Protectionism and Regional Specialization: Evidence from China's Industries*

Chong-En Bai†, Yingjuan Du‡, Zhigang Tao§, and Sarah Y. Tong¶

February 2002

Abstract

This paper investigates the determinants of regional specialization using a panel data set covering 32 two-digit industries in 29 Chinese regions over a period of 13 years (1985–1997), paying particular attention to the role of regional protectionism. It is found that there is less geographic concentration in industries where past profit and/or tax margins are high and where the share of employment by the state-owned enterprises is high, reflecting stronger local government protection of these industries. The evidence also supports the external-economies theory and the increasing-returns-to-scale theory, but not the resource-endowment theory, of regional specialization. Finally, the overall time trend of regional specialization of China's industries is found to have reversed an early drop in the mid 1980s, and registered a significant increase in the later years.

Key Words: protectionism, regional specialization, scale economy, external economy, resource endowment.

JEL Classification Number: F1, R12, P2

*We are grateful to Songnian Chen, David Li, and Richard Wong for their helpful comments.

†School of Economics & Finance, The University of Hong Kong, Pokfulam Road, Hong Kong. Tel: (852) 2859-1036. Fax: (852) 2548-1152. baic@hku.hk.

‡School of Business, The University of Hong Kong.

§School of Business, The University of Hong Kong.

¶Hong Kong Institute of Economics & Business Strategy and School of Economics & Finance, The University of Hong Kong.

1 Introduction

Trade facilitates specialization, which in turn leads to more gains from trade. To understand the pattern of trade among geographic units, one needs to investigate the determinants of international/regional specialization. For this reason, the study of geographic concentration in production has been an important area of research in both international economics and regional economics. Much of the empirical literature on this topic, however, is carried out using sub-national data, and hence focusing on the regional specialization of economic activities (see Hanson (2001), and Overman, Redding, and Venables (2001) for most recent surveys). Such an approach has two advantages. One is that comparable data is more readily available for sub-national units and the other is that it avoids the difficulty of controlling institutional differences across countries in international studies (Davis, Weinstein, Bradford and Shimpo (1997), Bacchetta, Rose, and van Wincoop (2001), and O'Connell and Wei (2002)).

A number of theories, mostly originated from the study of international trade and specialization, have been proposed to account for regional specialization of economic activities. One theory emphasizes the regional disparity in resource endowments (Ohlin, 1933). Second, for industries that enjoy increasing returns to scale, there is a natural tendency to have production clustered in a few places as opposed to scattered production in many places (Krugman, 1991). Third, even for industries that exhibit constant or decreasing returns to scale, it is possible that a firm's cost of production (or its ability to introduce new products and services) is reduced (or enhanced) by the presence in the same region of other firms in the same industry. Such spillover effects or external economies could then lead to the geographic concentration of production (Marshall, 1920). Using the U.S. data on regional specialization from 1860 to 1987, Kim (1995) finds support for the resource-endowment theory and the increasing-returns-to-scale theory, but not for Marshall's external-economies theory.

While the benefits of trade and specialization are well understood, a pre-condition for realizing these benefits — namely, free flow of goods and services across regions and countries — is not always satisfied due to possible protectionism at both international and regional levels. Therefore, the role of protectionism in determining specialization deserves careful study. In this regard, sub-national studies have not made much contribution, because in many countries, such as the United States, interregional trade barriers are prohibited by the national government and therefore local protectionism is not a factor. The case of China is different and it provides us with a unique opportunity to study the role of protectionism in regional specialization. China's economic reform since 1978 has introduced fiscal decentralization, which provided the local governments with a strong incentive to protect their tax base by shielding local firms and industries from interregional competition. The local governments also have incentives to protect local employment, especially that of state-owned enterprises. Meanwhile, there was no promulgation in the early years of economic reform, and no effective implementation in the later years, of a central-government policy that prohibits interregional trade barriers. Therefore, local protectionism is an important factor in China's regional specialization.

There is considerable controversy about the degree of local protectionism in China. Young (2000) provides anecdotal evidence on the rise of local protectionism in China during the reform era. He also supports his conclusion by evidence of reduced regional specialization based on the evolution of the five sectors in the socialist measure of national income (agriculture, industry, construction, transport, and commerce) and on the evolution of the three sectors in GDP accounting (primary, secondary, and tertiary). Naughton (1999), on the other hand, uses data from the input–output tables among Chinese provinces in 1992, and finds an increase in regional specialization in 1992 over that in 1987. A systematic study on the role of local protectionism as a determinant of regional specialization and a further investigation on the trend of regional specialization in China would shed useful light on this controversy.

We construct a panel data set of 32 two-digit industries in 29 Chinese regions[1] over the period of 13 years (1985–1997). Our data on regional specialization are more disaggregated than those used by Young (2000), and cover a longer and more recent time period than those used by Naughton (1999). Using the data, we study not only the overall trend but also the determinants of regional specialization. We pay particular attention to the role of local protectionism, in addition to those of resource endowment, increasing returns, and external economies. Specifically, it is conjectured that local gov-

[1]The sample includes 29 provinces, autonomous regions, and municipalities directly under the central government. Hanan gained the status of a province in 1988. However, its data are included in Guangdong province in this study.

ernments tend to protect industries that yielded high profit and/or tax in the past, thereby reducing the geographic concentration in those industries. Local protectionism is also expected to be significant for industries with high percentages of employment in state-owned enterprises.

Our empirical study finds strong support to our hypotheses about local protectionism. Other things being equal, regional specialization is found to be low for industries that yielded high profit and/or tax in the past, and for industries with high percentages of employment in state-owned enterprises. Our study also lends support to the increasing-returns-to-scale theory and the external-economies theory of geographic concentration. However, we find no evidence supporting the resource-endowment theory, despite the significant regional disparity in resource endowment in China. Presumably, as transportation costs decrease over time, the key condition for the resource-endowment theory — relatively immobile resources — no longer holds. Measurement problems related to resource endowment are another probable source of this result. Finally, the overall time trend of China's regional specialization of industrial production has reversed an early drop in the mid 1980s, and registered a significant increase in the later years, suggesting the increasing dominance of the conventional economic forces behind regional specialization over the specific ones for local protectionism.

The remainder of the paper is organized as follows. In Section 2, we discuss the theories of regional specialization in more detail and develop hypotheses based on them. In Section 3, we construct some variables for the testing of various hypotheses. Descriptive statistics of key variables are offered and compared with some of the findings in the existing literature. Section 4 presents econometric testing of the hypotheses and assesses the relevance of various theories in the context of China. The paper concludes with Section 5.

2 Theories and Hypotheses

Regional specialization of industrial production within a country share many common features with international specialization and has received considerable attention in the study of international trade and regional economics. Theories based on resource endowment, increasing returns to scale, and external economies have been proposed to account for regional specialization of industrial production within a country, and their predictions have been tested. In the first part of this section, we summarize these three theories and develop testable hypotheses based on them. The transition economy of China offers us a unique opportunity to study the role of local government protectionism in regional specialization. The second part of this section develops hypotheses arising from this consideration.

The first theory of regional specialization of industrial production within a country is a natural extension of the theory of international trade and specialization (Ohlin, 1933). Different countries are endowed with different sets of natural, physical, and human resources. In the extreme case where each country is an isolated island, it has to be self-sufficient by producing all goods and services in demand. Then, there is no trade among various countries and hence no specialization of production. When trade among different countries is possible, however, each country produces a subset of goods and services, and trades with the other countries. The pattern of specialization is determined by the comparative, not absolute, advantages of countries implied by resource endowments and technological capabilities. Assuming away any technological difference among countries, a country with relatively abundant labor supply focuses on labor-intensive industries, while another country with relatively abundant natural resources specializes in industries requiring extensive inputs of natural resources. The specialized production in conjunction with the demand profile determine the import–export patterns among various countries. It is important to note that this theory is based on a crucial assumption that factors of production are immobile.

The same logic may apply to regional specialization within a country. In particular, regional specialization of industrial production would emerge, provided that endowment of immobile resources varies from one region to another. To develop a testable hypothesis, we investigate the use of immobile resources by various industries in China. If industries with heavy employment of immobile resources are concentrated in a few regions, then we have support for the resource-endowment theory of regional specialization. Hence we have the following testable hypothesis:

Hypothesis 1: Industries with heavy employment of immobile resources are geographically concentrated.

The second theory of regional specialization — a direct extension of the new trade theory — is based on the technological feature of increasing returns to scale (Krugman, 1991). Increasing returns to scale arise when there is a significant fixed cost of production or there is a decreasing average variable cost of production. A firm in such an industry would enjoy a low average cost of production by producing a large volume of goods and services, which further enhances the firm's advantage in the marketplace. A small lead in the volume of production would give rise to a huge advantage in the dynamic competition. The positive feedback eventually leads to a high concentration of production. It is economically more efficient to have the production carried out in a few places than to scatter production in every region of the economy. The hypothesis arising from the increasing-returns-to-scale theory is as follows.

Hypothesis 2: Geographic concentration is more likely in industries that exhibit increasing returns to scale.

There is an element of indeterminacy in the increasing-returns-to-scale theory of regional specialization, in the sense that every region has the potential of being the location of large volume of production so long as its initial level of production is sufficiently high. This is in contrast to the first theory of regional specialization, in which resource endowment determines the specific locations for production of goods and services.

The third theory of regional specialization is that of external economies (Marshall, 1920). Marshall argued that there are three main channels through which the presence in the same region of other firms in the same industry may exert positive spillover effects: a cluster of an industry can support specialized suppliers, it allows labor-market pooling, and it helps foster knowledge spillover (Krugman and Obstfeld, 2000). The first two channels imply that a firm's cost of production is reduced by the presence in the same region of other firms in the same industry. For example, if there were only one investment bank in New York City, it would be very expensive or even impossible for the bank to recruit an expert in derivative trading, because of possible firm-level shocks. A geographic cluster of firms in the same business offers an incentive for the provision of specialized inputs including human capital, which in turn make the firms in the cluster more cost competitive than firms that are not affiliated with any cluster. Such positive feedback eventually leads to regional specialization of industrial production (Enright, 1990).

An alternative and complementary mechanism for the external-economies theory of regional specialization is that a firm is more likely to develop new products and services in the presence of other firms in the same industry and in the same region, the third channel identified by Marshall. Experiences from Silicon Valley in California suggest the importance of informal exchange of ideas among people of different firms in the same industry and in the same region. Because much of the knowledge generated in a firm is tacit and difficult to document, formal exchange across regions is not effective. In addition, such tacit knowledge from one firm is often complementary to that from another firm in the same industry. Hence new ideas can sprout up when there is informal exchange of ideas among people of different firms in the same industry and in the same region. For empirical tests of the theory, please see a widely-cited study by Rauch (1993).

We summarize the above discussion with the following hypothesis.

Hypothesis 3: Regional specialization is predicted for industries that enjoy significant external economies.

As in the increasing-returns-to-scale theory, there is also an element of indeterminacy regarding the location of specialized production in the external-economies theory of regional specialization, in the sense that every region has the potential of being the location of the specialized production so long as it starts with a sufficiently high level of concentration. In the case of Silicon Valley, the presence of some top-level universities may have played the key role in forming the clustering of the information-technology companies.

Underlying each of the above three conventional theories of regional specialization is the assumption of interregional trade of goods and services. If each region were an isolated island, then there would not be any specialization in industrial production among the regions, even if there were significant disparity in resource endowment, or increasing returns to scale in certain industries, or external economies. In general, there is interregional trade in goods and services, the ease of which, however, depends on the degree of local protectionism, among other factors.

It should be recognized that local governments in almost every country, whether it is economically developed or still developing, have the incentive to protect their local industries. This is because local governments rely on their local industries for tax revenue. They also care about local employment, which is important for elections in economically developed economies and for social stability in transition economies (Bai, Li, Tao, and Wang, 2000). To ensure a solid tax base and maintain local employment, local governments can erect various barriers of trade to protect local industries from interregional competition. This problem is similar to the protectionism in international trade, with one crucial difference. Compared with international trade among countries, it should be relatively easy to ensure smooth interregional trade as the national government does have authority over local governments. In the United States, the constitution prohibits interstate tariffs. This has greatly facilitated interregional trade of goods and services and led to regional specialization of industrial production.

During the economic transition in China, anecdotal evidence suggests that there is substantial flow of goods and services among regions, though local protectionism has been from time to time a serious problem. The main force behind local protectionism arises from some mismatch in the economic policies during the reform era since 1978. Prior to the economic reform in 1978, China had a highly centralized fiscal system. All the tax revenue collected had to go first to the central government. The planning commission of the central government had the authority to decide the expenditure of the local governments and allocate revenue from the central pool (Qian, 2000). Such a system delinked tax revenue and expenditure at the level of local governments, and provided little incentive for local protection or local production. Since 1978, fiscal decentralization has been introduced, which allows the local governments to retain a percentage of the revenue collected and therefore provides them with a strong incentive to protect local industries. What is lacking in the fiscal reform is the promulgation in the early years and effective implementation in the later years of a policy that prohibits barriers to interregional trade.

To develop testable hypotheses regarding local protectionism, we focus on the local governments' benefits from erecting barriers of interregional competition, and figure out which industries the local governments would like to protect. First of all, given the incentive from tax revenue, it is conjectured that the local governments want to protect industries that have high tax margins. In transition economies such as China, there remains significant state ownership in most industries. Being the holders of residual rights of control, if not the residual claimants of income, the local governments also care about the profits of state-owned enterprises. Furthermore, due to the lack of rule of law, even profits of privately owned enterprises are subject to various degrees of expropriation, in the form of ad hoc taxes and fees, by the local governments. Note that it may take time for the local governments to figure out which industries have high profit and/or tax margins. Thus, we have:

Hypothesis 4: Geographic concentration is low for those industries that had high profit and/or tax margins in the past.

Second, the local governments are concerned with employment in their respective regions of governance. While the benefits of maintaining local employment are quite obvious and even universal across countries of various degrees of economic development, we would like to stress some unique features of transition economies such as China. Introduced to address the inefficiency of centrally-planned economy, market reform in China since 1978 has managed to improve the efficiencies of some state-owned enterprises but also exposed the inefficiencies of many others. Indeed it is widely acknowledged that enormous amount of surplus labor exists in state-owned enterprises of all types of industries in China. For the sake of social stability, the local governments as well as the central government are compelled to maintain employment of these workers (Bai, Li, Tao and Wang, 2000). Furthermore, most state-owned enterprises are administered by the local governments and are important political constituents of local government officials. Because of these reasons, in face of interregional competition, the local governments are receptive to calls for protecting industries in which significant employment is in state-owned enterprises. Hence we have:

Hypothesis 5: Regional specialization is low for industries with high percentages of employment in state-owned enterprises.

3 Data and Measurement

In this section, we first discuss how to construct a measure of regional specialization of industrial production. Then we define and measure other variables that will be used for testing the five hypotheses discussed in the previous section. Finally, we present some summary statistics. In particular, we discuss the general trend of regional specialization in China, and comment on the related work by Naughton (1999) and Young (2000).

3.1 A measure of regional specialization

One way to measure regional specialization is to quantify the interregional trade patterns resulting from specialization. This approach is widely adopted for studying division of labor and specialization in the global economy. Compared with trade among different countries, however, data on interregional trade within a country are difficult to come by. Hence, in this paper, we take a more direct approach to measuring the degree of geographic concentration in various industries: namely, mapping out the geographic distribution of production activities in each industry and normalizing it with the geographic distribution of overall industrial activities.

There are two variables indicating the magnitude of production activities: output value and employment level. Output data of 32 industries in 29 regions in current prices are obtained from: the *China Statistical Yearbook* for 1985–1987, the *China Statistical Yearbook on Industrial Economy* for 1988–1994 and 1997, and the *China Industrial Census* for 1995.[2] Employment data of 32 industries in 29 regions are only obtained for 1988-1994 and 1997 from the *China Statistical Yearbook on Industrial Economy.* The employment data for other years are disaggregated only to the level of industry, not the level of industry by region. Overall, for each region and each industry, we have a time series of output for 1985–1995 and 1997, and a time series of employment for 1988–1994 and 1997.

With the output data, we can construct a measure of regional specialization called *Hoover coefficient of localization* (1936).[3] It is based on the location quotient with

[2] While the data are obtained from different statistical yearbooks, they are all compiled by the same China Statistics Bureau and are supposed to follow a common set of statistical criteria. In general, the most detailed industry-by-region data are provided by the *China Statistical Yearbook* in the early years, but by the *China Statistical Yearbook on Industrial Economy* in later years. There are a few exceptions, mostly for data on tax plus profit margins and share of SOE employment, which will be discussed in details. For 1995 and 1996, there was no publication of the *China Statistical Yearbook on Industrial Economy*, the reason being the publication of the *China Industrial Census* of 1995 (private communications with officials at China Statistics Bureau), the data of which are subsequently used. For a summary of data sources, please refer to the Data Appendix.

[3] As there are limited firm-level data on Chinese industries, we cannot use the localization index recently developed by Ellison and Glaeser (1997).

respect to output, which is defined as

$$L\_OUTPUT_{ij} = \frac{OUTPUT_{ij}}{OUTPUT_i} \Big/ \frac{OUTPUT_j}{OUTPUT}$$

where $OUTPUT_{ij}$ is output of industry i in region j, $OUTPUT_j$ is total output in region j, $OUTPUT_i$ is total output of industry i, and $OUTPUT$ is total output of China. If L_OUTPUT_{ij} is larger than one, then region j has a higher percentage of industry i than of total industrial output. Similarly, if L_OUTPUT_{ij} is smaller than one, then region j has a lower percentage of industry i than of total industrial output. Given the location quotients of industry i for all regions $j = 1, ..., R$, we rank regions by their location quotients in descending order and get a sequence of regions. Then, following that sequence, we calculate the cumulative percentage of output in industry i over the regions (y-axis) and the cumulative percentage of output in all industries over the regions (x-axis), and thus plot the localization curve for industry i. If the industry is evenly distributed across regions, then the location quotient will be equal to one for all regions, and the localization curve will be a 45-degree line. If the industry is more regionally concentrated, then the localization curve will be more concave. Analogous to the Gini coefficient for income distribution, the coefficient of localization is defined as the area between the 45-degree line and the localization curve divided by the entire triangular area. The higher the value of the Hoover coefficient, the more localized the industry is.

A Hoover coefficient of localization can also be constructed from the employment data. In fact, this approach is used in a number of studies on the regional specialization of economic activities in the United States (Kim, 1995; Ellison and Glaeser, 1997; Dumais, Ellison, and Glaeser, 2001). In this paper, we construct a Hoover coefficient using output data in addition to that using employment data, for two reasons. First, using employment data may result in biases, as state-owned enterprises (SOEs) in China have a common problem of redundant workers, and shares of SOEs are different across regions and across industries. The second reason is that we lose four years of data when employment is used to measure concentration.

3.2 Other variables

Next we turn to the challenge of finding variables for testing the five hypotheses discussed in the previous section: resource endowments, scale economies, external economies, and local protectionism in a transition economy.

A. Resources

An implicit assumption for the resource-based theory of regional specialization is that certain resources are immobile or their transportation costs are high. Thus it is more cost-efficient to locate an industry requiring extensive inputs of these resources in regions that are rich in them, which consequently leads to the geographic concentration of the industry. To test the hypothesis that resource-based industries tend to be localized (Ohlin, 1933), however, we need to find an appropriate measure of those immobile resources.

In a study on regional specialization in the United States, Kim (1995) uses the cost of raw materials divided by total value added as the measure of resource intensity. Note that the measure is a ratio of the value of all inputs to the industry's total value added. However, not all inputs are equally immobile; thus the measure used by Kim may not reflect the industry's true dependence on immobile resources. To illustrate this point, consider China's electronics industry, which uses expensive inputs such as embedded chips for low-value-added OEMs. According to Kim's measure, the resource dependence rate is very high, but the inputs involved, such as the embedded chips, are highly mobile. For the transition economy of China there is another drawback with Kim's measure, as the raw materials are often under government price control and therefore undervalued. Take for example China's tobacco industry. The prices for the raw materials are kept low due to the government policy of supporting industrial development at the expense of rural development (the problem of price scissors as studied by Sah and Stiglitz, 1984). Thus Kim's measure of resource intensity would be low for China's tobacco industry, though the actual degree of resource dependence is very high.

To demonstrate the potential problems with the use of Kim's measure for China, we compute the raw-material intensity, defined as the difference between output and value added divided by value added, for the year 1992. The data on output were the ones used in the construction of the Hoover coefficient. The data on value added were obtained from the *China Statistical Yearbook on Industrial Economy.*

In Table 1, we list the values of Kim's measure for the 32 industries in China. Highly resource-based industries such as coal mining and processing, metals mining and processing, nonmetal minerals mining and processing, petroleum and natural gas extraction, and tobacco processing have low values of Kim's measure. In contrast, less resource-based industries such as food processing and production, garments and other fiber products, and textiles have high values of Kim's measure. We conclude that this measure is not appropriate for Chinese industries.

As an alternative, we use energy consumption intensity to indicate an industry's dependence on resources. It is defined as the ratio of total energy consumption, measured in tons of standard coal, to total output. Coal is the most important energy source for industries in China. In addition, freight transportation of coal in China has been expensive. We therefore believe that energy consumption intensity is a more appropriate measurement for an industry's dependence on immobile resources than the measure used by Kim (1995). Using energy consumption data of 32 industries obtained from the *China Industrial Census* for 1995, we calculate the energy consumption intensities of the 32 industries and denote the variable by $ENERGY$. Unfortunately, data for other years cannot be found. This method is not without its own problem as more diverse energy sources are used in more recent years and some new energy sources are more mobile than coal. However, there does not seem to be any better alternative that is available.

B. Scale economies

To test the hypothesis that industries characterized by increasing returns to scale should be geographically concentrated, we use average firm size in an industry, denoted by $SCALE$, as a measure of scale economies. Consistent with our measure of geographic concentration, both employment and output data are used to calculate the average firm size. Data on output and number of firms at the industry level are obtained from the *China Statistical Yearbook on Industrial Economy* for 1986-1994 and 1997, and from the *China Industrial Census* for 1995. As in the construction of Hoover Coefficients, employment data are available only for 1988-1995 and 1997 from the *China Statistical Yearbook on Industrial Economy.*[4] When calculating the average firm size with the

[4]Though the *China Statistical Yearbook* provides the most updated industry-by-region data for the early years of the sample as stated in footnote 2, the 1993 *China Statistical Yearbook on Industrial*

output data, we need to construct the price deflator[5] for industrial output and use it to obtain the output value in constant terms. Two panel data sets are constructed across 32 industries for 1986–1995 and 1997, when output data are used, and for 1988-1995 and 1997, when employment data are used.

C. External economies

External economies, through specialized input suppliers, labor-market pooling and knowledge spillover, are generally difficult to measure directly. In his study on the geographic concentration in the U.S. industries, Kim (1995) bypasses this problem by ruling out the relevance of external-economies theory in view of the following three observations and the belief that the degree of external economies is positively related to the intensity of R&D activities. First, the overall level of localization in the U.S. decreased between World War II and 1987, while its overall R&D intensity was rising. Second, the tobacco and textile industries had become more localized while their R&D, skill intensity, and rates of technological innovations fell. The opposite is true for industries like electrical machinery and transportation. Third, localization levels for high-tech industries were lower than those for low-tech industries such as tobacco and textiles.

In this paper, we plan to carry out a more systematic examination of the effect of external economies on regional specialization. We do this for several reasons. First, in the period 1985–1997, the overall level of localization first decreased slightly and then increased significantly. (Details will be given in Section 3.3.) In addition, Kim's second and third observations are based very much on the specific example of the tobacco industry, which is highly resource-based in the case of China, and therefore are not very convincing. Indeed, many low-tech industries are highly resource-dependent. Finally, whether or not external economies have an effect on regional specialization should be judged through econometric analysis.

We adopt the same belief about the relationship between external economies and R&D as Kim and use R&D intensity as a proxy for of external economies. The theory of external economies would predict that industries with high R&D intensity should be

Economy gives the historical data on industry-level employment, output and number of firms, which are suitable for constructing variables for scale economies and are subsequently used due to their consistency.

[5]The price deflator for year t is calculated as $IND\_curt/(IND\_cur1978 \times INDEX_t)$, where IND_curt denotes gross domestic industrial product in current price for year t, and $INDEX_t$ is the index of gross domestic industrial product for year t in comparable prices ($1978 = 100$).

geographically concentrated. We obtain data on R&D expenditure for the 32 industries from the *China Industrial Census* for 1995 and compute the R&D intensity of an industry, denoted by RD, by dividing the total R&D expenditure by the total output of the industry. Unfortunately, data for other years cannot be found, and consequently we cannot investigate how the intertemporal change in R&D intensity affects the geographic concentration of industrial production.

D. Tax plus profit margin

Under central planning, most firms were state owned and both their profits and tax payments were counted as government revenue. As a legacy of central planning, the official statistics only reported tax plus profit as a combined item and did not report their separate figures for a number of years in our sample period. Data on *tax plus profit* and *sales* for the 32 industries are obtained from the *China Industrial Census* for 1995, and the *China Statistical Yearbook on Industrial Economy* for 1985–1994 and 1996.[6] A panel data set of the tax plus profit margin, defined as tax plus profit divided by total sales and denoted by TPM, is constructed across the 32 industries and for the period of 1985–1996. In our econometric analysis we use lagged TPM as a measure of the local government's incentive to protect an industry, because it takes time for the local government to figure out which industry is worth protecting. The use of the lagged value also mitigates the potential endogeneity problem associated with TPM.

E. Share of SOE employment

Unexpectedly, data for calculating the share of SOE employment are the most difficult to come by. As summarized in the Data Appendix, for 1986-1988, data on SOE employment and COE (collectively-owned enterprises) employment, and their combined share in the total employment are available from the China Statistical Yearbook. The above data are used to calculate the total employment in an industry. The share of SOE employment, defined as SOE employment divided by total employment and denoted by $SSOE$, follows immediately. For 1993-1994, SOE employment is not provided, but it can be calculated from SOE value added and SOE productivity (defined as SOE value

[6] 1996 data on tax plus profit and sales are obtained from the 1998 edition of *China Statistical Yearbook on Industrial Economy*, which covers 1996 and 1997. Unfortunately, in the latest edition, there is no data at the industry-by-region level on output and employment for 1996, which are essential for constructing the Hoover Coefficients. Given our use of lagged TPM, however, the availability of TPM_96 is still a plus.

added divided by SOE employment), both of which are available from the China Statistical Yearbook. As the total employment in an industry is also available, the share of SOE employment can be calculated. Finally, for 1988-1992, 1995 and 1997, data on both SOE employment and total employment are available from, respectively, China Statistical Yearbook, China Industrial Census, and China Statistical Yearbook on Industrial Economy. Hence the share of SOE employment can be readily calculated.

3.3 Summary statistics

Hoover coefficients of localization for the 32 two-digit industries over time are given in Table 2a (constructed using output data) and Table 2b (constructed using employment data). The tables provide information on the pattern of regional specialization for each industry over time. One way of examining the data is to trace the time trend of all industries as a whole. The bottom two rows of the tables are the simple averages and weighted averages (weighted by the output values or employment levels of industries, respectively) over all industries. Let's first examine the results on the Hoover coefficients calculated with output data (Table 2a). The simple average of the coefficient of localization was 0.3665 in 1985. It went down slightly till 1989 and then rose steadily to 0.4055 in 1997. The trend is similar for the weighted average. The weighted average was 0.3117 in 1985. It decreased in value to 0.3056 in 1988 and then increased for all later years. We then look at results on Hoover coefficients calculated using employment data (Table 2b). The overall trend is very similar to that of the output-based Hoover coefficients.

As shown in Figures 1a (output-based) and 1b (employment-based), the aggregate coefficients indicate that, over the 13-year period 1985–1997, regional specialization of Chinese industries increased quite substantially. Our results are in sharp contrast to those in Young (2000) but are consistent with those in Naughton (1999). Young offers anecdotal evidence on the rise of regional protectionism in China since the economic reform in 1978. His statistical evidence is based on the evolution of the five sectors in the socialist measure of national income (agriculture, industry, construction, transport, and commerce), and the evolution of the three sectors in GDP accounting (primary, secondary, and tertiary). Naughton (1999) uses data from the input–output tables among

Chinese provinces in 1992, and finds an increase in regional specialization in 1992 over that in 1987. Our data on regional specialization are more disaggregated than those of Young (2000), and cover a longer and more recent time period than those of Naughton (1999).

Another way of examining the data in Tables 2a and 2b is to compare the cross-time averages for various industries. The second (output-based) and third (employment-based) columns in Table 3 give the cross-time average concentration for each industry. There are large variations across industries in the level of localization. Take, for example, the output-based Hoover coefficient. It ranges from 0.2083 (metal products) to 0.8749 (logging and transport of timber and bamboo). Mining industries, which depend heavily on resources, are more localized than manufacturing industries: the average Hoover coefficient for mining industries over the 13-year period is 0.6558, while that for manufacturing is only 0.3278.[7] Even within manufacturing industries, there exist significant differences. Tobacco processing is the most localized, followed by stationery, educational and sports goods, and electronics and telecommunications. Metal products, machinery and equipment manufacturing, and raw chemical materials and chemical products are the three least localized industries.

Table 3 also lists the (cross-time average) values of other variables used in the analysis. Regarding the measures for the three conventional theories of regional specialization (R&D intensity, energy consumption, and average firm size), Table 3 shows that there are also large variations across industries. We first look at the R&D intensity. Petroleum and natural gas extraction has the highest R&D intensity. Its value (209 yuan of R&D expenses per 10,000 yuan of output) is more than five standard deviations larger than the mean value (22 per 10,000). On average, mining industries have higher R&D intensity than manufacturing industries. But if we exclude petroleum and natural gas extraction, then mining industries actually have lower R&D expenditure than manufacturing industries. Some manufacturing industries have rather high R&D intensity, such as transportation equipment manufacturing (48 per 10,000) and medical and pharmaceutical products (47 per 10,000). Furniture manufacturing, stationery, educational and sports goods, and metal products have the lowest R&D intensities. Second, the average

[7] The figures are obtained by taking simple averages of relevant data from Table 3. Mining industries include industries 6 to 12; manufacturing industries, 13 to 42.

energy consumption of all 32 industries is 4.68 tons of standard coal per 10,000 yuan of output. Electric power, steam, and hot water production and supply has the highest energy consumption (21.17), and electronics and telecommunications has the lowest (0.31).

Recall that the average firm size can be constructed using either employment or output data. The average firm size of all industries in terms of employment is 865. Petroleum and natural gas extraction has the largest firm size among the industries: the average number of employees per firm is 19,314, which is five standard deviations more than the mean value. Industries with large firm size are logging and transport of timber and bamboo, tobacco processing, and chemical fibers: their average firm sizes are 1,110, 891, and 648, respectively. The average firm size is less than 100 for the following seven industries: furniture manufacturing (63 employees), timber processing, bamboo, cane, palm fiber and straw products (80), tap water production and supply (83), food processing and production (87), printing and record pressing (90), plastic products (97), and metal products (99). The average firm size in terms of output gives a similar ranking of the industries; in particular, the highest (petroleum and natural gas extraction) and lowest (furniture manufacturing) are the same as those in terms of employment.

Finally, we discuss the two variables used for testing the hypotheses on local protectionism, TPM (tax plus profit margin) and $SSOE$ (share of SOE employment). As shown in Figure 2, the weighted average of $TPMs$ across all industries first underwent a dramatic decrease from 21.0% in 1985 to 11.6% in 1990, and then stabilized for a while until another drop to 9.1% in 1996. This is a result of the economic reform that began in late 1978. Between 1949 and 1978, the Chinese economy was characterized by a system of central planning. Two important manifestations of central planning were the lack of competition and the suppression of factor prices, both of which implied high profit margins for industrial production. The economic reform since 1978, however, has unleashed forces that have increased competition and raised the factor prices, resulting in lower profit margins for industrial production. The phasing-out of central planning has made it easier for both local governments and private entrepreneurs to enter various industries, increasing the competitive pressure in the product market. Meanwhile, the restrictions on prices of various inputs have gradually been eliminated, resulting in higher and more volatile market prices, which increase the cost for most industrial production. The stable profit margins since 1991 signal the maturing of the competitive markets in China.[8]

The time trend for the shares of SOE employment is quite similar to that of tax plus profit margins. As shown in Figure 3, the (cross-industry weighted average) share of SOE employment first decreased from 78.1% in 1986 to 72.7% in 1987 and then stabilized at the 71-72% level until 1992, when it plunged to 55.3% in 1993 and subsequently stabilized at the 50% level. The figures depict the commonly held perception that the state sector has declined substantially during the reform era. This illustrates the inefficiency of state-owned enterprises in face of increasing competition.

Despite the clear-cut time trends of TPM and $SSOE$, Table 3 shows that there remain significant differences in these two variables across different industries. The average tax plus profit margin was 14.0%. The industry with the highest TPM was tobacco processing (55.9%). Electric power, steam and hot water production and supply, and petroleum processing, coking products, and gas production came next, at 23.3% and 20.7% respectively. The industries with the lowest $TPMs$ were coal mining and processing (1.2%), food processing and production (6.06%), and leather, furs, down, and related products (6.12%). The average share of SOE employment was 61.3%. Industries with the highest $SSOEs$ were petroleum and natural gas extraction (98.1%), tap water production and supply (92.7%), and electric power, steam, and hot water production and supply (92.6%). Industries with the lowest $SSOEs$ were garments and other fiber products (12.1%), furniture manufacturing (14.2%), and plastic products (21.0%).

4 Regression Analysis

In this section, we carry out econometric tests of our hypotheses. As discussed in Sections 3.1 and 3.2, a panel data set for 32 industries and 13 years (1985–1997) has been constructed for the following variables: Hoover coefficient ($HOOVER$), average firm size ($SCALE$), tax plus profit margin (TPM), and share of SOE employment ($SSOE$).

[8] Without direct evidence on the time trend of tax margins, we focus mainly on that of profit margins, the data of which are directly available in the later years of our sample.

We do not, however, have panel data on R&D intensity (RD) and energy consumption intensity ($ENERGY$): such data are only available for 1995.

We use the following panel structure to test the five hypotheses listed in Section 2:

$$\begin{aligned} HOOVER_{it} = \; & \beta_0 + \beta_1 RD_i + \beta_2 ENERGY_i + \beta_3 SCALE_{it} \\ & + \beta_4 TPM_{i\ t-l} + \beta_5 SSOE_{it} + \alpha_i + \varepsilon_{it} \end{aligned}$$

where α_i is the industry-specific effect, and l indicates the years of lag that is used in the regression. In our estimation, l ranges from 1 to 4 and the number of observations included changes accordingly. It should also be noted that two sets of regressions are carried out separately for the Hoover coefficient calculated with output data and that with employment data.

As discussed earlier, we were only able to obtain information on R&D intensity and energy consumption intensity for one year. As a result, we can't estimate the equation using the fixed-effect method. To overcome this problem, we adopt the following approach. First, we estimate a random-effect model including all five variables. Second, we estimate a fixed-effect model in two steps. In the first step, we only include three variables with panel data, $SCALE$, TPM, and $SSOE$, and save the estimated intercepts for different industries. These estimates are then used as dependent variables in the second step of regression, where the two remaining variables, RD and $ENERGY$, are used as independent variables. Third, we carry out a Hausman test to decide which model, random effect or fixed effect, is preferred.

Table 4 summarizes the random-effect estimation results. In Table 4a, the results are obtained when the dependent variable, the Hoover coefficient of localization, is calculated using output data. The results provide weak support for the external-economies theory of regional specialization. The coefficients on RD (R&D intensity) are all positive. Industries with high R&D expenditure are more localized. The coefficients are only marginally significant at the 10% level and 15% level, respectively, when 1-year and 2-year lags are used for TPM. When the lag is 3 or 4 years, the coefficient for RD is no longer significant. However, there is strong support for the increasing-returns-to-scale theory. The regression coefficients on $SCALE$ (the average level of output per firm in constant price) are positive, and significant at the 5% level when 2-year lag is used for

TPM, at the 15% level when 1-year lag is used for TPM, and at the 1% level in the other two specifications. The empirical results, however, provide little support for the resource-based theory of regional specialization. Although the regression coefficients on $ENERGY$ (energy consumption intensity) are all positive for all specifications, they are never significant. It is possible that energy consumption intensity is not a good proxy for an industry's dependence on immobile resources. More generally, as transportation becomes less costly, the key assumption for the resource-endowment theory — immobile resources — may no longer hold.[9]

The results in Table 4a provide very strong support to the two hypotheses on local protectionism, namely, regional specialization is low for industries with high tax plus profit margins, and for industries with high percentages of SOE employment. The regression coefficients on past TPM with lags of one, two, three or four years are negative and significant at the 1% level. The coefficients on $SSOE$ (the share of SOE employment) are negative for all specifications: significant at the 1% level when 1-year or 2-year lag for TPM is used, and significant at the 5% level when 3-year or 4-year lag of TPM is used.

In estimating the models in Table 4b, we use the Hoover coefficient of localization calculated with employment data ($HOOVER\_Employment$). Accordingly, the average firm size ($SCALE\_Employment$) is also obtained using employment data. The results are, for the most part, consistent with those in Table 4a. The coefficients on RD are positive for all four specifications, but none is statistically significant. The coefficients on $ENERGY$ are positive except when 4-year lag of TPM is used, but none of the coefficients is significant. Among the three conventional theories on regional specialization, the strongest evidence is for the scale-economy theory. The coefficients on $SCALE$ are positive, and significant at the 5% level when 1-year or 2-year lag of TPM is used and at the 1% level when 3-year or 4-year lag of TPM is used.

In addition, the results also provide strong evidence supporting the two hypotheses on local protectionism. The coefficients on the lagged TPM (tax plus profit margin) are negative and significant at the 1% level in all of the four estimations. The coefficients on

[9] In their study on the geographic concentration in the U.S. industries, Dumais, Ellison and Glaeser (2001) use more recent and disaggregated data than Kim (1995) and find significant shifts in industrial activity across regions, which suggests increasing irrelevance of the resource-endowment theory.

$SSOE$ (share of SOE employment) are negative when 1-year or 2-year lag of TPM is used, and are significant at the 10% level and the 5% level, respectively. The coefficient on $SSOE$ is positive when 4-year lag of TPM is used, but with very little significance.

To conclude, random-effect estimation results provide relatively strong support for the scale-economy theory but weak support for the external-economy theory. There is little evidence supporting the resource-based theory of industrial localization. In addition, the results indicate strongly that low levels of industrial localization are linked to high past $TPMs$ (tax plus profit margins) and high $SSOEs$ (percentages of SOE employment), the two hypotheses on local protectionism.

In addition to the random-effect estimation, we also use a two-step fixed-effect model to account for the cross-sectional effects of some factors. In the first step, the dependent variable is the same as in the random-effect regression. We exclude the two variables that have data only for one year, RD and $ENERGY$. The estimated fixed effects for the industries are then saved and used as the dependent variable in the regression of step two. The two excluded variables in step one are the independent variables in step two. In summary, the estimation procedure is as follows.

Step one:

$$HOOVER_{it} = \gamma_1 SCALE_{it} + \gamma_2 TPM_{i\,t-l} + \gamma_3 SSOE_{it} + \mu_i + \varepsilon_{it}$$

Step two:

$$\widehat{\mu}_i = \theta_0 + \theta_1 RD_i + \theta_2 ENERGY_i + \varepsilon_i$$

Once again, two sets of estimation results are obtained. In the first set, the dependent variable is the Hoover coefficient of localization calculated using output data. In the second, the coefficient is obtained using employment data. Ordinary least-squares estimation techniques are employed. The estimation results summarized in Table 5 are consistent with those we obtained using the random-effect model.

We first examine the estimation results using Hoover coefficient of localization calculated with output data (Table 5a). Again, there is weak evidence supporting the external-economy hypothesis. The coefficients on RD are all positive, and significant at the 10% level when 1-year lag of TPM is used and at the 15% level when 2-year lag of TPM is used. There is little or no evidence supporting the resource-based theory of localization. The coefficients on $ENERGY$ are positive but insignificant. The results provide support for the scale-economy hypothesis. The coefficients on $SCALE$ are positive and significant at the 5% level in three of the four results. The results regarding local protectionism are also similar to those from the random-effect estimation. The coefficients on lagged TPM are negative and significant at the 1% level for all of the four specifications. The coefficients on $SSOE$ are negative, and significant at the 5% level when 4-year lag of TPM is used and at the 1% level in the other three specifications, consistent with our hypothesis that a high percentage of SOE employment calls for a local authority to erect barriers for interregional trade and and cause lower localization.

We now examine the estimation results using the Hoover coefficient of industrial localization calculated with employment data. The results are, for the most part, similar to those we just presented. There is slightly weaker support for the external-economy hypothesis and the hypothesis on the share of SOE employment, but much stronger support for the scale-economy hypothesis. In general, the random-effect and the two-step fixed-effect estimations produce very similar results.

Finally, we would like to compare the results from the two methods of estimation, random effect and fixed effect. In general, while estimating panel data, both the random-effect and the fixed-effect models have their advantages and drawbacks. Random-effect models impose structural restrictions on intercepts, but the estimation is more efficient. Fixed-effect models impose fewer restrictions, but may not be as efficient. To evaluate the benefit from efficiency against the loss of generality, we carry out a Hausman test as discussed in Greene (2000). Since we lack the panel data on two of the five independent variables, we include only the remaining three variables in our statistical tests. The results are summarized in Table 6.

We carry out the Hausman test for each combination of the Hoover coefficient and lag period of variable TPM. For all but one specifications, the p-value is less than 2%, indicating a preference of the fixed-effect model over the random-effect model under these specifications. When the employment based Hoover coefficient is used as the dependent variable and a 4-year lag is used for TPM, the p-value is 35.8%, giving no clear indication which of the two models is better.

5 Conclusion

This paper investigates the determinants of regional specialization in China's industries. The transition economy of China provides us with a unique opportunity to study the role of protectionism in regional specialization within a country. In addition, our study also yield new evidence about the resource-endowment theory (Ohlin, 1933), the increasing-returns-to-scale theory (Krugman, 1991), and the external-economies theory (Marshall, 1920) of regional specialization. Finally, our finding about the overall time trend of regional specialization shed useful light on the debate about the degree of local protectionism in China.

We use a panel data of 32 two-digit industries in 29 Chinese regions over a period of 13 years (1985–1997). We use both output and employment data to measure the size of firms and eventually calculate the Hoover coefficient of localization. Although the latter data are more commonly used in the literature, the former may be more useful in the context of China. One reason is that there is redundant labor in many firms and the degree of this problem varies across industries and regions. Thus, employment may not provide a consistent measure of the size of the firms. The other reason is that more data are available about output than about employment.

We estimate both the random-effect and the fixed-effect model. Because we only have cross-sectional instead of panel data for some independent variables, the fixed-effect model cannot be estimated in one step. We adopt a two-step approach. In the first step, we estimate the industry fixed effects, using only independent variables with panel data. In the second step, we regress the estimated values of the industry fixed effects on the remaining independent variables.

The results lend strong support to the role of protectionism. It is found that the degree of regional specialization is lower for industries with higher profit-plus-tax margins in the past and for industries with higher shares of employment in state-owned enterprises, reflecting stronger incentives for local governments to protect these industries. There are also evidence supporting the increasing-returns-to-scale theory and the external-economies theory of regional specialization. It is found that both the average size of firms and the R&D intensity in an industry have positive effects on the degree of geographic concentration in the industry. However, we find no support for the resource-endowment theory, despite the significant regional disparity in resource endowment in China. Presumably, as transportation costs decrease over time, the key condition for the resource-endowment theory — the immobility of resources — no longer holds. These results are in contrast to those in Kim (1995), where the external-economies theory is rejected but the increasing-returns-to-scale theory and the resource-endowment theory are supported.

Despite the strong evidence for the role of protectionism, the overall time trend of regional specialization of industrial production in China has reversed an early drop and registered a significant increase in the later years of the reform era. This finding is in contrast to that in Young (2000) but is consistent with that in Naughton (1999). It suggests that the conventional economic forces behind regional specialization have become dominant over the force of local protectionism.

References:

Bacchetta, P., Rose, A. K., and van Wincoop, E., "Intranational Economics and International Economics." Journal of International Economics, 55, 1:1, 2001.

Bai C. E., Li, D. D., Tao, Z. G., and Wang, Y. J., "A Multitask Theory of State Enterprise Reform." Journal of Comparative Economics, 28, 4:716–738, Dec. 2000.

Davis, D. R., Weinstein, D. E., Bradford, S. C., and Shimpo, K., "Using International and Japanese Regional Data to Determine When the Factor Abundance Theory of Trade Works." American Economic Review, 87, 3:421-446, 1997.

Dumais, G., Ellison, G., and Glaeser, E., "Geographic Concentration as a Dynamic Process." Review of Economics and Statistics, forthcoming.

Ellison, G., and Glaeser, E., "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." Journal of Political Economy, 105, 5:889–927, Oct. 1997.

Enright, M., Geographic Concentration and Industrial Organization. Ph.D. thesis, Harvard University, 1990.

Greene, W. H., *Econometric Analysis*, 4th edition. Prentice Hall, 2000.

Hanson, G. H., "Scale Economies and the Geographic Concentration of Industry." Journal of Economic Geography, 1, 255-276, 2001.

Hoover, E.M., "The Measurement of Industrial Localization." Review of Economics and Statistics, 18, 162–171, 1936.

Kim, S., "Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S. Regional Manufacturing Structure, 1860–1987." Quarterly Journal of Economics, 110, 881–908, 1995.

Krugman, P., and Obstfeld, M., *International Economics: Theory and Practice*, 5th edition. Reading, MA: Addison-Wesley, 2000.

Krugman, P., *Geography and Trade.* Cambridge, MA: MIT Press, 1991.

Krugman, P., "Increasing Returns and Economic Geography." Journal of Political Economy, 99, 483–499, 1991.

Marshall, A., *Principles of Economics.* New York: Macmillan, 1920.

Naughton, B., "How Much Can Regional Integration Do to Unify China's Market?" mimeo, 1999.

O'Connell, P. G. J., and Wei, S. J., "'The Bigger They Are, the Harder They Fall': Retail Price Differences across U.S. Cities." Journal of International Economics 56, 1:21–53, 2002.

Ohlin, B., *Interregional and International Trade.* Cambridge, MA: Harvard University Press, 1933.

Overman, H. G., Redding, S. J., and Venables, A. J., "The Economic Geography of Trade, Production, and Income: A Survey of Empirics." mimeo, 2001.

Qian, Y. Y., "The Process of China's Market Transition (1978–1998): The Evolutionary, Historical, and Comparative Perspectives." Journal of Institutional and Theoretical Economics, 156, 1:151–171, Mar. 2000.

Rauch, J. E., "productivity Gains from Geographic Concentration of Human Capital: Evidence from Cities." Journal of Urban Economics, 34, 3: 380-400, 1993.

Sah, R. K., and Stiglitz, J. E., "The Economics of Price Scissors." American Economic Review, 74, 1:125–138, 1984.

Young, A., "The Razor's Edge: Distortions and Incremental Reform in the People's Republic of China." Quarterly Journal of Economics, 115, 4:1091–1135, Nov. 2000.

Table 1: Raw Material Intensity for Chinese Industries in 1992

Industry Code	Industry	Raw Material Intensity
13	Food Processing & Production	5.3252
25	Petroleum Processing, Coking Products, & Gas Production & Supply	3.9547
17	Textile Industry	3.8828
19	Leather, Furs, Down & Related Products	3.5629
30	Plastic Products	3.3184
18	Garments & Other Fiber Products	3.2795
20	Timber Processing, Bamboo, Cane, Palm Fiber & Straw Products	3.2598
22	Papermaking & Paper Products	3.2004
41	Electronic & Telecommunications	3.1740
32	Smelting & Pressing of Metals	2.9839
34	Metal Products	2.9658
26	Raw Chemical Materials & Chemical Products	2.9256
37	Transportation Equipment Manufacturing	2.8912
21	Furniture Manufacturing	2.7851
40	Electric Equipment & Machinery	2.7743
24	Stationery, Educational & Sports Goods	2.6981
28	Chemical Fibers	2.6257
23	Printing & Record Pressing	2.5793
35	Machinery & Equipment Manufacturing	2.5335
29	Rubber Products	2.4215
27	Medical & Pharmaceutical Products	2.2479
44	Electric Power, Steam & Hot Water Production & Supply	2.0085
31	Nonmetal Mineral Products	1.9918
8	Metals Mining & Processing	1.9464
15	Beverage Production	1.8854
42	Instruments, Meters, Cultural & Official Machinery	1.8376
6	Coal Mining & Processing	1.7866
46	Tap Water Production & Supply	1.7219
10	Nonmetal Minerals Mining & Processing	1.3631
7	Petroleum & Natural Gas Extraction	1.2121
16	Tobacco Processing	0.7984
12	Logging & Transport of Timber & Bamboo	0.7494
All		2.6958

Note: Following Kim's measurement, the Raw Material Intensity is defined as the ratio of (Output-Value added) to Value added.

Table 2a: HOOVER_Output Coefficients for 32 Two-Digit Industries over the Period of 1985-1997

Industry	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1997
6 Coal Mining & Processing	0.6272	0.6364	0.5972	0.5898	0.5726	0.5959	0.6151	0.6289	0.6239	0.6236	0.6545	0.6404
7 Petroleum & Natural Gas Extraction	0.8154	0.8076	0.7981	0.8068	0.8017	0.8100	0.8126	0.8045	0.7824	0.7831	0.7683	0.7402
8 Metals Mining & Processing	0.6273	0.6340	0.6099	0.6697	.	0.6489	0.6260	0.6182	0.5891	0.5943	0.6153	0.6070
10 Nonmetal Minerals Mining & Processing	0.3550	0.3805	0.3532	0.3317	0.3653	0.3807	0.3789	0.3662	0.3515	0.3850	0.3817	0.4203
12 Logging & Transport of Timber & Bamboo	0.8519	0.8596	0.8311	0.8624	0.8806	0.8774	0.8779	0.8837	0.8913	0.8935	0.8942	0.8954
13 Food Processing & Production	0.2988	0.2711	0.2324	0.2203	0.2399	0.2345	0.2370	0.2400	0.2707	0.2663	0.2632	0.2615
15 Beverage Production	.	0.3186	0.2887	0.2935	0.2907	0.2717	0.2794	0.3068	0.3102	0.2903	0.3012	0.2767
16 Tobacco Processing	.	0.4917	0.5045	0.5165	0.5350	0.5457	0.5587	0.5789	0.5895	0.6469	0.6556	0.6365
17 Textile Industry	0.3123	0.3109	0.3172	0.3116	0.3140	0.3408	0.3447	0.3603	0.4004	0.3965	0.4036	0.3835
18 Garments & Other Fiber Products	0.2098	0.1957	0.2441	0.2581	0.2942	0.3163	0.3550	0.3837	0.4444	0.4409	0.4409	0.4506
19 Leather, Furs, Down & Related Products	0.1921	0.2042	0.2226	0.2481	0.2746	0.3242	0.3698	0.4220	0.4238	0.4275	0.4479	0.4552
20 Timber Processing, Bamboo, Cane, Palm Fiber & Straw Products	0.4043	0.4029	0.4193	0.4396	0.4566	0.4409	0.4172	0.4160	0.3720	0.3636	0.3354	0.3636
21 Furniture Manufacturing	0.2404	0.2391	0.2505	0.2567	0.2723	0.2916	0.2913	0.3076	0.3414	0.3277	0.2921	0.2536
22 Papermaking & Paper Products	0.2310	0.2294	0.2227	0.2333	0.2439	0.2405	0.2371	0.2533	0.2756	0.2831	0.2792	0.2721
23 Printing & Record Pressing	.	.	0.2216	0.2328	0.2427	0.2302	0.2344	0.2374	0.2377	0.2428	0.2597	0.2470
24 Stationery, Educational & Sports Goods	0.4220	.	0.5513	0.5628	0.5682	0.5675	0.5733	0.5598	0.5568	0.5420	0.5801	0.5537
25 Petroleum Processing, Coking Products, & Gas Production & Supply	0.5575	0.5418	0.4456	0.4458	0.4439	0.4344	0.4298	0.4375	0.4177	0.4241	0.4259	0.4120
26 Raw Chemical Materials & Chemical Products	0.2301	0.2308	0.2217	0.2095	0.2154	0.2109	0.2016	0.2071	0.1955	0.2037	0.1993	0.2170
27 Medical & Pharmaceutical Products	0.2162	0.2154	0.2133	0.2127	0.1911	0.2117	0.2101	0.2296	0.2180	0.2274	0.2172	0.2340
28 Chemical Fibers	0.5068	0.5001	0.4755	0.4610	0.4411	0.4401	0.4404	0.4377	0.4756	0.4551	0.4552	0.4389
29 Rubber Products	0.2257	0.2219	0.2113	0.1985	0.2167	0.2174	0.2119	0.2249	0.2717	0.2942	0.3190	0.3344
30 Plastic Products	0.3085	0.2956	0.3180	0.2971	0.3101	0.3201	0.3410	0.3338	0.3506	0.3648	0.3578	0.3516
31 Nonmetal Mineral Products	0.2055	0.2151	0.2000	0.2049	0.1916	0.2003	0.2034	0.2127	0.2290	0.2431	0.2449	0.2561
32 Smelting & Pressing of Metals	0.3716	0.3692	0.3498	0.3946	0.3905	0.3979	0.4227	0.4151	0.3698	0.3827	0.3791	0.3816
34 Metal Products	0.1936	0.1860	0.1865	0.1972	0.1977	0.1921	0.2023	0.2146	0.2167	0.2232	0.2478	0.2419
35 Machinery & Equipment Manufacturing	0.1614	0.1632	0.1791	0.1732	0.2731	0.1798	0.1843	0.1959	0.2284	0.2435	0.2600	0.2845
37 Transportation Equipment Manufacturing	0.3287	0.3385	0.3336	0.3511	0.3651	0.3549	0.3507	0.3667	0.3544	0.3679	0.3633	0.3885
40 Electric Equipment & Machinery	0.2939	0.2824	0.2724	0.2855	0.2764	0.2665	0.2786	0.2818	0.3057	0.3108	0.3261	0.3404
41 Electronic & Telecommunications	0.4322	0.4365	0.4316	0.4024	0.4106	0.4223	0.4841	0.4926	0.5144	0.5358	0.5537	0.5754
42 Instruments, Meters, Cultural & Official Machinery	.	.	0.3470	0.3331	0.3504	0.3313	0.3416	0.3701	0.3678	0.3827	0.3856	0.4664
44 Electric Power, Steam & Hot Water Production & Supply	0.2778	0.2847	0.2489	0.2577	0.2272	0.2433	0.2424	0.2476	0.2639	0.2787	0.2595	0.2583
46 Tap Water Production & Supply	.	.	0.2430	0.2631	0.2877	0.2765	0.3429	0.3219	0.3209	0.3410	0.3603	0.3379
Simple Average	0.3665	0.3665	0.3544	0.3600	0.3594	0.3692	0.3780	0.3861	0.3925	0.3996	0.4040	0.4055
Weighted Average	0.3117	0.31118	0.3085	0.3056	0.31774	0.3182	0.32714	0.33453	0.34867	0.35859	0.3615	0.36698

Table 2b: HOOVER_Employment Coefficients for 32 Two-Digit Industries over the Period of 1988-1997

Industry		1988	1989	1990	1991	1992	1993	1994	1995	1997
6	Coal Mining & Processing	0.4558	0.4518	0.4489	0.4551	0.4610	0.4775	0.5604		0.4761
7	Petroleum & Natural Gas Extraction	0.7062	0.7280	0.6897	0.7000	0.6744	0.7523	0.6676		0.6753
8	Metals Mining & Processing	0.5562	0.5525	0.5470	0.5430	0.5410	0.5274	0.5259		0.4971
10	Nonmetal Minerals Mining & Processing	0.2803	0.2757	0.2859	0.3233	0.2849	0.2868	0.2823		0.2853
12	Logging & Transport of Timber & Bamboo	0.8721	0.8739	0.8785	0.8840	0.8850	0.8866	0.8930		0.9159
13	Food Processing & Production	0.2461	0.2387	0.2374	0.2404	0.2356	0.2215	0.2216		0.2134
15	Beverage Production	0.2718	0.2554	0.2461	0.2512	0.2431	0.2668	0.2632		0.2623
16	Tobacco Processing	0.5011	0.5001	0.4959	0.5064	0.5091	0.4985	0.5171		0.4843
17	Textile Industry	0.2870	0.2838	0.2874	0.2888	0.2932	0.3056	0.3063		0.3131
18	Garments & Other Fiber Products	0.2916	0.3044	0.3288	0.3512	0.3676	0.4266	0.4275		0.4895
19	Leather, Furs, Down & Related Products	0.2364	0.2509	0.2930	0.3289	0.3724	0.3756	0.4100		0.4805
20	Timber Processing, Bamboo, Cane, Palm Fiber & Straw Products	0.5198	0.5388	0.5458	0.5368	0.5225	0.5114	0.4996		0.4537
21	Furniture Manufacturing	0.2339	0.2366	0.2435	0.2420	0.2547	0.2380	0.2592		0.2575
22	Papermaking & Paper Products	0.2012	0.1975	0.1914	0.2009	0.2078	0.2120	0.2330		0.2136
23	Printing & Record Pressing	0.1631	0.1633	0.1632	0.1627	0.1656	0.1558	0.1625		0.1707
24	Stationery, Educational & Sports Goods	0.5310	0.5321	0.5678	0.5749	0.5835	0.6180	0.6195		0.6645
25	Petroleum Processing, Coking Products, & Gas Production & Supply	0.3512	0.3687	0.3599	0.3775	0.3788	0.3834	0.3985		0.4052
26	Raw Chemical Materials & Chemical Products	0.1758	0.1757	0.1794	0.1821	0.1747	0.1847	0.2217		0.1783
27	Medical & Pharmaceutical Products	0.1905	0.1929	0.1922	0.1935	0.1890	0.1923	0.2538		0.1727
28	Chemical Fibers	0.4207	0.4239	0.4115	0.4122	0.4072	0.3762	0.3610		0.4083
29	Rubber Products	0.2037	0.2030	0.2041	0.2037	0.2119	0.2216	0.2358		0.2724
30	Plastic Products	0.2989	0.2911	0.2892	0.2873	0.2964	0.2970	0.2841		0.3038
31	Nonmetal Mineral Products	0.2113	0.2057	0.2023	0.2063	0.1999	0.1882	0.2098		0.2093
32	Smelting & Pressing of Metals	0.3725	0.3732	0.3710	0.3712	0.3891	0.3569	0.3565		0.3528
34	Metal Products	0.1771	0.1821	0.1780	0.1806	0.1811	0.1893	0.1888		0.2252
35	Machinery & Equipment Manufacturing	0.1698	0.1682	0.1692	0.1748	0.1750	0.1993	0.2276		0.2247
37	Transportation Equipment Manufacturing	0.2844	0.2827	0.2840	0.2850	0.2902	0.2982	0.3179		0.3234
40	Electric Equipment & Machinery	0.2602	0.2583	0.2579	0.2605	0.2583	0.2665	0.3152		0.3006
41	Electronic & Telecommunications	0.3979	0.3994	0.4128	0.4367	0.4313	0.4676	0.4528		0.4955
42	Instruments, Meters, Cultural & Official Machinery	0.3251	0.3295	0.3283	0.3334	0.3423	0.3342	0.3326		0.3532
44	Electric Power, Steam & Hot Water Production & Supply	0.2767	0.2734	0.2797	0.2673	0.2648	0.2663	0.2546		0.2450
46	Tap Water Production & Supply	0.2092	0.2016	0.2040	0.2036	0.2109	0.1979	0.2293		0.1875
	Simple Average	0.3337	0.3348	0.3367	0.3427	0.3438	0.3494	0.3590		0.3597
	Weighted Average	0.28315	0.28317	0.28454	0.2898	0.29071	0.30604	0.32504		0.32627

Table 3: Mean Value and Rank of All Variables

Industry Code	HOOVER_Output	Rank	HOOVER_Employment	Rank	SCALE_Output (100M Yuan per firm)	Rank	SCALE_Employment (10K people per firm)	Rank
6	0.6171	4	0.4733	7	0.0278	15	0.0606	6
7	0.7942	2	0.6992	2	6.3714	1	1.9314	1
8	0.6218	3	0.5363	4	0.0214	18	0.0247	12
10	0.3708	12	0.2881	18	0.0081	30	0.0110	24
12	0.8749	1	0.8861	1	0.0483	10	0.1110	2
13	0.2530	24	0.2318	23	0.0208	19	0.0087	29
15	0.2934	20	0.2575	21	0.0198	22	0.0104	25
16	0.5690	5	0.5016	6	0.8084	2	0.0891	3
17	0.3497	15	0.2956	16	0.0546	8	0.0371	7
18	0.3361	16	0.3734	11	0.0175	23	0.0137	20
19	0.3343	17	0.3435	13	0.0200	21	0.0137	21
20	0.4026	10	0.5161	5	0.0072	31	0.0080	31
21	0.2804	22	0.2457	22	0.0063	32	0.0063	32
22	0.2501	25	0.2072	25	0.0207	20	0.0148	19
23	0.2386	27	0.1634	32	0.0093	29	0.0090	28
24	0.5489	6	0.5864	3	0.0160	25	0.0132	23
25	0.4513	9	0.3779	10	0.1668	5	0.0281	9
26	0.2119	30	0.1840	31	0.0417	13	0.0208	16
27	0.2164	29	0.1971	28	0.0649	7	0.0270	11
28	0.4606	8	0.4026	9	0.2245	3	0.0648	4
29	0.2456	26	0.2196	24	0.0419	12	0.0234	13
30	0.3291	18	0.2935	17	0.0161	24	0.0097	27
31	0.2172	28	0.2041	27	0.0124	27	0.0135	22
32	0.3854	11	0.3679	12	0.1728	4	0.0640	5
34	0.2083	32	0.1878	30	0.0134	26	0.0099	26
35	0.2105	31	0.1886	29	0.0259	16	0.0215	14
37	0.3553	14	0.2957	15	0.0485	9	0.0270	10
40	0.2934	21	0.2722	19	0.0369	14	0.0186	17
41	0.4743	7	0.4368	8	0.0772	6	0.0322	8
42	0.3676	13	0.3348	14	0.0228	17	0.0212	15
44	0.2575	23	0.2660	20	0.0429	11	0.0158	18
46	0.3095	19	0.2055	26	0.0095	28	0.0083	30
Mean	0.3790		0.3450		0.2655		0.0865	
Std Dev.	0.1685		0.1660		1.1235		0.3376	

Table 3: Mean Value and Rank of All Variables (Continued)

Industry Code	RD (0.0001)*	Rank	ENERGY (ton per 10K Yuan)	Rank	TPM	Rank	SSOE	Rank
6	12.5701	16	13.0045	4	0.0115	32	0.8449	6
7	208.5804	1	6.2033	8	0.1543	11	0.9810	1
8	14.4716	13	6.1928	9	0.1443	12	0.7890	9
10	9.8120	21	8.6313	6	0.1577	8	0.4307	25
12	14.3231	14	3.8498	13	0.1544	10	0.9075	4
13	2.5922	27	2.8034	15	0.0606	31	0.7493	12
15	12.3129	17	2.7789	16	0.1889	5	0.7584	11
16	14.2183	15	0.3370	31	0.5588	1	0.8775	5
17	7.2907	23	1.2722	23	0.0775	29	0.6283	17
18	1.6366	29	0.4613	29	0.0776	28	0.1214	32
19	2.8340	26	1.3094	21	0.0612	30	0.2560	29
20	1.9500	28	3.5694	14	0.0785	27	0.3722	26
21	0.6060	32	1.3821	19	0.0818	26	0.1418	31
22	12.0869	20	5.5537	10	0.1126	20	0.5440	22
23	3.1519	25	0.5792	28	0.1410	13	0.5607	21
24	0.6816	31	0.3631	30	0.1250	18	0.2993	27
25	41.3155	4	18.3977	2	0.2067	3	0.8440	7
26	25.8125	9	8.5997	7	0.1363	15	0.7584	10
27	46.6132	3	2.0877	17	0.1401	14	0.7373	14
28	12.1735	18	3.9710	12	0.1601	7	0.7380	13
29	36.9443	5	1.3449	20	0.1545	9	0.5317	23
30	7.8999	22	0.8964	25	0.0911	25	0.2101	30
31	12.1464	19	10.9271	5	0.1290	16	0.5947	19
32	23.3214	10	13.7966	3	0.1632	6	0.8100	8
34	0.6883	30	1.2994	22	0.1040	23	0.2706	28
35	36.4425	6	1.5159	18	0.1112	21	0.6588	16
37	47.9226	2	1.1011	24	0.1105	22	0.6596	15
40	21.2388	11	0.6832	27	0.1189	19	0.5149	24
41	19.3870	12	0.3122	32	0.1032	24	0.5774	20
42	30.9408	7	0.7340	26	0.1289	17	0.6062	18
44	6.2866	24	21.1671	1	0.2332	2	0.9261	3
46	25.9228	8	4.7314	11	0.2051	4	0.9265	2
Mean	22.3180		4.6830		0.1401		0.6133	
Std. Dev.	36.7004		5.4487		0.0897		0.2441	

*: 10K Yuan of R&D expenditure per 100M Yuan of output

Table 4a: Random Effect GLS Regression Estimates for HOOVER_Output

Coeficient St. D	Regression 1		Regression 2		Regression 3		Regression 4	
SCALE_Output	0.0161	#	0.0221	**	0.0252	***	0.0256	***
	0.0100		0.0095		0.0098		0.0094	
TPM1	-0.1216	***						
	0.0449							
TPM2			-0.1811	***				
			0.0437					
TPM3					-0.1775	***		
					0.0478			
TPM4							-0.1714	***
							0.0492	
SSOE	-0.0664	***	-0.0627	***	-0.0575	**	-0.0521	**
	0.0236		0.0227		0.0249		0.0248	
RD	0.0013	*	0.0012	#	0.0011		0.0010	
	0.0008		0.0008		0.0008		0.0008	
ENERGY	0.0084		0.0072		0.0064		0.0055	
	0.0118		0.0118		0.0118		0.0118	
Constant	0.3858	***	0.3987	***	0.4013	***	0.4030	***
	0.0387		0.0386		0.0387		0.0386	
Number of Obs	340		313		282		251	
R^2 Within	0.1141		0.1639		0.1538		0.1599	
R^2 Between	0.1148		0.1115		0.1153		0.1176	
R^2 Overall	0.1170		0.1151		0.1184		0.1208	
Wald Chi2	42.81		57.45		48.19		44.52	
Prob>Chi2	0.0000		0.0000		0.0000		0.0000	

Note: ***, **, *, and # denote 1%, 5%, 10%, and 15% significant level, respectively.

Table 4b: Random Effect GLS Regression Estimates for HOOVER_Employment

Coeficient St. D	Regression 5		Regression 6		Regression 7		Regression 8	
SCALE_Employment	0.0460	**	0.0548	**	0.0683	***	0.1266	***
	0.0228		0.0239		0.0242		0.0311	
TPM1	-0.1334	***						
	0.0443							
TPM2			-0.1234	***				
			0.0461					
TPM3					-0.1624	***		
					0.0447			
TPM4							-0.2418	***
							0.0517	
SSOE	-0.0372	*	-0.0415	**	-0.0235		0.0029	
	0.0213		0.0212		0.0222		0.0233	
RD	0.0010		0.0010		0.0009		0.0004	
	0.0008		0.0008		0.0008		0.0007	
ENERGY	0.0012		0.0014		0.0005		-0.0012	
	0.0119		0.0119		0.0119		0.0119	
Constant	0.3542	***	0.3574	***	0.3580	***	0.3677	***
	0.0386		0.0387		0.0386		0.0388	
Number of Obs	251		251		251		220	
R^2 Within	0.1072		0.0986		0.1216		0.1627	
R^2 Between	0.0813		0.0862		0.0935		0.1164	
R^2 Overall	0.0814		0.0860		0.0935		0.1168	
Wald Chi2	28.24		26.23		32.68		39.47	
Prob>Chi2	0.0000		0.0001		0.0000		0.0000	

Note: ***, **, *, and # denote 1%, 5%, 10%, and 15% significant level, respectively.

Table 5a: Two-Step Fixed Effect Estimates for HOOVER_Output

First Step

Dependent Variable: Hoover_Output

	Regression 1		Regression 2		Regression 3		Regression 4	
SCALE_Output	0.0129		0.0194	**	0.0223	**	0.0230	**
	0.0101		0.0095		0.0098		0.0094	
TPM1	-0.1233	***						
	0.0448							
TPM2			-0.1809	***				
			0.0435					
TPM3					-0.1747	***		
					0.0476			
TPM4							-0.1666	***
							0.0489	
SSOE	-0.0719	***	-0.0687	***	-0.0650	***	-0.0598	**
	0.0236		0.0226		0.0248		0.0247	
R^2	0.9934		0.9945		0.9951		0.9960	

Second Step

Dependent Variable: Estimated Coefficients for Each Cross-Section Unit

RD	0.0014	*	0.0012	#	0.0012		0.0011	
	0.0084		0.0008		0.0008		0.0008	
ENERGY	0.0087		0.0075		0.0068		0.0059	
	0.0130		0.0130		0.0130		0.0129	
Constant	0.3874	***	0.4004	***	0.4032	***	0.4047	***
	0.0415		0.0415		0.0414		0.0412	
R^2	0.1301		0.1011		0.0888		0.0807	

Note: ***, **, *, and # denote 1%, 5%, 10%, and 15% significant level, respectively.

Table 5b: Two-Step Fixed Effect Estimates for HOOVER_Employment

First Step

Dependent Variable: Hoover_Employment

	Regression 1		Regression 2		Regression 3		Regression 4	
SCALE_Employment	0.0408	*	0.0495	**	0.0629	***	0.1198	***
	0.0226		0.0238		0.0241		0.0311	
TPM1	-0.1346	***						
	0.0441							
TPM2			-0.1229	***				
			0.0458					
TPM3					-0.1608	***		
					0.0445			
TPM4							-0.2382	***
							0.0517	
SSOE	-0.0405	*	-0.0452	**	-0.0274		-0.0016	
	0.0212		0.0211		0.0221		0.0232	
R^2	0.9960		0.9960		0.9961		0.9965	

Second Step

Dependent Variable: Estimated Coefficients for Each Cross-Section Unit

RD	0.0011		0.0010		0.0009		0.0004	
	0.0008		0.0008		0.0008		0.0008	
ENERGY	0.0014		0.0016		0.0007		-0.0010	
	0.0130		0.0130		0.0129		0.0128	
Constant	0.3552	***	0.3584	***	0.3589	***	0.3684	***
	0.0415		0.0414		0.0412		0.0408	
R^2	0.0626		0.0581		0.0456		0.0095	

Note: ***, **, *, and # denote 1%, 5%, 10%, and 15% significant level, respectively.

Table 6: Hausman Test

	Dependent Variable	Independent Variable 1	Independent Variable 2	Independent Variable 3	Chi2 Prob>Chi2
1	HOOVER_Output	SCALE_Output	TPM1	SSOE	14.63 0.0022
2	HOOVER_Output	SCALE_Output	TPM2	SSOE	13.20 0.0042
3	HOOVER_Output	SCALE_Output	TPM3	SSOE	27.70 0.0000
4	HOOVER_Output	SCALE_Output	TPM4	SSOE	20.11 0.0002
5	HOOVER_Employment	SCALE_Employment	TPM1	SSOE	275.91 0.0000
6	HOOVER_Employment	SCALE_Employment	TPM2	SSOE	16.17 0.0010
7	HOOVER_Employment	SCALE_Employment	TPM3	SSOE	10.08 0.0179
8	HOOVER_Employment	SCALE_Employment	TPM4	SSOE	3.22 0.3584

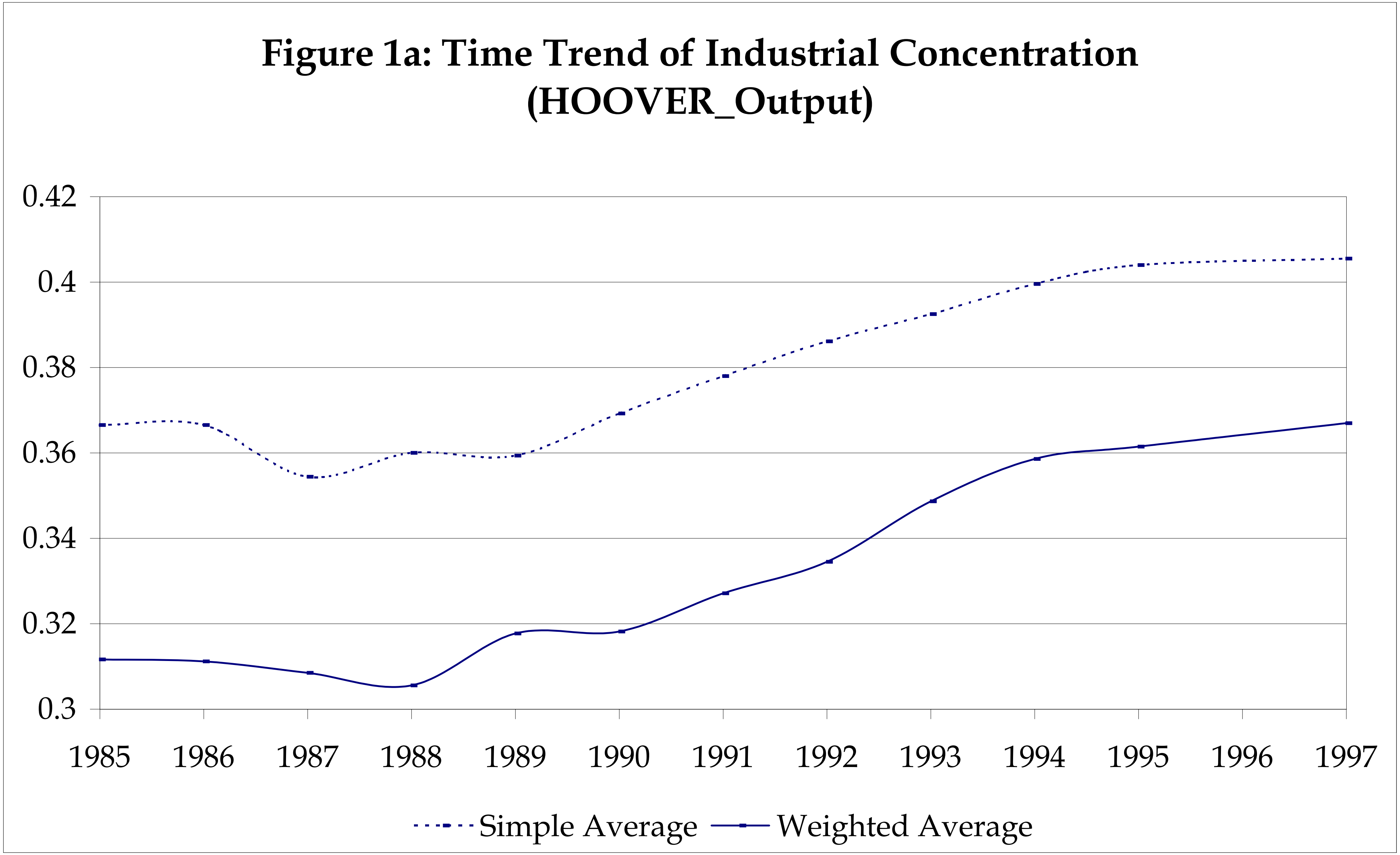


Figure 1a: Time Trend of Industrial Concentration
(HOOVER_Output)
0.42
0.4
0.38
0.36
0.34
0.32
0.3
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
Simple Average
Weighted Average

Figure 1b: Time Trend of Industrial Concentration (HOOVER_Employment)

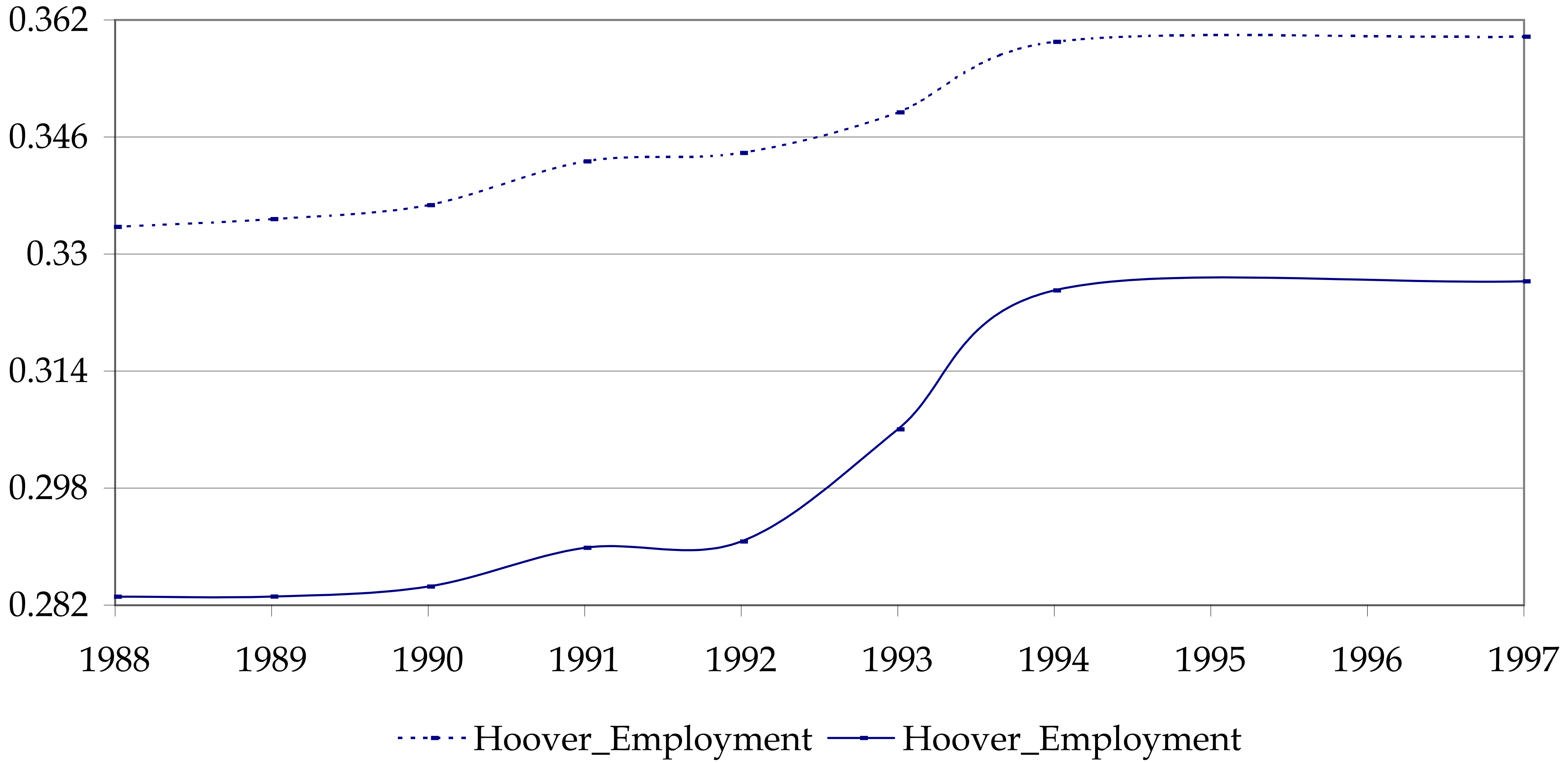


Figure 2: Time Trend of Tax Plus Profit Margin

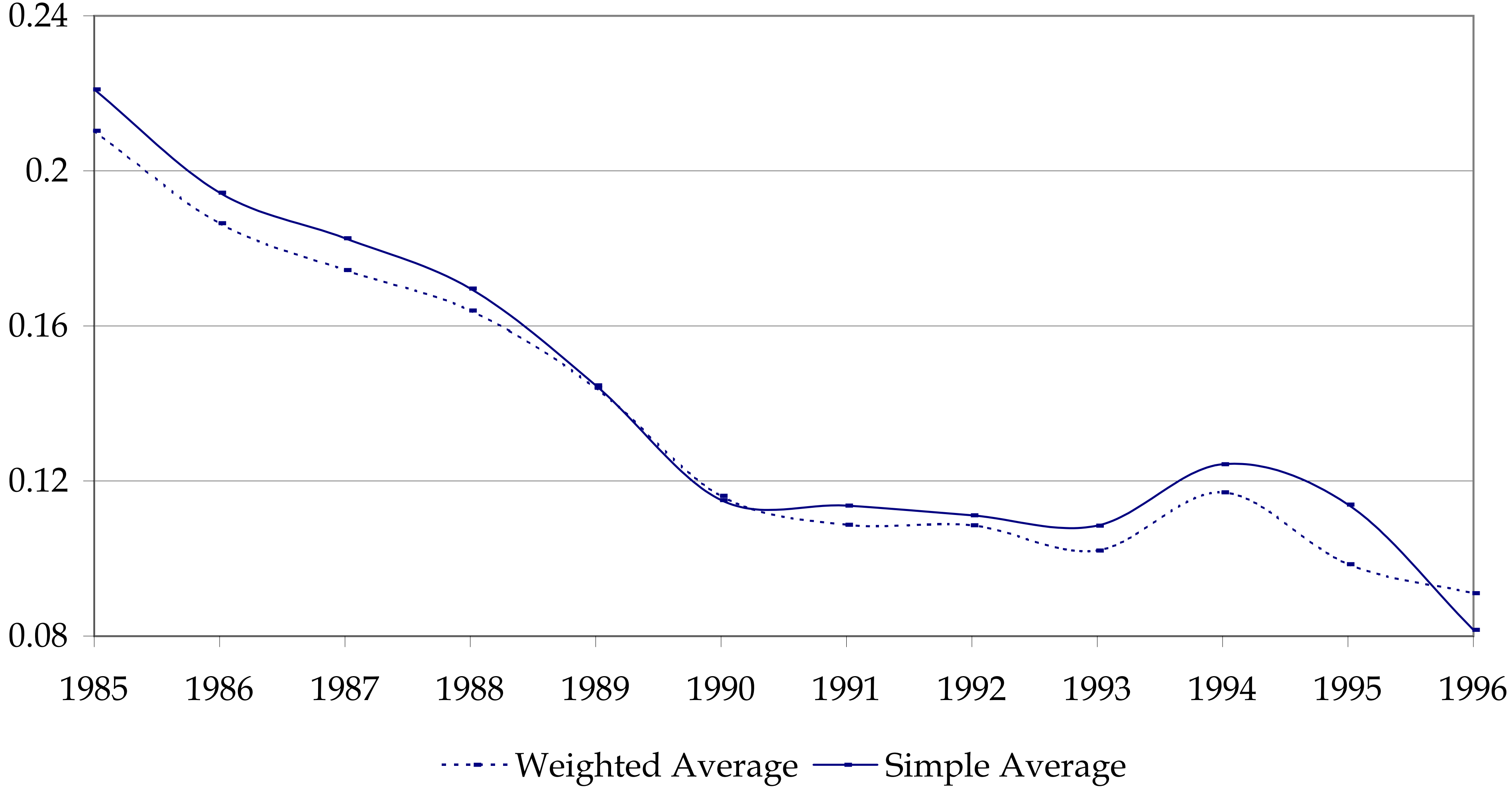
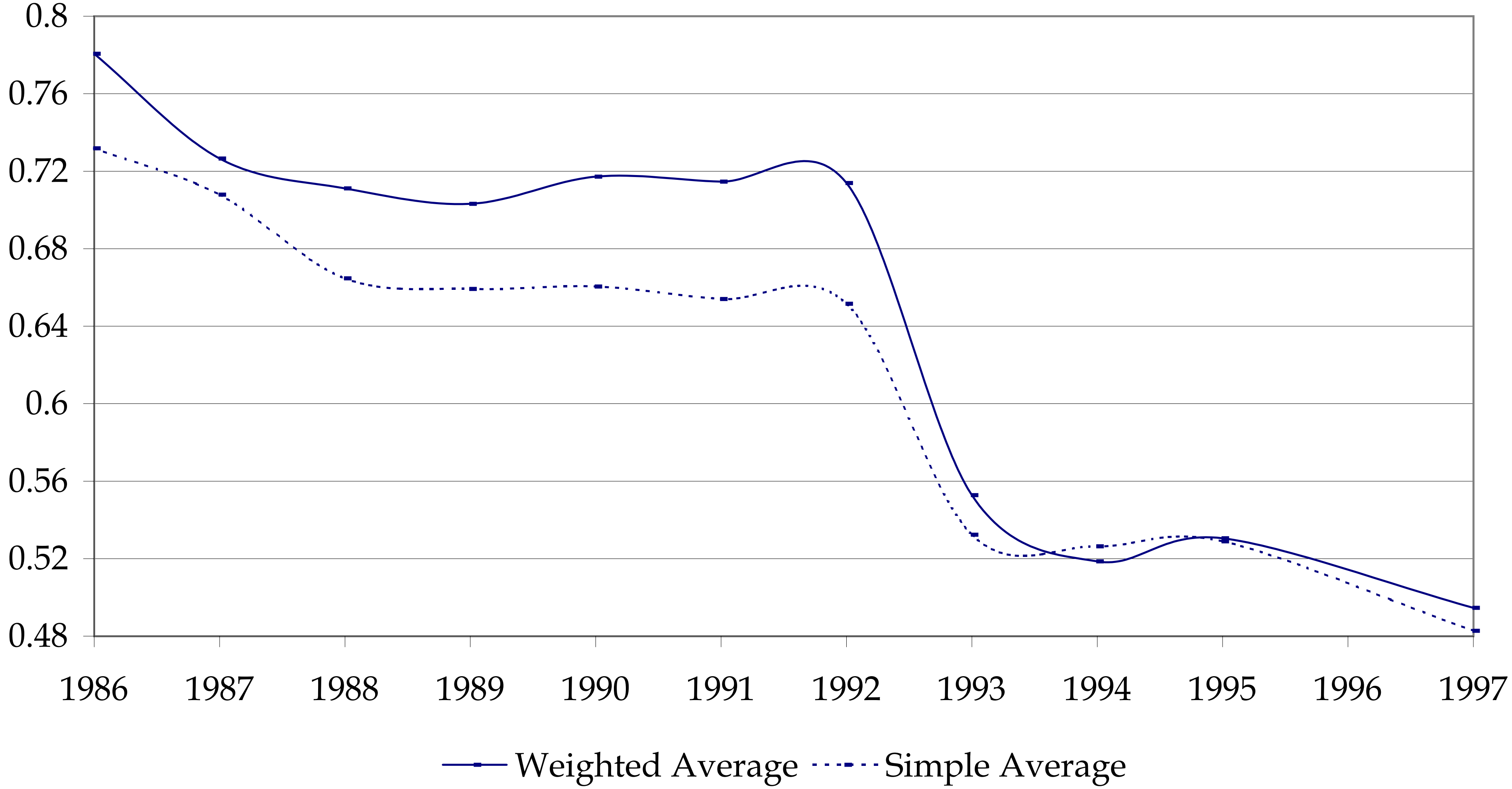


Figure 3: Time Trend of Share of SOE Employment



Data Appendix

	1985	1986	1987	1988	1989-1992	1993-1994	1995	1996	1997
HOOVER_Output									
Output for each industry and each region	2	2	2	1	1	1	3		1
HOOVER_Employment									
Employment for each industry and each region				1	1	1			1
SCALE_Output (= Output/Number of firms/Output deflator)									
Output for each industry		1	1	1	1	1	3		1
Number of firms in each industry		1	1	1	1	1	3		1
Price deflator		2	2	2	2	2	2		2
SCALE_Employment (=Employment/Number of firms)									
Employment for each industry				1	1	1	3		1
Number of firms in each industry				1	1	1	3		1
TPM(=Total Tax & Profit/Sales)									
Total tax & profit for each industry	1	1	1	1	1	1	3	1	
Total sales for each industry	1	1	1	1	1	1	3	1	
SSOE(=SOE Employment/Total Employment)									
Total Employment for each industry					2	2	3		1
The share of total SOE and COE (Collective-Owned Enterprises) Employment in Total Employment		2	2	2					
COE Employment for each industry		2	2	2					
SOE Employment for each industry		2	2	2	2		3		1
SOE Value-added for each industry						2			
SOE Productivity for each industry						2			
RD (= R&D Expenditure/Output)									
R&D expenditure for each industry							3		
ENERGY (=Energy consumption/Output)									
Energy consumption for each industry							2		

1. China Statistical Yearbook on Industrial Economy
2. China Statistical Yearbook
3. China Industrial Census