Economics Letters 131 (2015) 86–90

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

City scale and productivity in China

Chao Li, John Gibson *

Department of Economics, Private Bag 3105, University of Waikato, Hamilton 3240, New Zealand

HIGHLIGHTS

- Effects of city scale on output per worker in China are examined.
- Existing studies use incomplete measures of city scale and may give wrong results.
- Statistical yearbook data miss many private sector workers so Census data used here.
- Four-fifths of cities are close to estimated productivity-maximizing scale.
- Output losses from sub-optimal scale are typically below 10%.

ARTICLE INFO

Article history:
Received 12 December 2014
Received in revised form
23 March 2015
Accepted 3 April 2015
Available online 11 April 2015

JEL classification:
R12
O15

Keywords:
Agglomeration
Cities
Population
Productivity
Scale
China

ABSTRACT

We examine effects of city scale on output per worker in China. Existing studies use wrong measures of scale. Four-fifths of cities are close to the estimated productivity-maximizing scale. Output losses from sub-optimal scale are typically below 10%.

1. Introduction

A claimed legacy of China's command economy era, and the *hukou* restrictions on migration, is cities too small to fully exploit gains from agglomeration (Au and Henderson, 2006; Xu, 2009). But existing studies wrongly measure city scale.[1] For example, Xu (2009) uses 1990–1997 data that only counts people with local *hukou* registration for each city, ignoring more than 100 million city residents with *hukou* registration from elsewhere (Chan, 2012). Au and Henderson (2006) [henceforth, AH] measure city scale by 1997 employment but their data miss many private sector workers.[2] It is only since China's 2010 population census that reliable city-level estimates of the urban population and of total employment are available.

This paper uses these new, more reliable data, to estimate the relationship between output per worker and city scale. The flexible functional forms of AH are used, but with more controls for city industrial structure and more complete data on city scale (employment) that counts all private sector workers. Most

* Corresponding author. Tel.: +64 7 838 4289; fax: +64 7 838 4331.
E-mail addresses: cl95@waikato.ac.nz (C. Li), jkgibson@waikato.ac.nz (J. Gibson).

[1] Even simple descriptive claims are wrong. AH (2006, p. 557) say China had nine cities over 3 million and 125 of 1–3 million in 2000; a ratio of large to small cities (0.07) well below the global average of 0.27. But measured by residents in the 2000 census, China had 20 cities over 3 million and 89 cities of 1–3 million; a ratio of 0.23. The 2010 census has 38 cities with residents over 3 million versus 97 cities of 1–3 million, giving a large-to-small ratio of 0.39. Using the local *hukou* registered population rather than the resident population makes the ratio just 0.18.

[2] AH use *City Statistical Yearbook* (*CSY*) data, where 'private sector' employment was titled as self-employed (with a very low share); private sector employees are apparently excluded. Long-form census data on employment by sector show *CSY* substantially undercounts private sector employment.

http://dx.doi.org/10.1016/j.econlet.2015.04.001

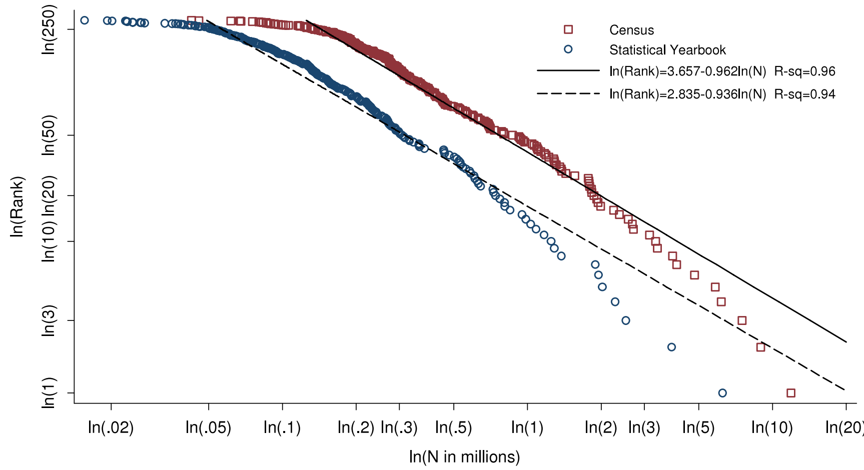

Fig. 1. Rank-size plots using census of population and statistical yearbook estimates of non-agricultural employment in 2010. *Notes*: *N* is non-agricultural employment (secondary and tertiary sectors). Number of observations is 286 cities.

cities are close to their productivity-maximizing scale, with losses from sub-optimal scale typically below 10%. Evidently, millions of non-*hukou* migrants are reshaping China's economic geography, forming an urban system with smaller productivity losses than previously thought. Other studies note relaxed *hukou* restrictions are changing China's city size distribution (Luckstead and Devadoss, 2014) but this is the first to challenge the often repeated claim that China foregoes agglomeration-based productivity gains from having too many small cities.

2. Data and model specification

The sample is 286 urban cores ('cities') in 2010. Prefectural cities (*diji shi*) have an urban core of districts (*shiqu*) equivalent to a city proper (Roberts et al., 2012) or to the urbanized portion of a Metropolitan Statistical Area (MSA) in the US (AH, 2006). Contiguous districts within the same prefectural city are merged so there is one core per prefecture. These data are at the same spatial level as AH (2006) and Xu (2009) but cover more of China; the prefectures these cores come from are the location of 97% of China's GDP. We use two main sources: the *City Statistical Yearbook* (*CSY*) (NBS, 2011), and the 2010 Population Census (NBS, 2012).[3] City-level GDP data are from *CSY* but not city scale estimates; *CSY* only counts local *hukou* holders so omits over 100 million urban residents with *hukou* registration from elsewhere, and also omits many private sector workers. Instead of *CSY*, more complete population and employment data from the 2010 Population Census long form are used to measure city scale.

While we mainly follow AH (2006) in estimating how big a city should be to maximize predicted output per worker, other benchmarks for an efficient city size distribution can be used, such as the fit to Zipf's Law (Gabaix, 1999). We therefore start with a log–log plot of city employment against city rank, which shows several features of the data (Fig. 1). First, the city size distribution is significantly ($p < 0.001$) closer to Zipf's Law if Census employment is used, with a coefficient of -0.962 compared with -0.936 for *CSY* employment (standard errors of 0.012 and 0.014), though both reject the coefficient of -1 needed for an exact fit to Zipf's Law. Second, the rank-size curve shifts left if using *CSY* employment, which averages just 43% of employment recorded for the same cities by the Census, due to the *CSY* data missing many private sector workers. Third, the tails of very small and very large cities are where there is most divergence from the linear fit of log rank on log size. To confirm that results are not due just to cities in the tails, results weighted by city employment are reported in addition to using equal weights for cities (so unequal weights for workers).

We use two flexible functional forms to relate output per worker, *GDP/N* to city scale, *N*, industrial structure, capital per worker, and other covariates, largely following AH (2006). First,

$$\ln\left(\frac{GDP}{N}\right) = \alpha_1 N^2 + \alpha_2 N + \alpha_3 N \times msgdp + \alpha_4 N \times gdp1 + \beta \ln\left(\frac{Capital}{N}\right) + \gamma X \quad (1)$$

allows the scale that maximizes output per worker of the *k*th city to be calculated as:

$$N_k^* = -\frac{\hat{\alpha_2} + \hat{\alpha_3} \times msgdp_k + \hat{\alpha_4} \times gdp1_k}{2 \times \hat{\alpha_1}}. \quad (2)$$

The industrial structure variables *msgdp* (ratio of secondary sector to tertiary sector GDP) and *gdp1* (share of primary sector GDP) are needed since productivity maximizing scale is higher in service-intensive cities. The vector of control variables, *X* includes population-weighted distance to all other cities (to proxy for domestic market potential), distance to the ten largest ports (to proxy for foreign market access), average years of schooling (to proxy for labor quality), and the output of foreign-affiliated firms relative to domestic firms (to proxy for technology).

With Eq. (1), estimated peak points for some cities are negative. This is less of a problem with a Generalized Leontief specification, which also was preferred by AH (2006):

$$\ln\left(\frac{GDP}{N}\right) = \alpha_1 N + \alpha_2 N^{0.5} + \alpha_3 N^{0.5} \times msgdp^{0.5} + \alpha_4 N^{0.5} \times gdp1^{0.5} + \beta \ln\left(\frac{Capital}{N}\right) + \gamma X \quad (3)$$

with the scale that maximizes output per worker of the *k*th city given by:

$$\sqrt{N_k^*} = -\frac{\hat{\alpha_2} + \hat{\alpha_3} \times msgdp_k^{0.5} + \hat{\alpha_4} \times gdp1_k^{0.5}}{2 \times \hat{\alpha_1}}. \quad (4)$$

When scale (total employment) for city k, N_k is less than N_k^* the city is deemed under-sized, otherwise it is considered over-sized. If N_k

[3] The 2001 Yearbook of the Ministry of Public Security (MPS, 2001) city counts of people with non-agricultural *hukou* for that city are used as an IV.