China Economic Review 30 (2014) 432–447

Contents lists available at ScienceDirect

China Economic Review

Measuring market concentration in China: the problem with using censored data and its rectification

Chong-En BAI [a], Jie MAO [b], Qiong ZHANG [c,*]

[a] Department of Economics, School of Economics and Management, Tsinghua University, Room 258 Weilun Building, Tsinghua University, Beijing, 100084, China
[b] Department of Public Finance and Taxation, School of International Trade and Economics, University of International Business and Economics, Box 106, Floor 11, Boxue Building, University of International Business and Economics, 10 Huixindong Street, Chaoyang District, Beijing 100029, China
[c] Department of Economics, School of Economics, Central University of Finance and Economics, Room 1309 Main Building, Central University of Finance and Economics, 39 South College Road, Haidian District, Beijing 100081, China

ARTICLE INFO

Article history:
Received 21 August 2013
Received in revised form 27 May 2014
Accepted 27 May 2014
Available online 5 June 2014

JEL classification:
D42
L11
C81

Keywords:
Zipf distribution
Statistical methods
Hirschman–Herfindahl index
Systematic bias

ABSTRACT

This paper utilizes the non-linear estimation method to simulate the Zipf distribution, and constructs an alternative measure of Hirschman–Herfindahl index (HHI), in order to reveal the real changes in monopoly of China's industrial markets. Based on the annual waves of the Chinese Industrial Enterprises Database between 1998 and 2009, it finds that: 1) systematic bias of deceptive declining concentration would be very easy to appear when directly using censored survey data with some invariant threshold; 2) with method in this article, an alternative measure of China's market concentration (namely, the estimated Zipfian parameter) can be produced to better depict monopoly trend, even though small firms are censored out in the market surveys and commonly used HHI cannot avoid such systematic bias; and 3) China actually experiences much less competition improvement or monopoly reduction in many industries during this period.

1. Introduction

The concentration of economic power is often simply referred to as a diversification in the relative size of the firms in an industry or in the aggregate economy (Adelman, 1951; Attaran & Saghafi, 1988; Bain, 1949; Lerner, 1934; Mankiw, 1985; Saving, 1970; Schwartzman, 1960). The concentration of economic power and its trend over time is among the most basic and important economic issues, particularly for empirical studies on market efficacy evaluation and the policy implications of Anti-Monopoly Laws (AMLs) (Ijiri & Simon, 1971; Nauenberg, Alkhamisi, & Andrijuk, 2004). However, many jurisdictions, even in developed countries with mature statistical systems, may be unable to collect detailed information for all firms on a regular basis, let alone on China. In fact, to our knowledge, the complete market share data in China, particularly for the industrial sector, are only available through three censuses, taken in 1995, 2004 and 2008. These censuses suffer from considerable between-wave intervals that limit their use in measuring concentration trends in China's rapidly growing economy. This paper measures China's recent concentration trend by using available annual market surveys. It is mainly motivated by two controversies. The first controversy is related to three scholars' arguments for more or less concentration during the 1970s and early 1980s in the United States. The second controversy refers to hot debates on profitability improvement and prosperity delusion in China's industrial sector since the late 1990s. The recent concentration trend in China would be a good way to rethink these disputes. What is interesting is that

* Corresponding author.
E-mail addresses: baichn@sem.tsinghua.edu.cn (C.-E. Bai), maojie@uibe.edu.cn (J. Mao), zhangqiong08@gmail.com (Q. Zhang).

http://dx.doi.org/10.1016/j.chieco.2014.05.013
1043-951X/

the data are all from sampling surveys, and a so-called systematic bias, which will be discussed later, may likely appear when directly using such censored data.

Comments, replies and further comments among three scholars, Mohsen Attaran, Massound Saghafi and Patric O'Neil, in *Applied Economics* between 1991 and 1993 remind scholars to use market surveys prudently to measure concentration trends, because reliable recall is often limited to large or listed firms for such types of surveys.[1] Thus, further work is required to follow their discussions as market surveys on a sampling portion of firms are commonly available in practice.

Hofman and Kuijs (2006), and the China Economic Quarterly report announced by the World Bank on May 31st in 2006 (World Bank Office, 2006) both found great profitability growth in China's industrial sector since 1999, contradicting the conventional wisdom of thin and falling profit margin due to capital inefficiency and enormous overcapacity. Shan (2006) critically says the World Bank is deluded due to data quality and rough analysis. The controversy then involves many other scholars, such as Hong Liang, Feng Lu, Steven Roach, Guoqing Song and Xiaonian Xu. Some scholars, such as Roach (2006), tend to support Shan's arguments, whereas other scholars, such as Liang (2006) and Lu, Song, Tang, Zhao, and Liu (2008), stand by the World Bank. Because answers to the questions "do China's firms make healthy profits or do Chinese still save and invest rationally" are so important, more efforts to reconcile these disputes are undoubtedly worthy making. Therefore, this article follows discussions on data quality and measures China's recent concentration trend, which should prompt scholars to consider why China tracks profits, changes in market structure, and productivity improvement.

China has conducted annual waves of the Chinese Industrial Enterprise Database (hereafter, CIED for short) since 1996. This database, established by the National Bureau of Statistics of China, includes more than 3 million firms from 40 industries in operation between 1998 and 2009, which are the best choice for evaluating the concentration of different industries or sectors over years.[2] For each year before 2007, all state-owned enterprises (SOEs) and non-SOEs with sales no less than 5 million RMB (*guimoyishang* in Chinese) non-SOEs with sales no less than 5 million RMB (*guimoyishang* in Chinese) and all state-owned enterprises (SOEs) are covered,[3] whereas only *guimoyishang* enterprises (for both SOEs and non-SOEs) are included after 2007. As censored with a certain time-invariant threshold, namely 5 million RMB, the data would spin misleading tales and policies. The reasons are as follows: China is a newly industrializing economy (NIE) and is currently experiencing rapid expansion in many industries, which is reflected by both more and larger firms. We would therefore naturally expect a continuously decreasing Herfindahl–Hirschman index (HHI) for these sectors.[4] The problem is that this decrease trend would very likely be magnified. Suppose all firms grow proportionally over years, which would indicate no real changes in concentration. The HHI figures from annual surveys will still show a downward concentration trend because more and more firms become larger than the time-invariant threshold and are thus added to the censored samples in later years.[5] Namely, the systematic bias of measuring concentration trends would appear when directly using censored data such as this.[6]

Such annual HHI figures could be misleading and potentially discard useful information regarding the entire market (Curry & George, 1983; Nauenberg, Basu, & Chand, 1997, Nauenberg et al., 2004). The surveyed firms are the larger ones and are market share leaders. Their structure will undoubtedly outline that of the whole market. Moreover, as suggested by many empirical studies, the sizes of firms in a certain industry follow a Zipf distribution.[7] Therefore, we can use corresponding simulated Zipf distribution from these censored data to "imitate" the entire market by "supplementing" smaller firms not surveyed. Therefore, we could possibly avoid the aforementioned systematic bias and reveal the "real" overtime concentration trend.

Thus, this paper adopts an alternative approach to confirm the systematic bias.[8] That is, it first estimates the shape parameter of Zipf distribution for each industry in each year and then compares these normalized estimators to the normalized HHI-censored figures—the ones directly computed from the censored data according to HHI's definition based on an algebraic proof of a one-to-one

[1] According to O'Neill (1991), although Attaran and Saghafi (1988) find a significant trend toward competition for the US manufacturing industry in the 1970s and the early 1980s, it would either over- or understate the true trend merely with data of the Fortune 500 companies. O'Neill (1993) further points out that his "corrected" indexes "represent a more accurate description of aggregate concentration" (footnote 7 on Page 1286) to further comment on the reply from Saghfi and Attaran (1991). O'Neil is justified in raising concern regarding the recalculation of market shares of the 500 largest firms, with denominators replaced by the sum of sales over all manufacturing firms. However, due to missing data, the summation of the residual smaller firms is neglected, therefore generating lower concentration levels than those reported by Attaran & Saghafi (A–S) in 1988 and 1991. The problem is that there is no way of knowing, a priori, to what extent the figures are underestimated. Even if O'Neill (1993) thinks that his "'corrected' values are much closer to the true index values" (footnote 7 on Page 1286). The true values could likely be located between his figures and A–S results, and the latter values could be even closer.

[2] CIED also includes data of firms surveyed in 1996 and 1997. These firms are not considered because those two waves are pre-researches for later large-scale surveys, and the sample sizes are much smaller.

[3] Note that CIED has an indicator named "prime operation revenue", which is different from and narrower than the sales indicator. According to the literature (Cai & Liu, 2009; Hall & Tideman, 1967; Ijiri & Simon, 1971), we use sales because it can more accurately reflect the size of a firm.

[4] This indicator is a widely used measure of market concentration (Hall & Tideman, 1967; Herfindahl, 1950; Hirschman, 1945).

[5] Many indicators are considered to measure firm size by scholars when studying economic concentration (Adelman, 1951; Hall & Tideman, 1967): sales, employment, net income, profit, and assets. We mainly consider here the sales one. We do robustness check later with the employment indicator, and find very similar results. The sales measure, seems to be better in our opinion, since it is more sensitive to market changes, as firms cannot easily and quickly adapt their quantities of employees to market changes, possibly due to employment contracts and high lay-off costs.

[6] Please refer to Table 1 for more details.

[7] A famous support for Zipf distribution, namely "The Zipf distribution is an unambiguous target that any empirically accurate theory of the firm must hit", is announced by Axtell (2001). Besides Axtell (2001), many other studies, such as Okuyama, Takayasu, and Takayasu (1999), Fujiwara, Di Guilmi, Aoyama, Gallegati, and Souma (2004), Luttmer (2007) and Gabaix and Landier (2008), all find firm size well follows Zipf distribution.

[8] Though not reported to save space, we also consider the approach to derive parallel data from the original CIED data, and then by each industry, compare the annual normalized HHI (denoted by HHI-parallel and HHI-censored) figures respectively estimated from these two data series. This method directly follows O'Neil and A–S's arguments and is very intuitive, while suffers from subjective and arbitrary choice on how to construct parallel data as lacking theoretical foundation. We find that this approach, similarly as done by Attaran and Saghafi (1988), is not a bad choice when only annual surveys are available; though still faces the systematic bias problem.