's labor regulations. On the other hand, the rural commune/township factories which survived from the Great Leap Forward were also controlled by local governments because there was no market system for their inputs and outputs. We then assume that in the pre-reform period, they all followed the baseline standard of working hours and there were no difference between occupation groups and across industries (Table 5).

In the reform period, there were two important policy changes: one was the shift from planned to market-oriented industrial development that gave more room for non-state enterprises to grow and the other was the opening up to foreign direct investment, which resulted in rapid growth of export-oriented, labor-intensive industries. As widely observed (but not officially surveyed), driven by increasingly fierce market competition in the situation of abundant labor supply, non-state enterprises tend to ignore working conditions and over use of workers. Therefore, we have two assumptions in setting up the working hour standard for the enterprises in this layer. Firstly, we assume that the non-state enterprises increased their working hours in response to the opportunities emerged following the industrial reform. Therefore, we upward adjust the weekly working hours by 4 or from 48 to 52 hours a week for the period 1985-92. Secondly, we assume that these enterprises did not follow the baseline standard to cut working hours in 1994 and 1995, instead, they further raised the weekly working hours since 1993 along with the government's more liberal measures towards foreign trade and direct investment in 1993.[20] We thus increase the number of working hours of "workers" by 4 to 56 for the period in 1993-2005. It should be noted that this assumption allows 5 percent more than the maximum number of working hours that is allowed by China's *Labor Law* in order to

[20] Following Deng's call for bolder market-oriented reforms during his to the southern China from 18 January to 21 February, 1992, the 14th National Congress of the communist Party of China decided to promote "socialist market economy" (Bai, 2002, pp. 309 and 323-324).

capture the effect of the over work of workers.[21] For other occupation categories, the standard remains the same as that used for the period 1985-1992.

For those classified as industrial laborers in the "outer layer", consisting of the employment in village (production brigade) factories, self-employed, and people in private firms, since many engage in farm sideline activities or activities that were to fill the idle season in farming, and gradually move to labor intensive activities, we assume that they generally do not work full time year round, but their working hours have increased since the 1990s in line with the change of the overall economy. We adopt a seasonal adjustment that discounts the baseline standard by 35 percent for the period 1949-84; that is, about four months of the year of the industrial labor force in this layer were spent on farm activities or just idle. The only exception is the time in 1958-59 for which we assume that all types of laborers were fully used driven by the rural industrialisation policy during the Great Leap Forward campaign. After the industrial reform, we assume that the idle time or the time spent on farming reduced from 35 to 20 percent; therefore, we adjust the number of working hours to 80 percent of the baseline standard for the period 1985-92. For the period 1993-2005, because of the accelerated marketization we assume 48 per week as given by the pre-1994 baseline standard. Table 5 also lists the standard of working hours for this layer.

Based on the working hour standards in Table 5, we convert the numbers employed cross classified by 3 occupations (*j*), 3 ownership type (*o*) and 24 industrial sector (*s*) into the hours worked, which can be defined as a time series of marginal employment matrices in hours as $\mathbf{H}^{M}_{s+++jot}$ for the period 1949-2005. The results will serve as the "control totals" in the construction of the full-dimension labor employment and compensation matrices.

4.2 Construction of Compensation Data (Marginal Matrices)

Official data on labor compensation and problems

The Construction of labor compensation data is even a bigger challenge to us. Most of the required data, especially compensation by human capital attribute, do not exist in Chinese official statistics. The available official data are aggregate wages paid to

[21] As required by the Labour Law inaugurated in 1995, the standard or the institutional working hours per week are 44 (Article 36) and the maximum additional working hours should not exceed 36 per month (Article 41). This means that the maximum working hours should be 53 per week. (There are reports on migrant labor conditions, though anecdotal, especially extremely long working hours).

employees at sectoral level with ambiguous definition and insufficient coverage. It is necessary to begin with an understanding of the available data and related problems so that we can highlight the key tasks in our data work.

There are two basic indicators of labor compensation in the official labor statistics constructed by DPES, namely, "total wage bill" (*gongzi zonge*) and "total insurance and welfare payment" (*baoxian fuli feiyong zonge*), covering only the "staff and workers" in the urban/industrial sector as defined by DPES (see Section 4.1). Statistics for the two indicators are collected through the reporting system, supported increasingly by sample surveys since the reform, and internally reported in the DPES quarterly and annually bulletins, based on which the finalized statistics, especially for "total wage bill", are published in *China Labor Statistical Yearbook* from 1988.[22] A closer examination of the two indicators helps understand their origin and nature. Traditionally, "total wage bill" was one of a few key indicators used by the planning authorities to monitor and control state enterprises and urban collective firms (semi-state), so called "total wage bill management". The planning authorities assigned "total wage bill" to enterprises estimated based on their current-year performance and next-year production plans. Enterprises were not allowed to pay more than the assigned limit. On the top of the "total wage bill", enterprises were also allowed to make "total insurance and welfare payment" to their employees but it should be equivalent to a maximum 14 percent of its "total wage bill". Therefore, under the "total wage bill management" labor compensation was conceptually a planned spending quota on labor rather than actual payment to labor. Of course, during the central planning period plans were usually "fulfilled" and the planned payment was equal to the actual payment because there was no leeway for enterprises to pay more to their employees than the state set standard wages or to hire more workers that was not as planned. This is, however, no longer the case since the reform.

Along with the reform, income incentives have played an increasingly important role when enterprises have to respond to market competitions including competition for labor. In 1985, a reform to the existing wage system that began in 1955 was introduced to large and medium-sized state enterprises linking "total wage bill" with total enterprise tax payment (including turn over profits), under which the growth of

[22] Historical data for 1949-1985 are published by the Department of Social Statistics (DSS), the predecessor of DPES, in 1987 and updated for 1978-87 also by DSS in 1989.

the former was not allowed to be faster than that of the latter, usually 30 to 70 percent of the latter (State Council, 1985a). To effectively curb state enterprises' incentives of maximizing labor income—a typical problem with ambiguous property rights, the authorities resorted to tax incentives. As required in two subsequent regulations issued following the wage reform, an enterprise's total bonus payment (part of its wage bill) would be subject to a progressive tax from 30 to 300 percent if it exceeded an amount equivalent to four-month standard wage bill and its total wage payment would also be charged by a wage adjustment tax up to 300 percent if it exceeded the approved "total wage bill" by more than 7 percent (State Council, 1985b and 1985c).

The problem was two-folded. In cases that the assigned "total wage bill" was insufficient, enterprises had to hide some labor compensation as non-labor cost, whereas in cases that the assigned "total wage bill" was more than actually needed, enterprises were either encouraged to overpay employees or to make inappropriate expenses under the name of labor compensation. Coherently, "total insurance and welfare payment", which was still limited to an equivalent of 14 percent of the "total wage bill", gave the same distorted incentives to enterprises. Instead of abandoning the "total wage bill management", the statistical authorities focused on improving statistics on actual wage paid aiming to align the "total wage bill" assignment with the reality. After various surveys by NBS since 1986, a revised version of the 1955 *Regulations on the Structure Total Wage Bill* was issued in 1990 (NBS, 1990, Decree No. 1), supplemented by another NBS regulation on the structure of total labor compensation including both wage and non-wage incomes in 1992 (NBS, 1992, Circular No. 257). These efforts were largely in vain because the planning authorities enhanced the "total wage bill management" in order to manage the unprecedented expansion of aggregate demand. Consequently, the taxable wage standard, which was a key variable in managing the "total wage bill", could not timely reflect changes in the cost of labor along with the marketisation of the economy.[23]

There are another two problems in the DPES labor statistics. First, data on "total wage bill" do not cover employees who are not classified as "staff and workers" in the

[23] The tax threshold was raised to 550 yuan in 1996 after a 500 yuan standard introduced since the early 1990s. In 1999 it was further raised to 800 yuan with a local floating band up to 20 percent of the baseline. This standard however remained intact till 2006 when a further adjustment increased it to 1600 yuan with the termination of local floating adjustment (STA, 2006).

urban sector (up to 2000)[24] and they also exclude rural enterprises completely. Second, data on "total insurance & welfare payment" do not have industry breakdowns that match the "total wage bill".

There are another two sources of labor compensation data provided by NBS, both with a full coverage. They are industrial censuses and national input-output tables. The Department of Industrial and Transportation Statistics (DITS) is responsible for conducting the 1985 and 1995 National Industrial Censuses as well as the industrial part of the 2004 Economic Census. The censuses provide labor compensation data for all enterprises, both urban and rural, and family based or self-employed activities, but with more details on enterprises at or above the township level/designated size. However, the enterprise data still categorized under "total wage bill" and "total insurance & welfare payment" that are in line with their regular accounting and reporting practice with DPES. Therefore, while coming with a wider coverage than the DPES annual data, the census data do not solve the conceptual problems in the DPES indicators.

The Department of National Accounts (DNA) is responsible for constructing the national input-output tables with one full table in every five years since 1987 and one reduced table between two full tables. DNA in principle follows UN's System of National Accounts (SNA) and has an integrated view on statistics from all specialized NBS departments including DPES, which provides it a better position to find and fix problems. In addition to the data provided by the specialized departments, DNA conducts its own input-output surveys to fill data gaps and missing parameters required for constructing the full tables. One of the main tasks in the input-output survey is to repair the DPES labor compensation data. Through crosschecking the input-output survey results with the annual DPES statistics it attempts to identify labor compensation items that are reported or disguised as others and adjusts them for measuring labor compensation in the input-output tables. Given the shortcomings in the "total wage bill management" system and hence the distorted incentives of enterprises in reporting data, what DNA can do is still limited. However, their labor

[24] In the 1994 *China Labour Statistical Yearbook* (p. 591), DPES for the first time introduced a wider concept "labour compensation to persons employed" (*congye renyuan laodong baochou*) that consists of two parts: 1) "total wage bills" paid to "staff and workers" and 2) "labour compensation" paid to others employed. This change also appeared in a publication by NBS and Ministry of Labour (NBS-MOL, 1994, p. 25). However, the first data related to this are published in *China Labour Statistical Yearbook* from the 2001 issue onwards without industry details.

compensation data are the best available and should be used as our benchmark control totals in constructing labor compensation time series for the post reform period.

Table 6 summarizes the features of the available data and required data work in constructing the compensation time series that matches numbers employed by industry and by ownership types, that is, $\mathbf{N}^{M}_{s++++ot}$.

TABLE 6
AVAILABLE OFFICIAL DATA ON LABOR COMPENSATION AND THE REQUIRED DATA WORK IN THIS STUDY

Available Data	Required Data Work
DNA/Input-Output Table "Labor Compensation": • 2-digit level industries, no ownership breakdown • One full table every five years and one reduced table in between, 1987-2007 (CSIC-inconsistent)	• Industry classification adjustment • Reconciliation and interpolation between full and reduced tables • Used as benchmark control totals at industry level for constructing the post-1987 series
DITS/Census "Total Wage Bill" and "Total Insurance & Welfare Payment": • The 1985 and 1995 Industrial Censuses, major ownership types, 3/4-digit level, some indicators back to 1980 • The 2004 Economic Census, major ownership types, 3/4-digit level industries	• For "total wage bill", used as much more detailed benchmarks that help reconcile different CSICs and decompose non-state aggregates to 2-digit level as well as check and adjust annual data • The ratio of "total wage bill" to "total insurance & welfare payment" is used to estimate welfare payment at industry level by ownership types • Data on 1980 reported in the 1985 Census are used for industrial and ownership structures in central planning
DPES[a]: • State "Total Wage Bill", 2-digit level industries, 1952-84, 1978-87, 1988-2009 (CSIC-inconsistent) • State "Total Insurance & Welfare Payment", industrial aggregate, 1959-64, 1975-85; mining, manufacturing, utilities, 1993-99	• Industry classification adjustment • Together with census data, estimate industry level welfare payment for benchmarks and fill the gaps • Average wage used for making assumptions for industries in other ownership types
DPES[a] "Total Wage Bill": • Urban collectives, industrial aggregate, 1952-84, 1978-87, 1988-2009 • Other ownership types (including FDI) industrial aggregate 1984-87, 1988-2009	• Together with census data, especially ratios of state to non-state in census, estimate industry level wage and welfare payment for benchmarks • Interpolations between benchmarks

Sources: See the text for adjustments and references.
Notes: a) DSS as predecessor and DPSSTS as successor.

Data construction by sector and ownership type

Based on the available data presented in Table 6 and their problems discussed above, we take six steps to construct the sector (24) by ownership types (3) labor

compensation marginal matrix for each year of the entire period. In what follows, we describe the procedures of the data construction for each step.

The state sector

Constructing the state sector component is crucial, especially for the central planning period or in the case of industrial reform for the period prior to 1985, because this is the only sector for which the best data available though still far from satisfaction. Since the available DPES compensation data refer to numbers employed and are classified in the 1972 CSIC, an industrial classification that is typically designed for vertical administrative control over resource allocation, to keep the compensation-number matching we need to re-classify the numbers employed in the state industry in 39 sectors into 11 two-digit sectors, with the "residual" further decomposed into 4 sectors. We then assume that sub-groups in each of these sectors have the same wage rate, which is not a strong assumption particularly for the central planning period, and allocate the two-digit level "total wage bills" into corresponding sub-sectors covering for the period 1952-87 (DSS, 1987, pp. 124-5; 1989, pp. 166-7). For 1949-51, we assume the "total wage bill" grew at the same rate of the total industrial employment and the total wage is decomposed by the industrial structure of 1952. As for the rest of the state series, for the period 2001-05 (DPSSTS, 2006), direct data are available, which are crosschecked by the 2004 Census. For the period 1988-2000, except for the 1995 Census, when wage data are only available for three broad sectors (mining, manufacturing and utilities), we decompose the sectoral totals with the intra-industry structure of each of the three broad sectors for each year, obtained by interpolations between the structures of 1987 (crosschecked by the 1985 Census data), 1995 (the census) and 2001.

We now need to estimate "total insurance & welfare payment" in the state sector. Due to scant information we have to rely on the census data which give both welfare value and its industrial composition for four benchmarks, i.e. 1980, 1985, 1995 and 2004. We first construct an industry-specific welfare-wage ratio, calculated using the industrial compositions of wage and welfare,[25] for the benchmarks. Interpolations are

[25] This approach imposes the relationship between welfare (F) and wage payment (W) at a benchmark year (*) on other years. Let ω_i be the share of wage payment and ϕ_i be the share of welfare in the ith industry, i.e. $\omega_i = W_i / \sum_{i=1}^{n} W_i$ and $\phi_i = F_i / \sum_{i=1}^{n} F_i$. The welfare-wage ratio, λ, for a

used to derive the series of the ratio between 1980 and 2004. We assume that the 2004 ratio can be used for 2005. The welfare-wage ratio of 1980 is then used to derive the ratio for the period 1949-80. The underlying assumption for the pre-planning 1949-52 may be a bit strong, but we have no choice. Finally, adding the results of the "total insurance & welfare payment" to the estimated "total wage bill" we arrive at the total labor compensation for the state sector.

Urban collectives

In constructing labor compensation for enterprises in the "township layer", we work separately on three components, namely, urban collectives, foreign invested enterprises (FIEs) and the rest as a residual that includes mainly rural enterprises, domestic private enterprises as well as various types of joint ventures among domestic firms. This disaggregating is for two reasons. First, we need different assumptions for these components because financing, market and the institutions determining labor compensation are rather different among them. Second, while no detailed compensation data by industry for any of the component for most of the period, there are data for each component that are further divided into mining, manufacturing and utilities, which should definitely be used in our estimation.

For urban collectives, while there are average wage data and total wage payment for three groups, i.e. mining, manufacturing and utilities, for every year, there are also detailed industry-specific average wage data by industry for seven years, 1989-92 and 2003-05. Our first take is to decompose the group data. To do so we calculate an industry-to-group ratio in average wage for each industry for each of the seven years, e.g. the textiles-to-manufacturing ratio. Assuming that the relationship (the ratio) for 1989 is held for the years prior to 1989, we use the 1989 ratio and the available average wage for the manufacturing as a whole in each year prior to 1989 to estimate the wage rate by industry, technically the same approach as explained in Footnote 26. Besides, we obtain average wage estimates for the period between 1992 and 2002 by interpolation.

Estimation of the welfare payment for urban collectives is based on information from the 1995 and 2004 Censuses. Like the above exercise for the state sector, we

---

benchmark is defined as $\lambda_i^* = \phi_i^* / \omega_i^*$. To obtain the welfare share for the *i*th industry of any other year, we use the following approach, $\phi_{i,t} = \omega_{i,t}\lambda_i^* = \omega_{i,t}(\phi_i^* / \omega_i^*)$.

first calculate the welfare-wage ratio for each industry in urban collectives using the 1995 Census data. Since our calculation shows that the ratio for all collectives is 0.172 compared with 0.336 of the state sector for this benchmark, we then assume a welfare-wage ratio for urban collectives as a whole as half of that of state enterprises for the period prior to 1995. The so-derived annual aggregate ratios are then used to adjust the industry-to-total ratio in 1995 which gives us an estimation of the welfare-wage ratio by industry for each year before 1995. For the period between 1995 and 2004 we again use interpolation and for 2005 we assume the ratio for 2004 is held. Finally, the labor compensation by industry is obtained by multiplying the numbers employed in each industry by an industry-specific compensation rate that is the wage rate adjusted by the estimated welfare-wage ratio.

Foreign invested enterprises

The official statistics on foreign invested enterprises (FIEs) began in 1984. Our approach used to estimate labor compensation for FIEs is the same as we used for urban collectives. The only data with industry details available are average wage from the 1995 Census and from labor statistics in 2003-05. "Total wage bill" and average wage for mining, manufacturing and utilities are available for each year. Based on the industry-to-group ratio in 1995, we estimate industry wage rate for each year prior to 1995 and then complete the series by interpolations between 1995 and 2003. To estimate the welfare-wage ratio, we rely on data the 1995 and 2004 censuses. The 1995 ratio is used to estimate annual ratios for 1984-94. We conduct interpolations between 1995 and 2004 to complete the series, assuming 2005 equal to 2004.

Rural enterprises

The last component of the "township layer", whose records began in 1956, is in fact the "residual" resulting from subtracting urban collectives and FIEs from all non-state enterprises in this layer. We use the same approach to estimate average wage by industry for each year prior to 1995 based on data from the 1995 Census and conduct interpolations between the 1995 and 2004 Census, assuming 2005 equal to 2004. For the welfare-wage ratio, since there is no information and welfare payment is not a standard practice in these enterprises, we assume that it is equivalent to half of that of urban collectives in 1956-92, and then rose to 60 percent in 1993-2000 and to 70 percent in 2001-05 along with the marketisation.

The outer layer

There is no information on labor compensation for this layer except for self-employed family business in the 2004 Census. From the census we know that labor compensation for the self-employed is in general less than that of rural enterprises, or about 65 to 80 percent of the latter (note that there is no distinction between wage and welfare payment for the self-employed). In the estimation, we assume that the relative compensation (i.e. the self-employed-to-rural enterprises ratio) in 2004 is held for the period 2001-05. For the period prior 1949-2000, we assume that laborers in this layer were even less paid but increased over time. Specifically, we assume that the labor compensation to the self-employed is 70 percent of the 2004 ratio over 1949-84 or before the industrial reform, 80 percent in 1985-92, and 90 percent in 1993-2000.

Reconciliation with input-output tables

The final step is to reconcile our results with the national input-output tables. As we argued in the previous discussion that in terms of both concept and practice in measuring labor compensation, the national input-output tables is the only information that could bring us closer to the true labor compensation. DNA has constructed so far four full input-output tables, i.e. 1987, 1992, 1997 and 2002, and also four reduced input-output tables, i.e. 1990, 1995, 2000 and 2005. However, bearing in mind the input-output tables do not provide breakdowns for ownership types as we have been working on that are very important in the case of China.

Our procedures contain four steps. The first step is to make reclassifications among these tables to ensure consistence in industrial classifications in line with the 2002 CSIC. The second step is to make interpolations between the benchmarks to arrive at a time series for 1987-2005. Fortunately, this almost covers the entire period of industrial reform when inconsistence between the controlled and actual labor compensation was more likely to happen. The third step is therefore to calculate the differences between our estimates and the input-output table totals for each industry in each year at the national level. In the last step, we assume that the relative compensations between all five ownership types (instead of three that is finally used) are held and hence used to redistribute the input-output table identified differences based on the relative compensations. This finally gives us a marginal labor compensation matrix cross classified by sector (s = 24) and by ownership type (o = 3),

which can be defined as $\mathbf{C}^{M}_{s++++ot} = \sum_{g=1}^{2}\sum_{a=1}^{7}\sum_{e=1}^{5}\sum_{j=1}^{3}\mathbf{C}^{M}_{sgaejot}$. Therefore, the total number of cells for each year (*t*) is $\mathbf{C}^{M}_{s++++ot} = 24 \times 3 = 72$.

5. CONSTRUCTION OF THE FULL-DIMENSION EMPLOYMENT AND COMPENSATION MATRICES

Our work so far has constructed three long time series of marginal matrices for the period 1949-2005, that is, 1) numbers employed cross-classified by sector (industry), occupation and ownership types, denoted as $\mathbf{N}^{M}_{s+++jot}$, 2) hours worked cross-classified by the same three attributes as in the numbers employed matrices, denoted as $\mathbf{H}^{M}_{s+++jot}$, and 3) labor compensation cross-classified by sector and ownership type, $\mathbf{C}^{M}_{s++++ot}$. These marginal matrices will be used as the "control totals" in constructing the annual full-dimension matrices of the three series, denoted as $\mathbf{N}_{sgaejot}$, $\mathbf{H}_{sgaejot}$ and $\mathbf{C}_{sgaejot}$, respectively. This means that we can therefore derive annual average labor compensation series for each industry cross-classified by ownership type in terms of compensation per laborer employed (*N*) and compensation per hour worked (*H*) as follows,

(12a) $$\mathbf{c}^{M(N)}_{s++++ot} = \frac{\sum_{g}^{2}\sum_{a}^{7}\sum_{e}^{5}\sum_{j}^{3}\mathbf{C}_{sgaejot}}{\sum_{g}^{2}\sum_{a}^{7}\sum_{e}^{5}\sum_{j}^{3}\mathbf{N}_{sgaejot}}$$

and

(12b) $$\mathbf{c}^{M(H)}_{s++++ot} = \frac{\sum_{g}^{2}\sum_{a}^{7}\sum_{e}^{5}\sum_{j}^{3}\mathbf{C}_{sgaejot}}{\sum_{g}^{2}\sum_{a}^{7}\sum_{e}^{5}\sum_{j}^{3}\mathbf{H}_{sgaejot}}.$$

The nature of our key problem in the construction of the full-dimension employment matrices is the same as that of the construction of the full-dimension compensation matrices. That is, how to best estimate the missing dimensions in the full-dimension matrices (Table 1) based on the available information. The more information available, the more reliable the full-dimension matrices can be constructed. It is therefore necessary to look for more information in addition to the "control totals" at least for some time points even though it may not be consistent with the "control totals" because of different sources or different sizes of sampling.

After combing through all available information, published and unpublished from the NBS archives, we decide to focus the additional data work on eight years as benchmarks. They are 1955, 1963 and 1982 for the pre-reform period and 1987, 1990, 1995, 2000 and 2005 for the post-reform period. We set the year 1982 as the point that divides the pre- and post-reform periods, which was the eve of China's industrial reform that began in 1984. Having fewer benchmarks for the pre-reform period is less satisfactory. But the three time points can just avoid the radical fluctuations in 1958-62 (the Maoist feverish Great Leap Forward and its aftermath) and in 1966-76 (the Cultural Revolution). In other words, they may better reflect the underlying trend of the employment dynamics and relative cost of labor with different human capital attributes in the long run, which may be more realistically anchoring the estimated time series.

5.1 Additional Marginal Matrices for the Benchmark Years

In this section, we describe how additional marginal matrices are constructed for the benchmarks using the sources other than those used for the construction of our three annual matrices. Note that the additional matrices may not comply with the "control totals" as defined and constructed in Sections 4.1 and 4.2 and may have missing cells or incomplete dimension. Filling these gaps will be the tasks of Section 5.2 where the full-dimension matrices are constructed.

The 1955 and 1963 benchmarks

Our basic data used for extending the 1955 benchmark marginal matrix are from an unpublished NBS survey on 13,591 industrial firms which were state-owned or controlled by the state through state-private jointed ventures (DPES, 1955, pp. 133-5). They are, however, only available in two separate tables: numbers employed cross classified by occupation, industry and gender, and by occupation and age groups. With adjustment to our industrial classification, we can have two separate marginal matrices $\tilde{\mathbf{N}}^{M,55}_{sg++j}$ and $\tilde{\mathbf{N}}^{M,55}_{++a+j}$, but only for the state sector. Where the tilde "~" on the top of the matrix **N** indicates incomplete coverage of a dimension (here the "ownership" dimension) because of missing category (here the non-state categories) or missing cells.

The source, however, does not provide any information on education attainment. The only information on the education of industrial employees that we could find is

from a brief summary of another unpublished NBS survey conducted in 1957, which reports total industrial employment with four levels of education attainment for workers and "other types of employees" (DPES, 1957, p. 41). To match the education data with the gender, age and industry data in 1955, we first decompose the "other types of employees" with education attainment into administrative, technical and service staffs, and then decompose the estimates by industry. This gives us an additional marginal matrix $\tilde{\mathbf{N}}^{M,55}_{s++ej}$ for the state sector. The data used in the decompositions are the occupation-by-industry information from the aforementioned 1955 survey and the occupation data for total industry from the 1982 population census (see work below on the 1982 benchmark).[26]

The only source that allows us to extend the existing employment marginal matrix for 1963 is from an unpublished NBS survey on 35,766 state industrial firms available from the NBS archives (DPES, 1963, pp. 38-49). The survey results are only available in one table on numbers employed cross-classified by gender and occupation, but no information available on age and education with detailed industry breakdown (equivalent to 3-digit level by the 1972 CSIC), that is, $\tilde{\mathbf{N}}^{M,63}_{sg++j}$. We fill the education and age data by assuming that the 1963 benchmark is given by the linear trend that links the 1955 and 1982 benchmarks, i.e. $\tilde{\mathbf{N}}^{M,63}_{++a+j}$ and $\tilde{\mathbf{N}}^{M,63}_{s++ej}$.

Additional data on labor compensation are even more limited. For the entire central planning period, there are only five "labor-wage surveys" from the NBS archives, i.e. 1955, 1959, 1960, 1961 and 1963, available in different formats. The 1959 survey is relatively more informative that covers almost all state enterprises with 19.98 million employees at the time of survey (DPES, 1959, pp. 42-47), very close to our mid-year estimate of 21.55 million for the state sector. This survey provides wage bills cross-classified by numbers employed and by occupation with industry breakdowns. After some classification adjustment, we can obtain a marginal compensation matrix for the state sector in 1959 as $\tilde{\mathbf{C}}^{M,59}_{s+++j}$. Available wage surveys for 1955 and 1963 only report total wage bills paid to all employees and to workers in particular. Assuming that the ratio of non-worker occupations to workers in wage in 1955 and 1963 is the same as in 1959, the 1959 survey data are then used to estimate

[26] The 1955 data should are given more weight in the estimation as they are very close to 1957.

wage bills paid to non-worker occupations (administrative staffs, technicians and service employees) in 1955 and 1963. Note that as discussed in Section 4.2, employees in services are removed from the industrial workforce but compensation to them are kept and treated as payment for their services by industrial employees. After such an adjustment, we have obtained an additional marginal compensation matrix for the state sector for 1955 and 1963, respectively, $\tilde{\mathbf{C}}^{M,55}_{s+++j}$ and $\tilde{\mathbf{C}}^{M,63}_{s+++j}$.[27]

The 1982 and 1987 benchmarks

After ten years of interruption of the routine statistical work due to the Cultural Revolution (1966-76) and a slow recovery, the 1982 Population Census, though far from being sufficient or satisfactory by international standard, provides invaluable information for both the employment and labor compensation at the eve of the industrial reform. Importantly, as all institutions regulating employment and compensation experienced little change since the Cultural Revolution, the 1982 Census can serve as a close proxy for the period from the mid 1960s to the early 1980s.

The 1982 Population Census data are available in two formats: 1) national summary tables based on full samples and 2) full reports using one-percent of household samples in the census. The main problem in the 1982 Census data, as in later conducted censuses, is that the census questionnaires do not allow a direct integration (or full cross-classification) of the tables with different attributes. The best table available from the national summary is the numbers employed of each industry by age and gender (1982 Population Census, pp. 440-447), which gives us a marginal matrix $\tilde{\mathbf{N}}^{M,82}_{sga+++}$. Another table is the numbers employed with different level of education attainment (1982 Population Census, p. 461), which gives us another marginal matrix $\tilde{\mathbf{N}}^{M,82}_{s++e++}$. The one-percent of household samples of the 1982 Census also provides two useful tables for numbers employed by age, education and occupation and by gender, age and education, respectively, which allows us to obtain two more marginal matrices after, i.e. $\tilde{\mathbf{N}}^{M,82}_{++aej+}$ and $\tilde{\mathbf{N}}^{M,82}_{+gae++}$ (1982 Population Census, Vol. 2, pp. 1560-1587 and 2072-2135; Vol. 3, pp. 2328-2343 and 2392-2399).

[27] Note that "ownership type" dimension is missing in the marginal matrices for 1955, 1959 and 1963 as all the available data refer to the state sector. Missing dimensions for other sectors will be estimated by iterative proportional filling approach discussed later.

Since the 1982 Census data on the education attainment by occupation are the only information available for the late central planning period, they are used to help decompose the education data for industrial workforce as a whole for 1957 that are the only information on education available for the central planning period as discussed above, and then construct a table of numbers employed cross-classified by occupation and education for 1955 ($\tilde{\mathbf{N}}^{M,55}_{s++ej}$).[28]

An important reason for us to choose 1987 as the first benchmark for the post-reform period is because the NBS constructed its first SNA-type input-output table for that year with detailed labor compensation by industry. Besides, there are also data from the NBS 1987 One-Percent Population Sample Survey and the 1988 CASS Household Survey conducted by the Institute of Economic Research of CASS, which covered about 20,000 households. The CASS 1988 Household Survey provides average wage data that are cross-classified by four attributes of gender, age, education and occupation, but available only for two broad sectors, "mining" and "manufacturing & utilities". We assume the average wages of the CASS 1988 Survey are close proxies for the average compensation of the same types of labor in 1987, that is, they satisfy the following relationship:

(13) $$\mathbf{c}^{M,87}_{+gaej+} = \sum_s^{24}\sum_o^3 \mathbf{c}^{87}_{sgaejo} = \frac{\sum_s^{24}\sum_o^3 \mathbf{C}^{87}_{sgaejo}}{\sum_s^{24}\sum_o^3 \mathbf{N}^{87}_{sgaejo}}$$

Therefore, for each of the three broad sectors we can have a marginal matrix $\tilde{\mathbf{c}}^{M,87}_{+gaej+}$.

For the 1982 benchmark, we just assume the *relative* wages of all types of employees are the same in 1982 as in 1987. We argue that although the reform had affected labor compensation, the impact was similar among industries because they were almost all controlled by the state then.

The 1990 and 1995 benchmarks

Our work on the 1990 benchmark is based on our work for the 1987 benchmark and the 1990 Input-Output Table. Assuming the *relative* wages between all types of laborer available in 1987 are held for 1990, the average wage of each type of labor in

[28] Note that these additional human capital marginal matrices for 1982 do not exactly matching the items in the 1957 survey. We apply the iterative proportional filling (IPF) approach to obtain the estimates that match the 1957 survey data.

1987 is adjusted such that the average wage by industry is equal to the compensation per worker given by our "control totals" for 1990 in both numbers employed ($\mathbf{N}^{M,90}_{s+++jo}$) and total compensation paid ($\mathbf{C}^{M,90}_{s++++o}$) that is aligned with the 1990 Input-Output Table. This gives us an additional marginal matrix for average labor compensation for 1990, $\tilde{\mathbf{c}}^{M,90}_{+gaej+}$.

There is no additional employment information for the 1995 benchmark matrix, but additional information on compensation from the 1995 CASS Household Survey. The procedure used is almost the same as that for the 1987 benchmark. In the CASS 1995 Household Survey, average wage is cross-classified by four attributes, namely, gender, age, education and occupation, which is available only for two broad sectors, "mining" and "manufacturing". Based on the CASS 1995 Survey, following the same assumption as used for 1987 (Equation 13), we can construct a marginal matrix for average compensation for 1995, $\tilde{\mathbf{c}}^{M,95}_{+gaej+}$.

The 2000 benchmark

The additional information for the 2000 benchmark comes from the 2000 Population Census. Two of the census tables are used in this study, which allows two additional marginal matrices to be constructed. One is the 2000 Census Table 4-1, "Provincial Employment by Sector and Gender" that provides numbers employed in 92 two-digit industries, of which 6 are for mining, 30 for manufacturing and 3 for utilities. With such detailed industry breakdowns we can easily regroup these industries so as to align with the sectoral classification in this study (Table 1) and obtain $\tilde{\mathbf{N}}^{M,00}_{sg++++}$. The other source is the 2000 Census Table 4-5, "National Employment by Sector, Gender, Age, and Education". However, industries are only broadly defined as mining, manufacturing and public utilities. Therefore, we can have a marginal employment matrix cross-classified by gender, age and education for each of the three sectors, which can be denoted as $\tilde{\mathbf{N}}^{M,00}_{+gae++}$.

There are no additional labor compensation data to what we need for the construction of our annual series.[29] However, data from the China Household Income

[29] MOLSS Labour Market Survey for 1999 and 2000 has a sample size of 724,000 employed people, which is the ever largest wage/salary survey conducted in China. Only the data for 2000 are used in this study. One of the main deficiencies in this survey is that it has high level aggregation for

Project (CHIP)'s 2002 Household Income Survey can be used as a close proxy for 2000. The CHIP data set contains 6,835 sample households and their members in both urban and rural areas, only urban households data are used in this study.[30] From this data set, 9,584 individuals are identified as wage earners, of which 2,872 individuals work in mining, manufacturing and public utilities. Based on the CHIP income data we can construct average compensation by gender, age, occupation and ownership types for three broad sectors, namely, mining, manufacturing and public utilities. Thus, for each sector we now have $\tilde{\mathbf{c}}^{M,00}_{+gaej+}$, where "00" stands for the 2000 benchmark for simplicity.

The 2005 benchmark

The construction of the 2005 benchmark makes use of a rich data set based on national one-percent population sample survey for 2005. By assuming the sample survey data can represent the national situation, many difficulties encountered in earlier benchmarks are no longer a problem.

A summary of all the constructed marginal matrices

For readers to have a quick look of all constructed marginal matrices or "control totals" that will be used in the final construction of full-dimension matrices, we provide in Table 7 a summary of these marginal matrices for each benchmark year. In the last column of the table we also present matching marginal matrices from the annual marginal employment and compensation matrices constructed in Sections 4.1 and 4.2, as well as the average labor compensation matrices derived from them.

low-level education attainment, lumping together all types of labour with middle school-level education or below.

[30] We exclude the rural household sample though it contains individuals working in the industrial sector. This is because part of rural household income is earned jointly by all family members especially in farming activities and it is impossible to separate such income among members.

TABLE 7
MARGINAL MATRICES AVAILABLE FOR THE BENCHMARK YEARS

Benchmarks	Additional Marginal Matrix		Source of Data	Marginal Matrix used as the "Control Totals"
	Employment	Compensation		
1955	$\tilde{\mathbf{N}}^{M,55}_{sg++j}$, $\tilde{\mathbf{N}}^{M,55}_{++a+j}$, $\tilde{\mathbf{N}}^{M,55}_{s++ej}$	$\tilde{\mathbf{C}}^{M,55}_{s+++j}$, $\tilde{\mathbf{c}}^{M,55}_{s+++j}$	NBS archives: DPES *Bulletin*, 1955, 1959, 1960, 1961 and 1963	$\mathbf{N}^{M,55}_{s+++jo}$, $\mathbf{H}^{M,55}_{s+++jo}$ $\mathbf{C}^{M,55}_{s++++o}$, $\mathbf{c}^{M,55}_{s++++o}$
1963	$\tilde{\mathbf{N}}^{M,63}_{sg++j}$, $\tilde{\mathbf{N}}^{M,63}_{++a+j}$, $\tilde{\mathbf{N}}^{M,63}_{s++ej}$	$\tilde{\mathbf{C}}^{M,63}_{s+++j}$, $\tilde{\mathbf{c}}^{M,63}_{s+++j}$	As above…	$\mathbf{N}^{M,63}_{s+++jo}$, $\mathbf{H}^{M,63}_{s+++jo}$ $\mathbf{C}^{M,63}_{s++++o}$, $\mathbf{c}^{M,63}_{s++++o}$
1982	$\tilde{\mathbf{N}}^{M,82}_{sga+++}$, $\tilde{\mathbf{N}}^{M,82}_{s++e++}$, $\tilde{\mathbf{N}}^{M,82}_{++aej+}$, $\tilde{\mathbf{N}}^{M,82}_{+gae++}$	--	1982 Population Census	$\mathbf{N}^{M,82}_{s+++jo}$, $\mathbf{H}^{M,82}_{s+++jo}$ $\mathbf{C}^{M,82}_{s++++o}$, $\mathbf{c}^{M,82}_{s++++o}$
1987	--	$\tilde{\mathbf{c}}^{M,87}_{+gaej+}$	1988 CASS Household Survey; 1987 1%-Population Survey	$\mathbf{N}^{M,87}_{s+++jo}$, $\mathbf{H}^{M,87}_{s+++jo}$ $\mathbf{C}^{M,87}_{s++++o}$, $\mathbf{c}^{M,87}_{s++++o}$
1990	--	$\tilde{\mathbf{c}}^{M,90}_{+gaej+}$	Based on assumptions, see text…	$\mathbf{N}^{M,90}_{s+++jo}$, $\mathbf{H}^{M,90}_{s+++jo}$ $\mathbf{C}^{M,90}_{s++++o}$, $\mathbf{c}^{M,90}_{s++++o}$
1995	--	$\tilde{\mathbf{c}}^{M,95}_{+gaej+}$	1995 CASS Household Survey	$\mathbf{N}^{M,95}_{s+++jo}$, $\mathbf{H}^{M,95}_{s+++jo}$ $\mathbf{C}^{M,95}_{s++++o}$, $\mathbf{c}^{M,95}_{s++++o}$
2000	$\tilde{\mathbf{N}}^{M,00}_{sg++++}$, $\tilde{\mathbf{N}}^{M,00}_{+gae++}$	$\tilde{\mathbf{c}}^{M,00}_{+gaej+}$	2000 Population Census; 2002 CHIP Household Survey	$\mathbf{N}^{M,00}_{s+++jo}$, $\mathbf{H}^{M,00}_{s+++jo}$ $\mathbf{C}^{M,00}_{s++++o}$, $\mathbf{c}^{M,00}_{s++++o}$
2005	Full dimension	Full dimension	Full dimension	$\mathbf{N}^{M,05}_{s+++jo}$, $\mathbf{H}^{M,05}_{s+++jo}$ $\mathbf{C}^{M,05}_{s++++o}$, $\mathbf{c}^{M,05}_{s++++o}$

Note: For simplicity, $\mathbf{c}^{M}_{s++++o}$ indicates compensation per unit of labor either per person (**C**/**N**) employed or per hour worked (**C**/**H)**.

5.2 Construction of the Full-dimension Employment and Compensation Matrices

This section explains how the missing cells and dimensions are filled and estimated in order to construct full-dimension benchmark employment and compensation matrices, based on which the time series of the benchmark matrices can be constructed. We will first review the nature of the gaps and then introduce the iterative proportional filling (IPF) approach for these tasks.

The missing cells in the additional marginal matrices

There are two types of missing cells. One is a natural type simply because some classifications in employment do not or seldom exist naturally (e.g. managers or technicians in the youngest age group or young female workers in heavy industries). The other type is caused by some sampling biases, which is more likely when the sample size is small. Missing cells are a typical problem encountered by the additional marginal matrices for labor compensation because surveys on compensation are usually conducted on much smaller samples than those on employment that often take the form of industrial census, population census or one-percent population survey.

A regression approach is used to fill the missing compensation data in the available marginal matrices listed in Table 7. In the regression, the available compensation (often referring to wage or salary[31]) is regressed on a set of dummy variables that distinguish the effects of human capital attributes (Table 1). The results are fairly satisfactory with reasonably high $\overline{R}^2$ for cross-section analysis. The estimated coefficients of the dummy variables are then used to predict the missing values.

However, the likely biases contained in small sample size data are maintained in such a regression exercise, which cannot be easily fixed. Since the compensation surveys are reasonably designed, we argue that despite small samples the biases are moderate and will not cause significant distortion in the final results.

Estimation of the full-dimension matrices for benchmarks

The estimation of a full-dimension benchmark matrix is the key step for the construction of a complete time series matrix for labor employment and compensation. In our case, we need to estimate eight full-dimension benchmark matrices at this step.

Understanding the nature of the estimation especially the underlying assumptions of the estimation: the results are determined by the information from the "control totals" and additional marginal matrices in a given benchmark.

It is a scaling process to estimate a full-dimension matrix with the scope or size exactly the same as given by the "control totals". For example, the scaling up of the CASS survey results on compensation to the level of the 1987 "control totals",

[31] Compensation paid to auxiliary (non-industry) workers in industry should be considered as part of the compensation of industrial workers. This is discussed in the construction of the time series marginal matrices in 4.2 .

$\mathbf{C}^{M,87}_{s++++o}$, which is based on China 1987 Input-Output Table and the average wage/salary level by sector and occupation, gives us another marginal compensation matrix, $\mathbf{C}^{M,87}_{+gaej+}$, which is consistent with $\mathbf{C}^{M,87}_{s++++o}$.

Since the scaling does not affect the given structure of the marginal matrix involved, it assumes the structure of the "control totals" is held. Therefore, it is does not introduce any new problem.

Also, we assume that all the constructed marginal matrices, those used as the "control totals" and those as additional, are the best available.

In this study, we use the iterative proportional filling (IPF) approach to perform the task of estimating a full-dimension benchmark matrix from the available marginal matrices. The IPF is designed to link disintegrated tables containing partial data on cross-classification by generating the maximum likelihood estimate of each element of a matrix. It can also be used to fill missing cells (Bishop, Fienberg and Holland, 1975).[32]

6. Preliminary Results and Discussions

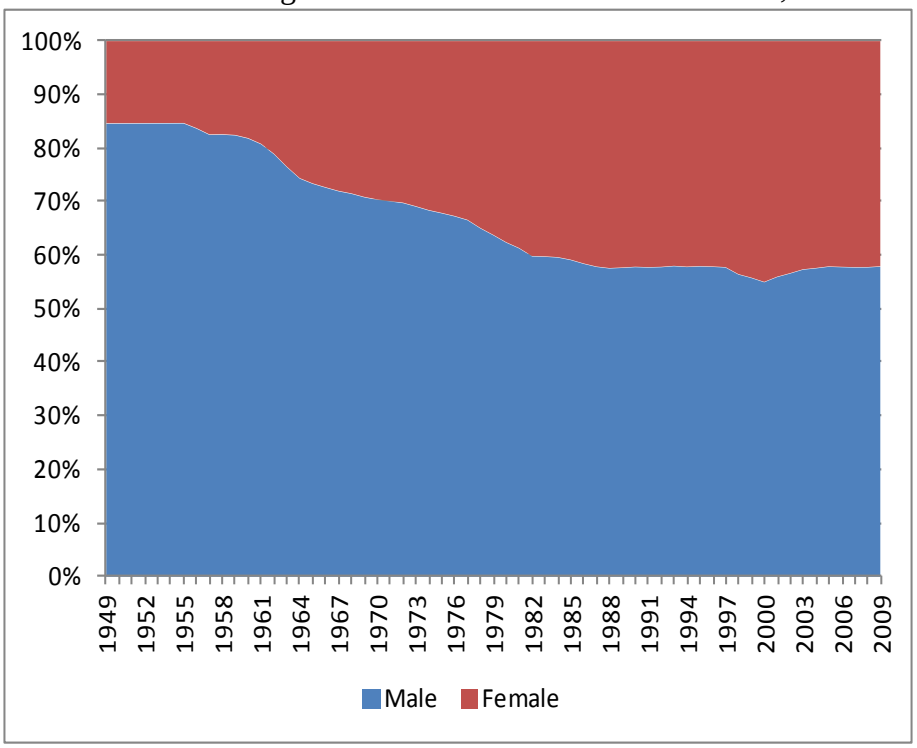
In this section we present the preliminary results of the work. Our discussion focuses on the plausibility of the results by examining the background of the macroeconomic performance, reforms and policy regime shifts, rather than interpreting the results as if the data work and estimation are problem free.

Structural changes of industrial workforce

The first picture we would like to show is the change of the gender structure of the Chinese industrial workforce over time. Figure 1 suggests that a significant increase in female participation in industrial production began with the implementation of the first five year plan in the early 1950s. It had continued over the entire central planning period. The re-structuring at the beginning of economic reform still saw the increase in female employment till the late 1980s. Further reforms in the 1990s, especially the reform of the over-staffing state enterprises finally stopped this trend. What we can see for 2009 is perhaps a nearly steady status.

[32] Interested readers may refer to for example the discussions of a five-dimension model with missing variables (Bishop, Fienberg and Holland, 1975, pp. 45-46), of the rules for detecting existence of direct estimates (pp. 76-77), and of the basic elements of the IPF (p. 83).

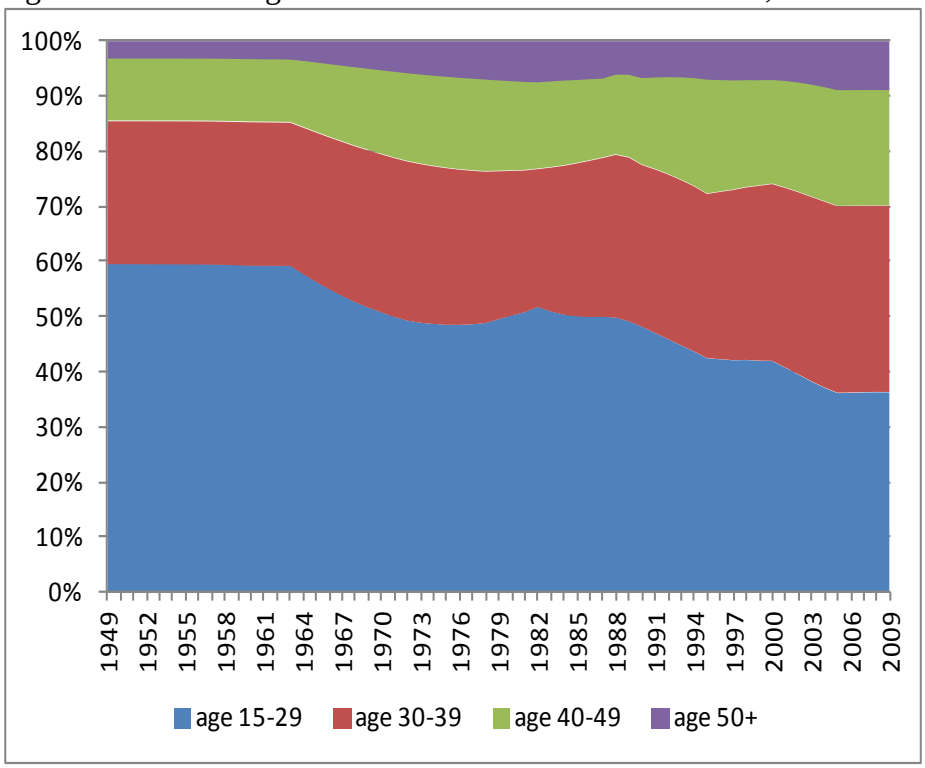
Figure 1
Gender Structural Change of the Chinese Industrial Workforce, 1949-2005



Sources: Authors' estimates.

Considering the age structure of the Chinese industrial workforce, with very limited data, we could not show any age structural change for the 1950s, or we could not capture the very likely younger structure in the 1950s (similar problem for the period after 2005, see Figure 2 for both periods). As shown in Figure 2, the resurgence of the youngest part of the industrial workforce (regrouped to include aged 15-29) from the mid 1970s reflected the post-war baby boomers entering the working age. If that group is considered the one that has benefited China with the so-called demographic dividend, it already started to decline in industry from the early 1980s. From the 1980s onwards, the average age of the industrial workforce has been increasing. And this will be continuing.

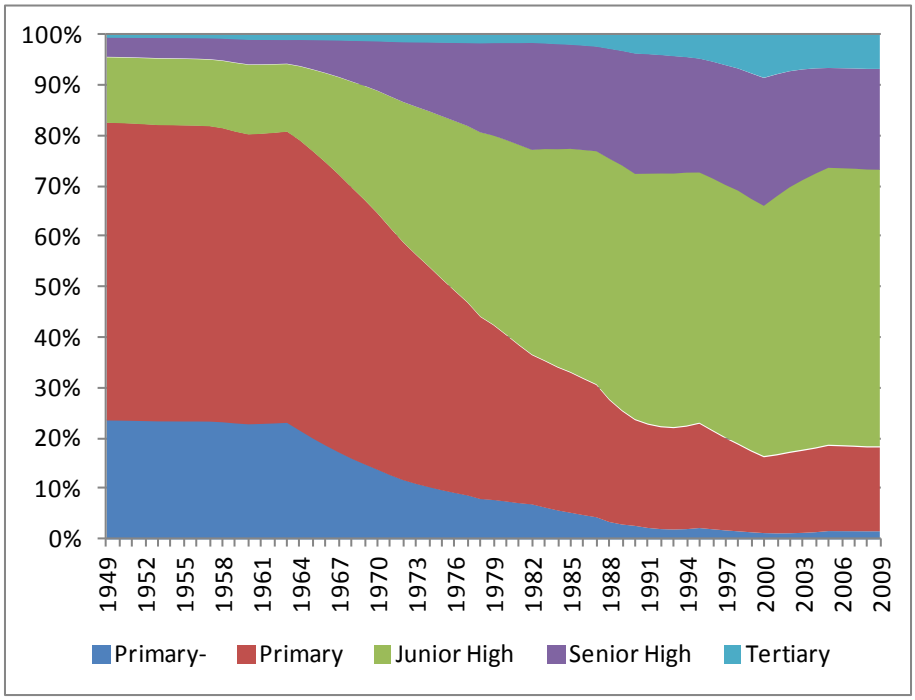
Figure 2
Age Structural Change of the Chinese Industrial Workforce, 1949-2005



Sources: Authors' estimates.

As shown in Figure 3, the education attainment status of the industrial workforce has been improving rapidly. By the late 2000s, the share of the workforce who obtained education up to junior high school level increased to 55 percent from 13 percent in the early 1950s and during the same period whose who obtained the senior high school level reached from 4 percent to 20 percent. This trend seems to be stable from the late 2000s and will not change significantly in the near future unless there is a significant technological upgrading in the industrial sector.

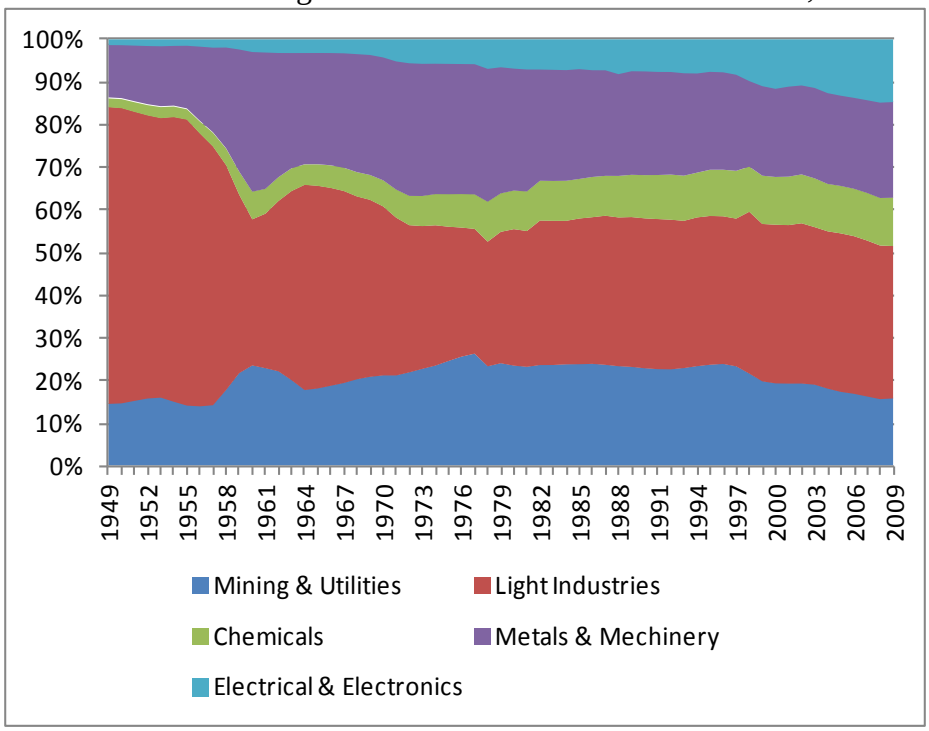
Figure 3
Structural Change of Education Attainment of the Chinese Industrial Workforce, 1949-2005



Sources: Authors' estimates.

The change in the industrial structure of Chinese industry has been dramatic. As Figure 4 demonstrates, during the second half of the 1950s the forced Soviet-style heavy industrialization quickly drove out a substantial share of the labor-intensive light manufacturing where China has comparative advantage. After a short-lived retreat, this decline continued till the reform began when this sector bounced back and enjoyed a continuous expansion since the reform. Benefiting from China's WTP entry, by the late 2000s, this sector regained its share in the late 1950s. The most significant development has been the electrical and electronic industries where Chinese cheap but qualified labor working for assembling lines for the export market. On the other hand, the reform has also stopped China's rapid expansion in the machinery industry. Figure 4 also suggests that primary resource based industries have been declining, which also show China's shift away from its comparative disadvantage industries driven by the reform.

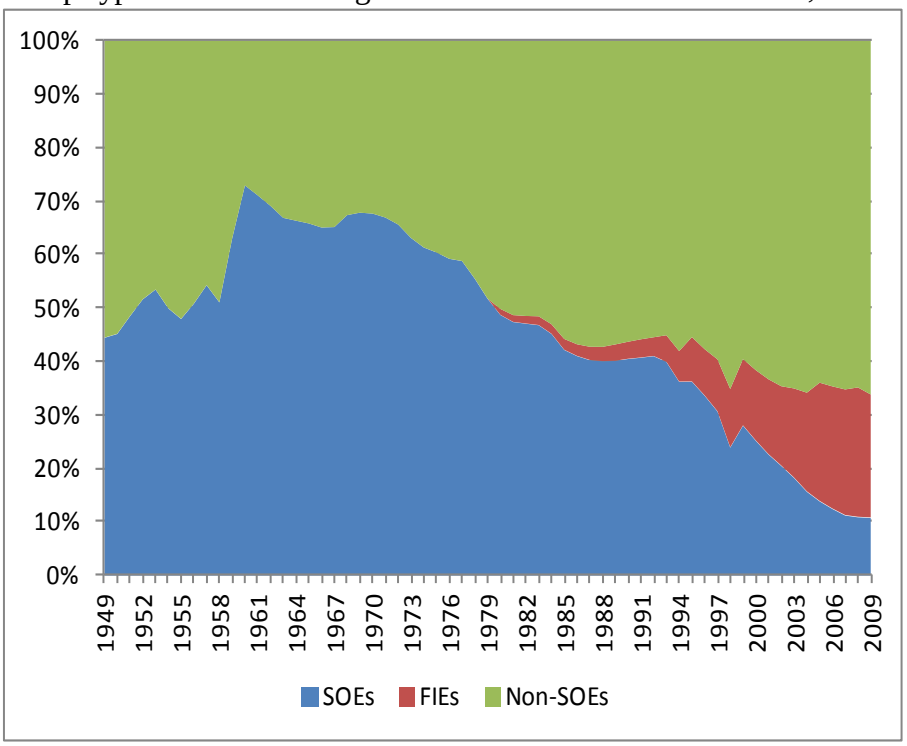
Figure 4
Industrial Structural Changes in the Chinese Industrial Workforce, 1949-2009



Sources: Authors' estimates.

However, the most dramatic change has happened in the ownership structure of the Chinese industrial workforce (Figure 5). The government's "socialist industrialization" campaign in the 1950s raised the state share of the industry from 45 to over 70 percent, and at the same time collectivized all "non-state" enterprises. In fact, until the industrial reform began in the mid 1980s, the changes observed between the state and the non-state sector did not make a real sense because of the planning controls over resource allocation in all industries and in all ownership types. Besides, under central planning there existed no foreign invested firms. The change began along with the reform in the late 1970s, not only were foreign investors allowed to invest in China, but along the SOE reform the collectives were also allowed to detach from the planning controls. By the late 2000s, the share of the state sector in the industrial employment declined to only about 10 percent of the national total if measured in hours worked, representing an absolutely decline in numbers employed from 45 million in 1993 to 18 million in 2009. As will be shown in our estimates of the labor input, the change of ownership structure over time has had a significant impact on the composition of labor in the Chinese workforce.

Figure 5
Ownership-type Structural Change of Chinese Industrial Workforce, 1949-2005



Sources: Authors' estimates.

Growth of labor input, hours worked and the composition of labor

Our preliminary estimates for the constant-price value of labor input and related basic indicators are reported in Tables A1 and A2 in the appendix. Their growth rates are reported in Tables 8 and 9 for the planning and reform periods and their sub-periods. The cost-weighted all types of labor in our exercise allow us to measure of the labor input in Chinese industry at the 1987 constant prices. By incorporating it with the constructed labor price index, we can obtain the total compensation for the total labor input, i.e. the measure of "outlay". The difference between the quantitative index and the labor input index is the effect of labor composition that may imply the "quality" of the workforce though subject to some strong assumptions about the efficiency of the labor market. Finally, in these results tables we also report weekly hours per person employed and compensation paid for per hour worked. Note that these are average measures that have already taken into account the labor composition effect (i.e. all types of labor being made homogeneous).

Table 8: Compound Annual Growth Rate of Labor Input (Hours-based) and Basic Labor Indicators in Chinese Industry, 1949-2009

(Percent)

	Labor Input	Price	Outlay	Composition	Numbers	Hours	Weekly Hours Per Person	Compensation per Hour
Planning:								
1949-1952	11.8	8.6	22.2	2.1	14.7	9.7	-4.0	10.9
1953-1957	6.8	11.3	19.1	1.1	6.6	5.6	-0.7	12.6
1958-1965	3.0	0.7	3.8	0.3	2.5	2.7	0.2	1.0
1966-1971	10.9	-3.2	7.9	1.6	9.2	9.3	0.4	-1.6
1972-1978	8.2	0.4	9.0	0.3	9.6	7.9	-1.3	0.7
1949-1978	6.9	***2.3***	***10.2***	***0.8***	***7.5***	6.2	***-0.7***	***3.3***
Reform:								
1979-1984	5.8	8.4	14.8	-0.6	7.2	6.4	-0.6	7.7
1985-1991	4.7	12.8	18.2	0.2	4.4	4.5	0.2	13.0
1992-2001	0.7	20.9	21.8	0.2	0.5	0.5	0.1	21.1
2002-2009	5.2	7.2	13.0	0.0	4.1	5.2	1.2	7.2
1979-2009	3.8	***13.0***	***17.3***	***0.0***	***3.6***	3.8	***0.3***	***13.0***

Sources: Authors' estimates.

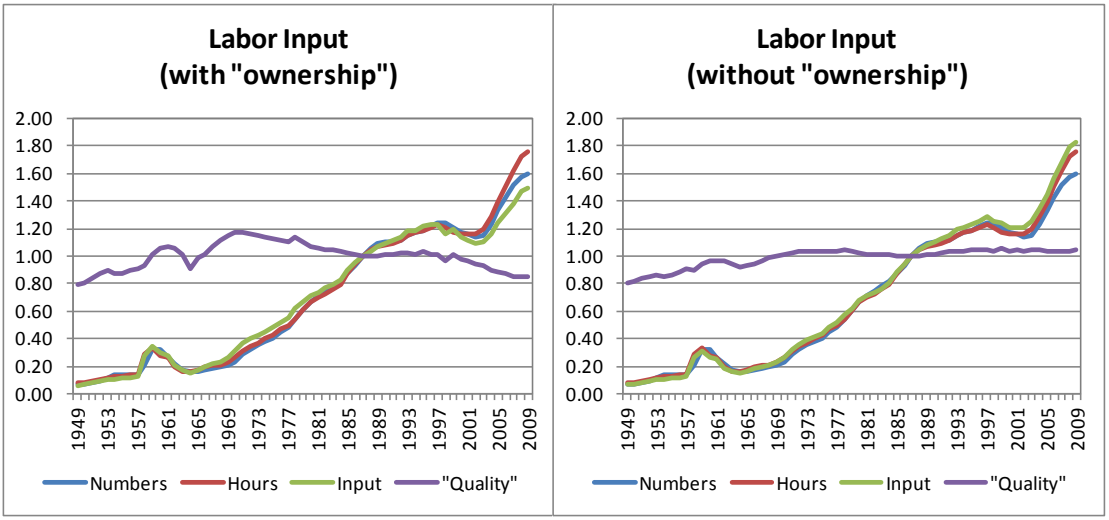
Notes: The estimation of labor input and composition is hours based.

As summarized in Table 8, the growth of labor input in Chinese industry slowed down from nearly 7 percent per year in the central planning period to less than 4 percent during the reform period. We argue that this was likely caused by a rapid shift to more labor intensive manufacturing which increased the demand for young, unskilled labor and a substantial correction effect to the previous overstaffing in the state sector under central planning. However, the composition of labor shows almost no change in the reform period compared with that of the planning period. One likely reason is that China's shift to labor-intensive manufacturing mainly relied more on disciplined young labor, even females, with basic education. While the quantity of labor did not increase in the reform period (3.8 percent per annum) as fast as in the planning period (6.2 percent), the average hours worked per week increased during the reform rather than declined on average as in the planning period. This finding also fits the nature of the post-reform labor-intensive expansion of Chinese industry.

One of the important findings is that changes in the ownership structure of the economy driven by the changes in policy regime have a significant downward impact on labor composition and hence reducing labor input, specifically, the decline of the state sector following the reform and, meanwhile, the increase in the labor market competition brought by non-state enterprises, especially foreign firms. As Figure 6 shows, if the ownership dimension is included (the left-hand panel), the composition effect will demonstrate a very different pattern compared with the results without it (the right-hand panel). This is because the pre-reform overstaffed SOEs paid much

more to the same type of labor, i.e. the same human capital attributes (gender, age and education) and engaged in the same industry (the same technology implied the cost of the same labor), and underwent an absolute decline since the SOE reform began in the early 1990s. But reallocation of the existing labor to other ownership enterprises does not mean the real composition of the Chinese industrial workforce has changed. Although in the longer term, the behaviour of the non-state and foreign enterprises in the labor market will unquestionably induce changes in labor supply, including the ownership dimension in the estimation of labor input and the effect of labor composition will distort the true picture.

Figure 6
Quantity, Composition and Labor Input in Chinese Industry
(1987=1)



Sources: Authors' estimates.

Change of labor composition: The main effects

This part of results show how the main human capital effects changed over the period and against the backdrop of the shifts of the main policy regime. As reported in Table 9, changes in industrial structure (sector) and age structure played the most important role in the planning period. Not surprisingly, education played a negative role during the Cultural Revolution. During the reform period, especially during the reform of the state sector in the 1990s, education showed a significant positive effect, but China's WTO entry turned it around, likely due to the expansion of labor-intensive industries. Structural change was no longer as important as in the planning period because labor was reallocated to labor intensive industries that did not have a high demand for high level of human capital. Changes in age structure played a positive role in both the

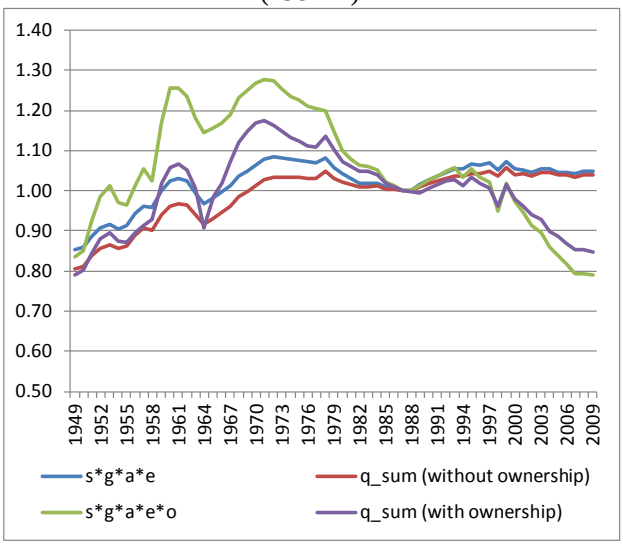
planning and reform period. This is likely because seniority captures the effect of experiences in production, technology and management.

Table 9: Annual Change of Main Human Capital Factors on the Composition of Labor in Chinese Industry
(Percent per annum)

	Composition without ownership	*Composition with ownership*	Sector (*s*)	Gender (*g*)	Age (*a*)	Education (*e*)	*Ownership (o)*
Planning:							
1949-1952	2.06	*3.56*	2.03	0.00	0.01	0.01	*3.51*
1953-1957	1.14	*0.75*	1.14	-0.02	0.01	0.02	*0.21*
1958-1965	0.35	*0.93*	0.11	-0.16	0.24	0.09	*0.91*
1966-1971	1.60	*2.93*	1.04	-0.10	0.62	-0.01	*0.09*
1972-1978	0.29	*-0.48*	0.35	-0.10	0.08	-0.29	*-0.97*
1949-1978	***0.77***	***0.98***	***0.59***	***-0.10***	***0.24***	***-0.05***	***0.08***
Reform:							
1979-1984	-0.59	*-1.48*	-0.38	-0.15	-0.26	-0.21	*-1.13*
1985-1991	0.16	*-0.32*	-0.03	-0.06	0.27	0.04	*-0.42*
1992-2001	0.18	*-0.54*	-0.26	-0.06	0.26	0.22	*-1.09*
2002-2009	-0.01	*-1.59*	0.03	0.10	0.12	-0.30	*-2.22*
1979-2009	***-0.02***	***-0.94***	***-0.15***	***-0.03***	***0.13***	***-0.04***	***-1.24***

Sources: Authors' estimates.

Figure 7
Estimated Main Effects on Labor Composition ("Quality") Change
(1987=1)



Sources: Authors' estimates.
Notes: s=industry/sector, g=gender; a=age; e=education; o=ownership type (see Table 1). q_sum=total effect of labor composition ("quality")

Table 9 also shows that by dropping the "ownership", the annual growth of labor composition changes from 0.98 to 0.77 percent per annum for the planning period. However, as for the reform period, it changes from -0.94 to -0.02 percent per annum. The results show that the most significant positive ownership effect is in the early

planning period and the most significant negative ownership effect is seen in the post reform period, especially in the state sector reform and China's WTO entry. The main effects with and without the ownership effect are depicted in Figure 7.

7. CONCLUDING REMARKS

Following the user cost theory on measuring labor input, after a careful scrutiny of available information we construct employment and compensation matrices for China's industrial workforce over the period 1949-2009. Our measures are able to capture both individual and interactive effects of changes in gender, age, education, industry and ownership types of China's industrial workforce, and decompose the growth of labor input in Chinese industry into quantity and composition ("quality") effects. We find that the annual growth of the labor input in Chinese industry experienced a substantial decline from 6.9 percent per annum in the pre-reform period to 3.8 percent per annum in the post-reform period. This could be explained by an effect of correction to various distortions under central planning, including over-manning and over-focusing on heavy industries.

Change of labor composition accounted for about 12 percent in the planning period (or 0.8 percent growth per annum), but it made little contribution during the reform period. We also find that the change in industrial structure and age structure (reflecting the effects of seniority and experience) almost explained for the entire (positive) change in labor composition in the planning period. However, the change of education turned into negative after 1965 which made the average contribution of education negative in the planning period. Following the reform, education showed the most important contribution in the 1990s when the reform deepened, but the effect turned into negative again alongside China's WTO entry. This could be explained by the labor-intensive and export-orientation nature of China's post-WTO industrial expansion.

APPENDIX

This appendix is devoted to give a detail description of use of iterative proportional fitting technique proposed Bishop et al. (1975), IPF technique for short below, by to compile compensation matrix.

Traditional way of using the IPF in compiling labor input is to derive employment matrix under a situation that one is cross-classifying workers by more characteristics of workers than those in data available. For instance, one might want to cross-classify total workers by six characteristics: sector, gender, age, educational level, status of employment and occupation, while data sources available may only provided data on employment that are cross-classified by characteristics of less six. In order to obtain employment matrix with workers fully cross-classified by the six characteristics, the IPF technique has so far been employed. The application of this technique, however, doesn't confines to generation of employment matrix, and can also be made to estimate compensation matrix in the same setting. That is, if data or each of several data sets available on wage rates of workers that are not to be cross-classified by as many characteristics as we need, then the IPF technique can be employed to estimate compensation matrix for workers cross-classified by full dimensions of workers' characteristics, given availability of the number of corresponding types of workers. In the following we taking 1987 as example, gives a detail of how the iterative proportional fitting technique is applied for estimating compensation matrix.

The ultimate goal of constructing compensation matrix for 1987 is to estimate wage rate of each type of workers cross-classified by five characteristic used in this study. Data sources, however, don't provide employment data on worker at such great detail. There is broad accordance in categories of each of all characteristics except sector between data source of 1987 CASS household survey and that required in the study. Regarding the characteristics of sector, there is no further subdivision of mining and manufacturing in the CASS survey, while mining is further divided into 4 sub-sectors and manufacturing into 19 sub-sectors in this study. Furthermore, the sector of public utilities is included in manufacturing in 1988 CASS survey, and mining, manufacturing and public utilities are lumped together into one sector named industry in 1995 survey. So we need special device to estimate compensation matrix

with workers classified by gender, age education and occupation for sub-sectors of each of mining and manufacturing, with the latter including public utilities.

Actually, two types of data on wage of workers are available for constructing compensation matrix in 1987. One, coming from CASS household survey, is wage rates of workers for whole mining, as well as whole manufacturing, with the workers cross-classified by other four characteristics: gender, age, education and occupation. Please note that the public utilities is included in manufacturing. We denote each element of the wage rates by w^{mi}_{gaeo} for mining and by w^{ma}_{gaeo} for manufacturing, inclusive of the public utilities. Other data is average compensation of workers for each sub-sector of mining and manufacturing, as well as the public utilities, derived from 1987 IO Table as total compensation of each sector divided by the number of workers in the corresponding sub-sector. The number of sub-sectors, clearly, is 24, of which 4 for mining, 19 for manufacturing and 1 for public utilities. We denote a cell in vector of the sectoral average compensation by w^{mi}_{s} in mining and w^{ma}_{s} in manufacturing. Again w^{ma}_{s} includes the public utilities. With the two types of data, the maximum likelihood estimate of each cell in compensation matrix for each of mining and manufacturing can be derived by implementing the IPF technique. Detail of how to implement the technique is as follows.[33] Because it is exactly the same for mining and manufacturing, we only give description to manufacturing and omit it for mining.

If we use $\hat{w}^{ma}_{sgaeo}$ to denotes estimate of element of compensation matrix that we want to obtain, then it must meet two constraints given by the two types of wage data described just above. That is,

(A1) $$\sum_{s=1}^{20} \hat{w}^{ma}_{sgaeo} * e^{ma}_{sgaeo} = w^{ma}_{gaeo} * e^{ma}_{+gaeo}$$

(A2) $$\sum_{g=1}^{2}\sum_{a=1}^{7}\sum_{e=1}^{5}\sum_{o=1}^{4} \hat{w}^{ma}_{sgaeo} * e^{ma}_{sgaeo} = w^{ma}_{s} * e^{ma}_{s++++}$$

where 20 stands for the number of sub-sectors in manufacturing, including the public utilities. The first formula means that the sum of products of estimated wages and the

[33] See Chapter 3 of Bishop and etc. (1975) for detail of implementing the iterative procedure in a situation that is suitable to generate an employment matrix.

number of workers over sub-sectors must equal total wage given by W_{gae}^{ma} for each type of workers classified by gender, age, education and occupation. And in the same way, the second formula means that the sum of products of estimated wages and the number of workers over gender, age and education must be equals to sectoral total of compensation that is given by w_s^{ma} for each of the 20 sub-sectors. Simple alteration to the two equations can produce other expressions that make the estimate $\hat{w}_{sgaeo}^{ma}$ much easier understand,

(A1-1) $$\sum_{s=1}^{20} \hat{w}_{sgaeo}^{ma} * \frac{e_{sgaeo}^{ma}}{e_{+gaeo}^{ma}} = w_{gaeo}^{ma}$$

(A2-1) $$\sum_{g=1}^{2}\sum_{a=1}^{7}\sum_{e=1}^{5}\sum_{o=1}^{4} \hat{w}_{sgaeo}^{ma} * \frac{e_{sgaeo}^{ma}}{e_{s++++}^{ma}} = w_s^{ma}$$

first of two equations implies that weighted average of estimated wages $\hat{w}_{sgaeo}^{ma}$ cross sectors within each type of workers cross-classified by gender, age and education is equals to wage rate available from CASS survey. Second of the two equations, of cause, can be explained in the same way.

Returning to how $\hat{w}_{sgaeo}^{ma}$ is estimated, the iterative proportional fitting procedure, starting with initial values of $\hat{w}_{sgaeo}^{ma}$, which we denote by $\hat{w}_{sgaeo}^{ma}(0)$ and usually set as $\hat{w}_{sgaeo}^{ma}(0) = 1$, first fits to first constraint specified by equation (A1) by calculating $\hat{w}_{sgaeo}^{ma}$ as $\hat{w}_{sgaeo}^{ma}(1,1)$:

(A3) $$\hat{w}_{sgaeo}^{ma}(1,1) = \hat{w}_{sgaeo}^{ma}(0) * \frac{\hat{w}_{gaeo}^{ma} * e_{+gaeo}^{ma}}{\sum_{s=1}^{22} \hat{w}_{sgaeo}^{ma}(0) * e_{sgaeo}^{ma}}$$

where the first of two units in parentheses in left of the equation represent first constraint given by equation (A1), and the second 1 represents the time of iteration. It takes values of 1 because this is first cycle in iterative process. And it is clear from equation (A3) that after fitting to first constraint for the first time, all the $\hat{w}_{sgaeo}^{ma}(1,1)$ are assigned with the same value of average wage within the same type of workers cross-classified by gender, age and education. After the first fitting, $\hat{w}_{sgaeo}^{ma}(1,1)$ became to meet the first constraint specified by equation (A1), but doesn't meet the

second constraint specified by equation (A2). So the procedure continues to fit to the second constraint by calculating $\hat{w}^{ma}_{sgaeo}(1,1)$ as $\hat{w}^{ma}_{sgaeo}(2,1)$:

$$\text{(A4)} \qquad \hat{w}^{ma}_{sgaeo}(2,1) = \hat{w}^{ma}_{sgaeo}(1,1) * \frac{w^{ma}_{s} * e^{ma}_{s++++}}{\sum_{g=1}^{2}\sum_{a=1}^{7}\sum_{e=1}^{5}\sum_{o=1}^{4} \hat{w}^{ma}_{sgaeo}(1,1) * e^{ma}_{sgaeo}}$$

where the first figure of 2 in parentheses in left of equation represents the second constraint given by equation (A2) above. And the second figure in parentheses represents the time of iteration. It is takes value of 1 because one iteration is considered finished after fitting all the constraints and fitting the second constraint is second step of the first iteration.

After second fitting in the first iteration, $\hat{w}^{ma}_{sgaeo}(2,1)$ comes to meet the second constraint, but not meet the first constraint. The iterative procedure enters second cycle of iteration, fitting to the first constraint by repeating calculation shown in equation (A3) and then fitting to the second constraint by repeating calculation shown in equation (A4). And furthermore, the cycle itself is iterated as many times as desired. The estimate of $\hat{w}^{ma}_{sgaeo}(2,i)$ converge in the sense that for any small figure of ε, we can find an integer of n such that absolute value of difference $\hat{w}^{ma}_{sgaeo}$ between two successive cycles are smaller than ε when the times of iteration exceeds n. That is, $\left|\hat{w}^{ma}_{sgaeo}(2,n) - \hat{w}^{ma}_{sgaeo}(2,n-1)\right| < \varepsilon$. When ε is set as $\varepsilon = 1\text{E-}06$, the absolute value of difference $\hat{w}^{ma}_{sgaeo}$ between two successive cycles usually become to be smaller than $1E-06$ after about 30 times of iteration.

REFERENCES

Bai, Hejin (2002), *A Guideline to the Major Economic Events of the People Republic of China* [*Zhonghua renmin gongheguo jingji dashi jiyao*], China Planning Press, Beijing

Barro, R. J., & Lee, J. W. (1997). International measures of schooling years and schooling quality. *American Economic Review*, Papers and Proceedings, 86(2), 218–223.

Barro, R. J., & Lee, J. W. (2000, April and August). International data on educational attainment: updates and implications. CID Working Paper No. 42. Center for International Development at Harvard University, Cambridge, MA.

Bishop, Yvonne M. M., Stephen E. Fienberg, and Paul M. Holland [1975], *Discrete Multivariate Analysis*, Cambridge, MA: MIT Press.

Borensztein, Eduardo and Jonathan D. Ostry, 1996 "Accounting for China's Growth Performance", *American Economic Review* 86 (2), 224-28.

CASS [Chinese Academy of Social Sciences], 1988 and 1995, *Household Survey*, 1988 and 1995. Beijing: Institute of Economic Research, Chinese Academy of Social Sciences

Chen, Kuan, Gary H. Jefferson, Thomas G. Rawski, Hongchang Wang, and Yuxin Zheng, 1988, "Productivity Change in Chinese Industry: 1953-1985", *Journal of Comparative Economics* 12 [4], 570-91

Chinloy, Peter T., 1980, "Sources of Quality Change in Labor Input", *America Economic Review* 70 [1], 108-19

CHIP (China Household Income Project), 2002 and 2007, *China Household Income Survey*. Beijing: Institute of Economic Research, Chinese Academy of Social Sciences

Chow, Gregory C. (1993). "Capital Formation and Economic Growth in China", *Quarterly Journal of Economics* 108 (3), 809-42.

Christensen Laurits R., Dale W. Jorgenson and Lawrence Lau, 1973, Transcendental Logarithmic Production Frontiers. *Review of Economics and Statistics* 55, no. 1 (February): 28-45.

CSIC (China Standard of Industrial Classification), *China Standard of Industrial Classification*, 1975, 1985, 1994 and 2002. Beijing: China Statistical Press.

Denison, Edward F., 1961, "Measurement of Labor Input: Some Questions of Definition and the Adequacy of Data". In *Output, Input, and Productivity Measurement*, 347-72. NBER Studies in Income and Wealth. Princeton, N.J.: Princeton University Press

Denison, Edward F., 1962, *Sources of Economic Growth in the U.S. and the Alternatives before Us*, New York, Committee for Economic Development

Denison, Edward F., 1974, Accounting for United States Economic Growth, 1929-1969, Washington DC, Brookings

DITS [Department of Industrial and Transportation Statistics, National Bureau of Statistics, China], various years, *Gongye Jiaotong Tongji Nianbao* [Annual Statistical Report on Industry and Transportation] [Internal Publication]. Beijing: National Bureau of Statistics

DITS, 1978, 1979, 1980, 1981, 1982 and 1983, *Annual Bulletin on Commune Factories*. Beijing: NBS Archives

DITS, 1984 and 1985, *Annual Bulletin on Township Enterprise*. Beijing: NBS Archives

DITS, 1999, *Xinbian Gongye Tongji Gongzuo Zhinan* [New Handbook of Industrial Statistical Work], Beijing: China Statistical Publishing House

DITS, 2000, *Zhongguo Gongye Jiaotong Nengyuan Wushinian Tongji Ziliao Huibian: 1949-1999* [Comprehensive Statistics of Fifty Years in Chinese Industry,

Transportation and Energy: 1949-1999], Beijing: China Statistical Publishing House

DPES, 1955, 1957 and 1963 *Labor and Wage Annual Bulletin* (Internal), National Bureau of Statistics of China

DPSSTS, 1998, [Department of Population, Social and Science and Technology Statistics], *Xinbian Laodong Tongji Shiyong Shouce* [A New Handbook of Practical Labor Statistics], Beijing: China Statistical Publishing House

DPSSTS, 2001 and 2006, *China Labor Statistical Yearbook*. Beijing: China Statistical Publishing House

DSS [Department of Social Statistics, National Bureau of Statistics, China], 1987, *Laodong Gongzi Tongji Ziliao 1949-1985* [Labor Employment and Wage Statistics 1949-85]

DSS, various years, *Laodong Gongzi Tongji Nianbao* [Annual Statistical Report on Labor and Wage] [Internal]. Beijing: National Bureau of Statistics

Gollop, Frank M. and Dale W. Jorgenson, 1980, "U.S. Productivity Growth by Industry, 1947-1973", in John W. Kendrick and Bertrice Vaccara [eds.], *New Developments in Productivity Measurement*, Chicago, University of Chicago Press, 17-136

Gollop, Frank M. and Dale W. Jorgenson, 1983, "Sectoral Measures of Labor Cost for the United States, 1948-1978", in Jack E. Triplett [ed.], *The Measurement of Labor Cost*, Chicago, University of Chicago Press, 185-235, 503-20

Holz, A.C. and Yi-min Lin, 2001, "Pitfalls of China's Industrial Statistics: Inconsistencies and Specification Problems", *The China Review*, 1 [1], 29-71

Hu, Zuliu F. and Mohsin S. Khan, 1997, "Why is China growing so fast?", *IMF Staff Papers*, 44 (1), 103-131.

Jorgenson, D.W. 1990, Productivity and Economic Growth, in Ernst R. Berndt and Jack E. Triplett (eds.), *Fifty Years of Economic Measurement*, NBER, Studies in Income and Wealth Volume 54. Chicago and London: University of Chicago Press

Jorgenson, D.W. and Zvi Griliches, 1967, "The Explanation of Productivity Change," *Review of Economic Statistics*, Vol. 34 [3], 249-283

Jorgenson, Dale W., Frank Gollop and Barbara Fraumeni, 1987, *Productivity and U.S. Economic Growth*, Harvard University Press, Cambridge, MA

Kendrick, John W., 1961, *Productivity Trends in the United States*. Princeton, N.J.: Princeton University Press

Kendrick, John W., 1973, Post-war Productivity Trends in the United States, *1948-1969*, New York: NBER

Li, Jingwen, Dale W. Jorgenson, Zheng Youjing and Masahiro Kuroda, 1993, *Productivity and Economic Growth in China, USA and Japan*, China Social Science Publishing House, Beijing

MOLSS [Ministry of Labor and Social Security, China] 1999 and 2000, *Zhongguo Laodongli Shichang Gongzi Zhidao Jiawei 2000* [Guidance of Wage Rates for

Labor Market in China]. Beijing: Zhongguo Laodong Shehui Baozhang Chubanshe [China Labor and Social Security Press].

NBS (National Bureau of Statistics, China) (various years), *China Statistical Yearbook*, China Statistical Press, Beijing

NBS, 1959, *The Great Ten Years*, People's Press, Beijing.

NBS, 1990, *A Revision the 1955 Regulations on the Structure Total Wage Bill*, Decree No. 1, 1990. Beijing: National Bureau of Statistics

NBS, 1992, *On the Structure of Total Labor Compensation Including Both Wage and Non-wage Incomes*, Circular No. 257. Beijing: National Bureau of Statistics

NBS and MOL [National Bureau of Statistics and Ministry of Labor], 1994, *Laodong tongji Zhibiao Jieshi ji Wenjian Xiuanbian* [Explanation of Labor Statistical Indicators and Selected Documents]. Beijing: China Statistical Publishing House.

SC/DRC (State Council/Development Research Center), 1962, Adjustment for Industrial Employment to Support Agricultural Production and Reduce Demand for Foodstuff, in SC/DRC, 2000, V.4 (II).

STA (State Taxation Administration), 2006, On Further Adjustment of Local Floating Band for Wage Rate. Beijing: State Taxation Administration

State Council, 1985a, 1985b, 1985c, *Circular on Regulation of Wage Bills*. Beijing: Sate Council

Wang, Haibo and Dong Zhikai, 1995, *Xin zhongguo gongye jingjishi* [History of Industrial Economy of the New China]. Beijing: Jingji Guanli Chubanshe

Wang, Yan and Yudong Yao, 2002, "Sources of China's Economic Growth 1952-1999: Incorporating Human Capital Accumulation," *China Economic Review*

Wu, Harry X. 2002, "Industrial Output and Labor Productivity in China 1949-1994: A Reassessment," in Maddison, A., W.F. Shepherd and D.S. Prasada Rao [eds.], *The Asian Economies in the Twentieth Century*, Edward Elgar Publishing Ltd., London

Young, Alwyn (2003), "Gold into Base Metals: Productivity Growth in the People's Republic of China during the Reform Period," *Journal of Political Economy*, 111, 1221–61

Yue, Ximing (2005), "Problems of Current Employment Statistics in China", *Economic Research Journal* (Jingji Yanjiu)

Zhu, Cishou, 1999, *Zhongguo gongye laodong shi* [Industrial Labor History of China]. Shanghai University of Finance and Economics Press

Table A1: Labor Input and Basic Labor Indicators in Chinese Industry, 1949-2009

Year	Volume of Labor (ml. 1987 yuan)	Labor Price Index (1987=1)	Outlay of Labor (ml. yuan)	Numbers (1000s)	Hours (ml.)	Weekly Hours per Person	Compensation per Hour (yuan)
1949	8,249	0.240	1,979	6,555	17,943	54.7	0.11
1950	8,667	0.253	2,191	6,795	18,703	55.0	0.12
1951	10,055	0.272	2,737	7,882	21,044	53.4	0.13
1952	11,738	0.307	3,608	9,896	23,998	48.5	0.15
1953	13,210	0.369	4,877	11,907	26,783	45.0	0.18
1954	13,578	0.412	5,590	13,402	27,822	41.5	0.20
1955	14,274	0.423	6,040	14,157	28,960	40.9	0.21
1956	15,262	0.496	7,576	13,940	30,033	43.1	0.25
1957	16,457	0.525	8,637	13,593	31,775	46.8	0.27
1958	33,244	0.353	11,722	21,064	64,518	61.3	0.18
1959	39,577	0.387	15,305	32,058	73,615	45.9	0.21
1960	33,691	0.480	16,169	32,020	61,331	38.3	0.26
1961	31,569	0.451	14,240	26,834	57,124	42.6	0.25
1962	23,821	0.504	12,006	22,062	43,258	39.2	0.28
1963	19,699	0.562	11,077	17,493	36,747	42.0	0.30
1964	19,288	0.576	11,114	16,063	36,811	45.8	0.30
1965	20,933	0.555	11,615	16,563	39,317	47.5	0.30
1966	23,056	0.538	12,414	17,740	42,641	48.1	0.29
1967	24,771	0.539	13,350	18,932	45,071	47.6	0.30
1968	26,075	0.525	13,701	19,727	46,387	47.0	0.30
1969	28,581	0.512	14,625	20,846	50,175	48.1	0.29
1970	33,570	0.480	16,101	23,455	58,043	49.5	0.28
1971	40,251	0.456	18,358	28,157	68,661	48.8	0.27
1972	45,037	0.464	20,915	32,031	76,301	47.6	0.27
1973	48,461	0.463	22,454	35,217	82,101	46.6	0.27
1974	51,432	0.454	23,356	37,811	87,252	46.2	0.27
1975	55,452	0.445	24,669	40,666	94,038	46.2	0.26
1976	60,603	0.436	26,415	44,562	102,940	46.2	0.26
1977	64,498	0.443	28,580	48,359	109,701	45.4	0.26
1978	71,405	0.469	33,480	53,566	119,361	44.6	0.28
1979	78,421	0.514	40,332	61,063	133,217	43.6	0.30
1980	84,670	0.586	49,577	66,867	145,159	43.4	0.34
1981	88,768	0.607	53,841	70,902	153,282	43.2	0.35
1982	92,166	0.629	58,008	74,600	159,921	42.9	0.36
1983	95,667	0.646	61,771	77,668	165,888	42.7	0.37
1984	101,127	0.759	76,727	81,406	175,090	43.0	0.44
1985	110,263	0.906	99,858	86,645	192,401	44.4	0.52
1986	118,111	1.037	122,422	92,868	206,186	44.4	0.59
1987	125,009	1.000	125,009	99,274	218,940	44.1	0.57
1988	130,469	1.243	162,180	104,767	228,539	43.6	0.71
1989	135,190	1.388	187,665	108,176	234,693	43.4	0.80
1990	137,590	1.743	239,783	109,335	236,923	43.3	1.01
1991	140,295	1.767	247,912	109,889	240,186	43.7	1.03
1992	143,303	1.898	272,059	111,107	243,592	43.8	1.12
1993	149,391	2.906	434,093	113,715	252,180	44.4	1.72
1994	151,475	3.962	600,142	115,882	255,961	44.2	2.34
1995	154,116	5.086	783,890	117,917	258,739	43.9	3.03
1996	157,044	6.925	1,087,510	120,708	263,730	43.7	4.12
1997	160,613	8.671	1,392,750	123,237	268,314	43.5	5.19
1998	155,917	9.173	1,430,220	122,729	263,688	43.0	5.42
1999	155,786	9.571	1,491,110	119,888	257,973	43.0	5.78
2000	150,684	11.241	1,693,810	116,857	253,706	43.4	6.68
2001	150,594	11.806	1,777,830	115,011	253,242	44.0	7.02
2002	150,205	12.392	1,861,390	112,488	253,699	45.1	7.34
2003	155,997	12.427	1,938,590	113,694	261,232	46.0	7.42
2004	168,075	12.615	2,120,210	122,650	281,496	45.9	7.53
2005	181,345	12.977	2,353,360	132,808	305,608	46.0	7.70
2006	195,222	15.271	2,981,180	141,065	329,212	46.7	9.06
2007	210,016	18.347	3,853,130	150,190	355,508	47.3	10.84
2008	224,520	20.214	4,538,350	156,805	377,992	48.2	12.01
2009	228,626	20.628	4,716,090	159,080	384,701	48.4	12.26

Table A2: Labor Input Indices in Chinese Industry, 1949-2009
(1987=100)

Year	Labor Input Index	Quantity (Hours) Index	Composition Index
1949	6.6	8.2	80.5
1950	6.9	8.5	81.2
1951	8.0	9.6	83.7
1952	9.4	11.0	85.7
1953	10.6	12.2	86.4
1954	10.9	12.7	85.5
1955	11.4	13.2	86.3
1956	12.2	13.7	89.0
1957	13.2	14.5	90.7
1958	26.6	29.5	90.2
1959	31.7	33.6	94.2
1960	27.0	28.0	96.2
1961	25.3	26.1	96.8
1962	19.1	19.8	96.4
1963	15.8	16.8	93.9
1964	15.4	16.8	91.8
1965	16.7	18.0	93.2
1966	18.4	19.5	94.7
1967	19.8	20.6	96.3
1968	20.9	21.2	98.4
1969	22.9	22.9	99.8
1970	26.9	26.5	101.3
1971	32.2	31.4	102.7
1972	36.0	34.9	103.4
1973	38.8	37.5	103.4
1974	41.1	39.9	103.2
1975	44.4	43.0	103.3
1976	48.5	47.0	103.1
1977	51.6	50.1	103.0
1978	57.1	54.5	104.8
1979	62.7	60.8	103.1
1980	67.7	66.3	102.2
1981	71.0	70.0	101.4
1982	73.7	73.0	100.9
1983	76.5	75.8	101.0
1984	80.9	80.0	101.2
1985	88.2	87.9	100.4
1986	94.5	94.2	100.3
1987	100.0	100.0	100.0
1988	104.4	104.4	100.0
1989	108.1	107.2	100.9
1990	110.1	108.2	101.7
1991	112.2	109.7	102.3
1992	114.6	111.3	103.0
1993	119.5	115.2	103.8
1994	121.2	116.9	103.6
1995	123.3	118.2	104.3
1996	125.6	120.5	104.3
1997	128.5	122.6	104.8
1998	124.7	120.4	103.6
1999	124.6	117.8	105.8
2000	120.5	115.9	104.0
2001	120.5	115.7	104.1
2002	120.2	115.9	103.7
2003	124.8	119.3	104.6
2004	134.5	128.6	104.6
2005	145.1	139.6	103.9
2006	156.2	150.4	103.9
2007	168.0	162.4	103.5
2008	179.6	172.6	104.0
2009	182.9	175.7	104.1