ANNALS OF ECONOMICS AND FINANCE **18-1**, 1–28 (2017)

A Structural Method to Estimate Firm-level Capacity Utilization and Application to Chinese Heavy Industries*

Shengyu Liu

Beijing Normal University, School of Business, Beijing, China
E-mail: liushengyu-2005@163.com

and

Heng Yin

Renmin University of China, National Academy of Development and Strategy, Beijing, China†
E-mail: yheng@ruc.edu.cn

We develop a structural model to estimate firm-specific capacity utilization (CU). Firm's productivity heterogeneity can be controlled in this framework. We demonstrate that the lowest point of short-run average total cost curve is the most suitable definition of potential output. The method is then tested by firm-level data including 15 Chinese heavy industries during 1998-2008. Intuitive findings provide strong evidence for the reliability and robustness of our estimation framework: CU keeps rising during the sample period, while slowing down after 2004; there are substantial differences in CU across industries, regions and ownerships; CU is strongly pro-cyclical; firm investment and government intervention deteriorates the problem of excess capacity.

Key Words: Capacity utilization; Firm heterogeneity; Structural estimation.
JEL Classification Numbers: D24, D61, E32, L11, O47.

1. INTRODUCTION

Capacity utilization (CU) is one of the most important indicators that have been widely used to measure macroeconomic performance. It's also a key variable in short-run macroeconomic analysis. For example, Green-

* We thank Yang Longjian, Liu Di, Li Shigang, and numerous seminar audiences for comments. This study is funded by the National Science Foundation of China (No.71373026, 71673305)

† Corresponding author.

1529-7373/2017

wood et al. (1988) introduce CU into their model, noticing that it is crucial to economic fluctuation. Basu and Kimball (1997) find that changes in CU can explain 40-60% of short-run economic fluctuation. Auernheimer and Trupkin (2014) develop a dynamic model to examine the role of CU in the propagation of business cycle.

Excess capacity has gradually become the top priority that Chinese economy needs to deal with. China has suffered from at least 5 rounds large-scale excess capacity in the past 20 years. The list of excess capacity industry keeps increasing in the government's energy conservation and emission reduction plan. The expansion of newly-built capacity exceeds the speed of cutting backward capacity, which makes the excess capacity problem even worsen. Excess capacity not only hurts the efficiency of resource allocation, but also hinders the upgrading of industrial structure. To locate the determinants of CU and implement appropriate policies, measurement of CU at firm level becomes an urgent issue.

The accuracy of CU estimation depends heavily on the quality of information at hand. It certainly helps that regularly large-scale and high-frequency firm-level surveys are performed specifically for measuring CU. For example, Quarterly Survey of Plant Capacity Utilization (QPC) are conducted and provide periodical statistics on CU for U.S. manufacturing. QPC collects detail information even on work patterns such as days in operation per week.[1] In contrast, due to the shortage of related firm-level survey, China's CU estimation mainly builds on macro and industry level data.[2] This may sometimes confound the picture. For example, OECD reports that China's average CU is 85.6% in 2011, while the IMF's statement being only 60%.[3]

Although CU estimation based on macro and industry data can provide some preliminary information, it smooths away lots of micro-level heterogeneity. After all, firm is the basic decision-making unit to balance the marginal revenue and cost of increasing capacity. Therefore, CU estimation should build on firm's cost minimization behaviour.

This paper aims at estimating heterogeneous firm-level CU in more reliable way. We develop a framework drawing on structural approach in the estimation of production functions and firm-specific productivity. Firm's productivity heterogeneity can be explicitly controlled in our CU estimation framework. In this paper, CU is defined as the ratio of firm's actual output to its potential output. We find that the suitable definition of potential output is the lowest point of short-run average total cost curve. We model capital stock as quasi-fixed input and derive theoretical expression of potential output according to firm's short-run cost minimization. Pa-

[1] http://www.census.gov/manufacturing/capacity/about_the_surveys/index.html.

[2] Cette et al. (2015) also note that France is short of firm-level CU information.

[3] See IMF Country Report No.12/195. OECD data is from http://stats.oecd.org/mei/.

rameters of production function and firm's heterogeneous productivity are estimated by proxy method developed firstly by Olley and Pakes (1996, henceforth OP). Firm-level CU is then calculated. We use firm-level data including 15 Chinese heavy industries during 1998-2008 to test the estimation procedure. Three robust and sensitive analyses are performed. The estimation results show median value of these manufacturing firms' CU is 90.36%. CU keeps rising during the overall sample period, while slowing down a little after 2004. There are substantial variations in CU both across industries, regions and ownerships. Five industries' CUs maintain at relatively lower level; firms located in China's Central and Western regions have lower CU than their counterpart in Eastern regions; and SOEs in general accumulate more capacity. CU is strongly pro-cyclical and there is an inverted U-shaped relationship between CU and firm age. Firm investment and government intervention have an adverse effect on CU. These intuitive findings provide strong evidence for the reliability and robustness of our estimation framework.

The rest of the paper proceeds as follows. Section 2 reviews CU estimation method. Section 3 introduces the structural estimation framework. Section 4 describes data we use in the test of our method. Section 5 presents estimation results and examines basic patterns of CU. Section 6 discusses the sensitivity of estimation results. Section 7 performs an empirical application on the determinants of CU. Section 8 provides concluding remarks.

2. LITERATURE REVIEW

It is a common consensus that CU is defined as the ratio of actual output to some measure of potential output.[4] In principle, if CU is less than 1, there exists somewhat excess capacity. Taking into account that it needs time to adjustment production capacity, firms usually maintain certain degree of surplus capacity to prepare for unexpected demand shock. Therefore, there is a thumb rule that the desirable CU lies in between 79%-82%.[5]

Disparities in CU estimation mainly stem from the definition of potential output, which can be divided into two groups: technology (engineering) approach and economic approach. Technology approach defines potential output as the maximum output that may be produced given a firm's all inputs (including equipment) are fully utilized under current technology. In U.S. the Wharton measure of potential output, for example, is based on "trend through peaks" and is supposed to capture the maximum attainable

[4]Lee (1995) defines CU as ratio of actual capital stock to its optimal capital stock. Following most literature, we select output definition. One consideration is that it's easier to measure. Output is also the most direct indicator of firm performance.

[5]See e.g. Corrado and Mattey (1997).

output.[6] Specifically, peaks in output are used as proxy of full utilization. Defining potential output as something under the minimum ratio of capital to output also belongs to this category. Nonparametric approaches like DEA (data envelopment analysis) or stochastic production frontier are often employed to estimate firm's potential output.

Technology approach defines potential output simply from the technical point of view, without considering the market demand under which firms operate. However, firm's real output is the rational choice that combining both market and cost information, such as input and output prices. Production technology is only one of the factors that affect firm's cost side. In other words, short-run fluctuation in CU is firm's reasonable response to output market shocks under constraints of input prices and production technology. Measuring CU without taking neither demand shocks nor input prices into account has almost nothing to do with the real purpose to discuss it, i.e. to deepen our understanding of firm behavior at micro level and overall economic fluctuation. Just in this sense Cassels (1937) point that "Potential output is conditioned in most cases by economic circumstances and must be interpreted as being the optimum output from the economic point of view."[7] Gajanan and Malhotra (2007) also argue that CU measures such as minimum capital-output ratio and peak-to-peak are not appropriate to analyse firm's supply decision-making.

Instead, economic approach takes into account all these factors and defines potential output as the optimal output under the current constraints of capital stock, production technology and market prices.[8] In general, economic approach sets the following analytical circumstance: capital can only be adjusted insufficiently in the short run; while labor and intermediate inputs (including materials, energy and power, etc.) can be fully adjusted. Therefore, expenditure related to capital is considered as fixed costs, while expenses on labor and intermediate inputs being variable costs. Then firm's capacity is determined by its capital stock. Thus, potential output can be defined as the level of output when firm's capital (capacity) has been fully made use of. The literature has discussed two definitions of "make fully use of capital". The first one, suggested by Cassels (1937), Hickman (1964) and Morrison (1985), corresponds to output that reaches to the minimum point of short-run average total cost ($SRATC$) curve, given current capital stock. Firms have no incentive to adjust it. The second definition of poten-

[6] Linear interpolation of Q/K ratios between peak years is employed to estimate potential output for between-peak years.

[7] See Hickman (1964) for more detailed discussion about the connection between economic definition and the engineering definition of capacity.

[8] For example, Baltagi et al. (1998) focus on the problem of excess capacity in US airline industry. They find that CU defined from the economic angle is more informative than engineering measure, such as seat occupancy.

tial output is the tangent of long-run average cost ($LRAC$) and $SRATC$ curves, advocated by Klein (1960), Berndt and Morrison (1981).

FIG. 1. Minimum short-run average total cost

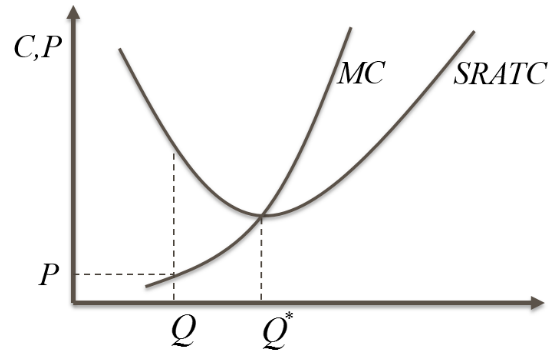


Figure 1 shows the first economic definition of potential output. Given current capital stock and input prices, $SRATC$ will be pinned down. Under current market demand (i.e. price is P), firm will select actual output Q to maximize profit (price equals marginal cost). In this case firms' revenue will be less than total costs, i.e. they lose money. Firms have incentive to adjust their capital stock at this point. Some firms will exit, others will reduce their capital stock. Industrial supply will move to the left and price recovers. This process stop until the MC curve and $SRATC$ curve get crossed (i.e. the lowest point on $SRATC$ curve). That is, only at the $SRATC$ curve's lowest point (output equals Q^*), current capital stock reaches equilibrium level, and firms have no incentive to adjust their capacity anymore. In this sense capital can be considered "fully utilized". This is just the logic that defines Q^* as potential output under current capital level.

Figure 2 illustrates the second economic definition of potential output suggested by the literature. According to envelope theorem, $LRAC$ is the envelope curve of $SRATC$ curve clusters. Since in the long-run firm can fully adjust its capital stock, capital has been adjusted to the optimal level at every point on $LRAC$. The tangent point (output equals $Q^{'}$) is not only on $SRATC$ curve, but also on $LRAC$ curve, which means output is optimal (capital being fully utilized). On any other point (such as output Q), however, the average cost using capital K is higher than the long-term average cost. In this sense the use of capital is inefficient.

In fact, the two economic definitions above are closely related together (this will be discussed in more detail in next section). As showing in Figure 3, if return to scale is constant, then $LRAC$ curve is horizontal, and the two economic definitions of potential output are equivalent. For increasing re-

FIG. 2. Tangent of long-run and short-run average total cost curves

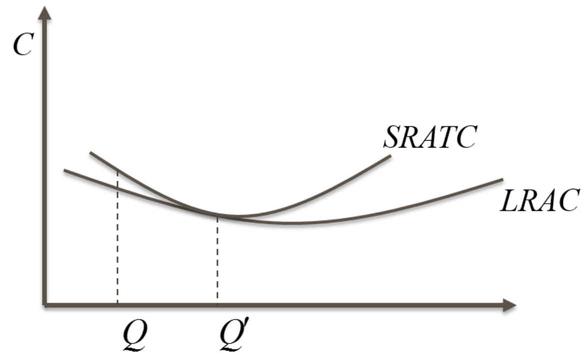
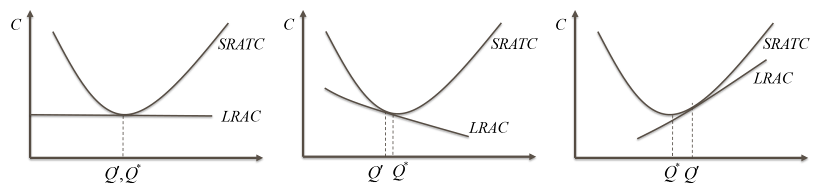


FIG. 3. Comparison between two economic definitions of potential output



Constant return to scale Increasing return to scale Decreasing return to scale

turn to scale, $LRAC$ curve slopes downward, and $Q^{'} < Q^{*}$; For decreasing return to scale, $LRAC$ curve slopes upward, then $Q^{'} > Q^{*}$.

However, we maintain that defining potential output as the lowest point of $SRATC$ curve (the first definition of economic approach) is better than the second (the tangent of $LRAC$ and $SRATC$ curve) for the CU estimation. The reasons are as follows. Firstly, CU is in nature a kind of short-run analysis. Capital can't be fully adjusted like other inputs such as labor and materials in the short run. Therefore firm's capacity in the short run is constrained by its capital stock. Firms adjust timely how much they produce according to the market demand. Thus CU fluctuates with demand shocks. The measurement and analysis of CU can help us understand firms' behavior like investment at micro level and economic fluctuation at macro level. This is just what the concept of CU aims to capture. As time horizon prolongs and capital can be fully adjusted, firm's capacity also changes with capital stock, then CU concept will be blurred. In this sense, the second definition based on capital adjustment is incompetent for CU measurement. Secondly, according to the first definition (Figure 1), firms will choose lower output under adverse demand shock, and it's CU is insufficient. To the opposite, if market demand is booming, firms will produce greater output than Q^*, which means over utilization. However, CU insufficient or excessive is not so direct and clear in the second definition of potential output. Thirdly, in industry with increasing return, firms have the incentive to build over capacity to threat potential entrants. This kind of strategical behavior make insufficient CU a natural tendency for this industry. However, as we have showed in Figure 3, CU will be over-estimated according to the second definition in return-increasing industry. Therefore, the second definition tends to make over capacity a seemingly less serious problem and even reverse the result. Of course, the extent to which the second definition bias the real picture is in nature an empirical problem. One of this paper's interesting findings is that the bias can be tremendous. We'll show this in section 6.1.

To our knowledge, until now literature on CU measurement doesn't satisfactorily deal with problem caused by firm productivity heterogeneity. Taking Nelson (1999)'s CU estimation of US privately owned electric utilities and Gajanan and Malhotra (2007)'s estimation on selected Indian manufacturing for example, they rely on time trend to control the overall trend in productivity fluctuations and neglect the effects of productivity heterogeneity on CU measurement. Thus, changes in productivity and changes in CU mix together which confounds CU estimation. Changes in CU stems from firms' active response to market demand, which is completely different from productivity changes. If firm productivity is heterogeneous, the gap between actual output and estimated potential output (based on economic or technology approach) is not solely caused by fluctuation in demand, it may also include difference between firm productivity. To take an example, assume two firms using the same amount of capital and producing the same quantity of output. Now consider one firm is more efficient (with higher productivity) than the other, then actually its potential output (CU) should be bigger (smaller). Therefore, if we ignore the difference in heterogeneous productivity, CU of firms with higher efficiency will be over-estimated. We will see this more clearly in section 3.1.

In the past 20 years, structural estimation of production function and productivity proposed by Olley and Pakes (1996) and Levinsohn and Petrin (2003, henceforth LP) has been widely used in empirical industrial organization and trade literature. It's still a fascinating field that attracts more and more researchers' attention. Aw, et al. (2011); De Loecker et al. (2012); Doraszelski and Jaumandreu (2013) and Ackerberg et al. (2015, henceforth ACF) are recent examples. The basic idea of OP and LP techniques is using observable variables as a proxy of unobserved productivity, thus solving the endogenous problem that long beset production function estimation. Drawing on these structural approach, this paper develops a structural framework to estimate CU at firm level, using the economic definition of potential output and trying to control firm-specific productivity.

3. STRUCTURAL MODEL AND ESTIMATION

3.1. Definition of potential output

Firm i at time t plans to produce output Q_{it} combining capital (K_{it}), labor (L_{it}) and intermediate inputs (M_{it}) with production function

$$Q_{it} = Q(K_{it}, L_{it}, M_{it}, \Omega_{it}), \tag{1}$$

Productivity (Ω_{it}) is assumed to be firm-specific and can only be observed by firm itself. In short-run firm's capital is assumed to be quasi-fixed factors, labor and intermediate inputs can be fully adjusted. Short-run variable cost is

$$VC_{it} = W_t L_{it} + P_t^M M_{it} \tag{2}$$

where W_t and P_t^M are wage and intermediate input price respectively. Specifically, considering Hicks-neutral Cobb-Douglas production function, firm's short-run cost minimization problem is

$$\begin{aligned} &\underset{L_{it},\ M_{it}}{Min}\ VC_{it} = W_t L_{it} + P_t^M M_{it}, \\ &s.t.\ Q_{it} = \Omega_{it} K_{it}^{\alpha_K} L_{it}^{\alpha_L} M_{it}^{\alpha_M}. \end{aligned} \tag{3}$$

From the first-order condition (FOC), we get

$$\begin{aligned} L_{it} &= \left(\frac{Q_{it}}{\Omega_{it} K_{it}^{\alpha_K}}\right)^{\frac{1}{\alpha_L+\alpha_M}} \left(\frac{W_t}{P_t^M}\frac{\alpha_M}{\alpha_L}\right)^{\frac{-\alpha_M}{\alpha_L+\alpha_M}}, \\ M_{it} &= \left(\frac{Q_{it}}{\Omega_{it} K_{it}^{\alpha_K}}\right)^{\frac{1}{\alpha_L+\alpha_M}} \left(\frac{W_t}{P_t^M}\frac{\alpha_M}{\alpha_L}\right)^{\frac{\alpha_L}{\alpha_L+\alpha_M}}. \end{aligned} \tag{4}$$

Then short-run average total cost ($SRATC$) is

$$SRATC_{it} \equiv \frac{VC_{it}}{Q_{it}} + \frac{P_t^K K_{it}}{Q_{it}}$$

$$= (\alpha_L + \alpha_M)\left(\frac{P_t^M}{\alpha_M}\right)^{\frac{\alpha_M}{\alpha_L+\alpha_M}} \left(\frac{W_t}{\alpha_L}\right)^{\frac{\alpha_L}{\alpha_L+\alpha_M}} \frac{(Q_{it})^{\frac{1}{\alpha_L+\alpha_M}-1}}{(\Omega_{it} K_{it}^{\alpha_K})^{\frac{1}{\alpha_L+\alpha_M}}} + \frac{P_t^K K_{it}}{Q_{it}}. \tag{5}$$

P_t^K is the user cost of capital. Defining firm's potential output (Q_{it}^*) as the lowest point of firm's $SRATC$ curve, i.e. $\partial SATC/\partial Q^* = 0$, which means that

$$Q_{it}^{*} = \Omega_{it} K_{it}^{\alpha_K+\alpha_L+\alpha_M} \left(\frac{\alpha_L}{W_t}\right)^{\alpha_L} \left(\frac{\alpha_M}{P_t^M}\right)^{\alpha_M} \left(\frac{P_t^K}{1-\alpha_L-\alpha_M}\right)^{\alpha_L+\alpha_M}. \quad (6)$$

Once firm's potential output is determined, CU can be measured as

$$CU_{it} \equiv \frac{Y_{it}}{Q_{it}^{*}} = \frac{Y_{it}}{\Omega_{it} K_{it}^{\alpha_K+\alpha_L+\alpha_M} \left(\frac{\alpha_L}{W_t}\right)^{\alpha_L} \left(\frac{\alpha_M}{P_t^M}\right)^{\alpha_M} \left(\frac{P_t^K}{1-\alpha_L-\alpha_M}\right)^{\alpha_L+\alpha_M}}. \quad (7)$$

where Y_{it} is the actual output. Note that if we define potential output ($Q_{it}^{'}$) as the output where the $SRATC$ curve is tangent to the $LATC$ curve, i.e. $Q_{it}^{'}$ is the output that satisfies $\partial SRATC/\partial K = 0$, then

$$Q_{it}^{'} = \Omega_{it} K_{it}^{\alpha_K+\alpha_L+\alpha_M} \left(\frac{\alpha_L}{W_t}\right)^{\alpha_L} \left(\frac{\alpha_M}{P_t^M}\right)^{\alpha_M} \left(\frac{P_t^K}{\alpha_K}\right)^{\alpha_L+\alpha_M}. \quad (8)$$

Compare equation (6) and (8), it is clear that under constant returns to scale ($\alpha_K + \alpha_L + \alpha_M = 1$) the two economic potential output defined above is equivalent. If returns to scale increase ($\alpha_K + \alpha_L + \alpha_M > 1$), then $Q^{'} < Q^{*}$. When returns to scale decrease ($\alpha_K + \alpha_L + \alpha_M < 1$), $Q^{'} > Q^{*}$.

(7) shows clearly that ignoring firm's productivity heterogeneity will bias CU measurement. Other things being equal, the higher firm's productivity is, the greater estimated potential output, and the lower CU estimation. Thus, ignoring productivity heterogeneity will overestimate high-productivity firms' CU.

(7) also shows that CU estimation depends on the actual output, capital, and three input prices, which can be observed directly. CU also relates to three parameters (i.e. α_K, α_L and α_M) and unobserved Ω_{it}, which we turn to the structural approach of production function and productivity estimation. Specifically, we use ACF method to estimate production function and OP method to deal with sample selection problem caused by exit.

3.2. Estimation

The firm's actual output is

$$Y_{it} = Q_{it} \exp\left(\mu_{it}\right). \quad (9)$$

where μ_{it} is shock to output. For Hicks-neutral Cobb-Douglas production function

$$y_{it} = \alpha_K k_{it} + \alpha_L l_{it} + \alpha_M m_{it} + \omega_{it} + \mu_{it}. \quad (10)$$

where k_{it}, l_{it}, m_{it} and ω_{it} are logs of corresponding variables. We assume ACF structure: capital k_{it} is determined one period ahead, current productivity ω_{it} realizes, and firm chooses labor after observing ω_{it}, then intermediate input is chosen. Therefore, labor is given by $l_{it} = l\left(k_{it}, \omega_{it}\right)$, and intermediate input demand is $m_{it} = m\left(k_{it}, l_{it}, \omega_{it}\right)$. Substitute them into the production function leads to

$$y_{it} = \varphi\left(k_{it}, l_{it}, m_{it}\right) + \mu_{it}. \tag{11}$$

We use two-stage estimation. The first stage just separates $\varphi\left(k_{it}, l_{it}, m_{it}\right)$ from output shock μ_{it}. By constraction, k_{it}, l_{it} and m_{it} are independent to μ_{it}. We use a complete polynomial of order three to approach the unknown function $\varphi\left(\cdot\right)$ and get the estimation $\widehat{\varphi}_{it}$ of $\varphi\left(k_{it}, l_{it}, m_{it}\right)$. In the second stage, the productivity is assumed to follow the first order Markov process, i.e. ω_{it} includes its conditional expectation at time $t-1$ and an innovation ξ_{it}. That is

$$\omega_{it} = E_t\left[\omega_{it} | \omega_{it-1}\right] + \xi_{it} = g\left(\omega_{it-1}\right) + \xi_{it}. \tag{12}$$

where $g\left(\omega_{it-1}\right)$ represents an unknown function of productivity expectation.

Since we can only observe samples of continuous operation, information omission of exitors may bias the estimation. In particular, the evolution of productivity (12) only represents continuous firm. That is

$$\omega_{it} = E_t\left[\omega_{it} | \omega_{it-1}, \theta_{it} = 1\right] + \xi_{it}. \tag{13}$$

θ_{it} is indicator function, which equals to 1 if firm remains active and 0 if it exits. Following OP,[9] we assume that firm' enter and exit decision depends on its current productivity level and a threshold $\underline{\omega}_{it}$. When $\omega_{it} > \underline{\omega}_{it}$, it continues to operate, otherwise it exits. Then

$$E_t\left[\omega_{it} | \omega_{it-1}, \theta_{it} = 1\right] = E_t\left[\omega_{it} | \omega_{it-1}, \omega_{it} > \underline{\omega}_{it}\right] = \phi\left(\underline{\omega}_{it}, \omega_{it-1}\right). \tag{14}$$

On the other hand, firm's survival probability can be expressed as

$$P_{it|t-1} = \Pr\left(\theta_{it} = 1 | \underline{\omega}_{it}, \omega_{it-1}\right) = \Pr\left(\omega_{it} > \underline{\omega}_{it} | \underline{\omega}_{it}, \omega_{it-1}\right) = \Psi_{t-1}\left(\underline{\omega}_{it}, \omega_{it-1}\right). \tag{15}$$

[9]As a recent example, Collard-Wexler and De Loecker (2015) also uses this method to deal with sample selection.

As OP notices, if the density of ω_{it} conditional on ω_{it-1} is positive in a region about $\underline{\omega}_{it}$ (for every ω_{it-1}), the selection equation (15) can be inverted to express $\underline{\omega}_{it}$ as a function of ω_{it-1} and $P_{it|t-1}$, i.e. $\Psi_{t-1}^{-1}\left(\omega_{it-1}, P_{it|t-1}\right)$. Then

$$E_t\left[\omega_{it}|\omega_{it-1}, \theta_{it}=1\right] = \phi\left(\omega_{it-1}, \Psi_{t-1}^{-1}\left(\omega_{it-1}, P_{it|t-1}\right)\right) = h(\omega_{it-1}, P_{it|t-1}). \tag{16}$$

Thus, to solve sample selection problem caused by firms' exit, we only need to add the predicted survival probability to Markov process (12), which becomes

$$\omega_{it} = h(\omega_{it-1}, P_{it|t-1}) + \xi_{it}. \tag{17}$$

Since capital is built one-period ahead, it doesn't relate to current innovation of productivity (ξ_{it}), which means

$$E_t\left[\xi_{it}\cdot\left[k_{it}\ l_{it-1}\ m_{it-1}\right]^{'}\right] = 0. \tag{18}$$

This is just our second stage moment condition to identify 3 production function parameters. We use GMM routine to estimate α_K, α_L and α_M. Then productivity can be recovered by

$$\omega_{it} = \widehat{\varphi}_{it} - \alpha_K k_{it} - \alpha_L l_{it} - \alpha_M m_{it} \tag{19}$$

Finally, bringing all these estimators into (8), we get an estimation of firm level CU.

4. DATA

The source of our data is the Annual Census of Industrial Production during period 1998–2008, a firm-level survey conducted by the National Bureau of Statistics (NBS). This annual census includes all industrial non-state firms with more than 5 million RMB (about $600,000) in annual sales plus all industrial state-owned firms (SOEs).[10] The source is the same used for example in Hsieh and Klenow (2009) and Song et al. (2011).

We have done an intensive work (in the style of Brandt et al., 2012) to link over time the data of the firms that presumably had the ID changed. This process has used extensively information such the firm's name, corporate representative, 6-digit district code, post code, address, telephone number, industry code, year of birth, and has been implemented in several steps:

[10] After 2006 SOEs with less than 5 million RMB are excluded from the survey.

first checking on neighbor years two by two, then longer panel sequences with the following/previous years. Description of the linking process can be found in Jaumandreu and Yin (2014).

TABLE 1.

Descriptive statistics*

Industry	Obs.	Output		Capital		Labor		Material	
		Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
20	36536	9.787	0.948	7.790	1.292	4.498	0.788	9.434	0.992
25	13109	10.115	1.514	8.885	1.896	4.591	1.267	9.716	1.548
26	88015	10.012	1.138	8.351	1.523	4.436	1.028	9.668	1.182
29	21034	10.008	1.060	8.290	1.410	4.827	1.013	9.457	1.092
30	49603	9.859	0.999	8.135	1.369	4.427	0.878	9.432	1.028
31	132342	9.899	1.041	8.560	1.366	4.760	0.881	9.398	1.100
32	36452	10.471	1.372	8.550	1.642	4.713	1.103	10.233	1.411
33	13335	10.146	1.225	8.170	1.635	4.488	1.054	9.926	1.269
34	83093	9.889	1.002	7.947	1.351	4.488	0.876	9.456	1.045
35	136479	9.851	1.009	8.038	1.357	4.541	0.903	9.443	1.045
36	69108	9.975	1.077	8.228	1.396	4.633	0.951	9.496	1.127
37	64669	10.200	1.286	8.442	1.615	4.797	1.032	9.715	1.336
39	71091	10.183	1.139	8.136	1.471	4.585	0.957	9.743	1.181
40	53337	10.809	1.378	8.707	1.753	5.121	1.134	10.216	1.449
41	15730	9.972	1.085	7.863	1.573	4.503	0.999	9.458	1.142
All	883933	10.035	1.145	8.265	1.486	4.628	0.976	9.598	1.189

*: Input and output variables are the logarithm forms.

To test our estimation method, we focus on all firms in 15 heavy industries defined by the National Bureau of Statistics.[11] We clean up the data according to the following four criteria. First, drop observations with missing value in key variables, such as revenue, labor, capital and intermediate materials. Second, delete observations with obvious reporting errors, such as negative value in key variables, or illogic values such as intermediate material being greater than revenue. Third, drop all firms with less than 8 employees. We also delete firms that variable cost are greater than revenue. Finally, delete the top and bottom 1 percent extreme values of the four main variables, i.e. revenue, labor, capital and intermediate materials. Table 1 provides the main variables' descriptive statistics.

5. ESTIMATION RESULTS

5.1. Production function parameters

[11]Appendix A lists the heavy industry classification, and detail sub-industries included.

TABLE 2.

Production function estimates[a]

Industry	Capital	Labor	Materials	Short RTS	Long RTS	Obs.[b]
20	0.035***	0.036***	0.939***	0.975	1.009	22345
25	0.054***	0.033***	0.907***	0.940	0.994	7804
26	0.044***	0.037***	0.929***	0.966	1.010	55976
29	0.043***	0.037***	0.939***	0.976	1.019	13772
30	0.038***	0.042***	0.931***	0.972	1.010	31923
31	0.040***	0.039***	0.925***	0.964	1.004	85338
32	0.029***	0.032***	0.941***	0.973	1.002	21920
33	0.032***	0.056***	0.932***	0.988	1.020	5837
34	0.038***	0.050***	0.929***	0.979	1.017	52263
35	0.047***	0.042***	0.932***	0.974	1.021	88155
36	0.057***	0.027***	0.915***	0.941	0.998	44424
37	0.046***	0.038***	0.921***	0.959	1.005	41628
39	0.031***	0.057***	0.929***	0.986	1.017	46618
40	0.036***	0.051***	0.911***	0.962	0.997	34242
41	0.056***	0.044***	0.892***	0.936	0.991	9995
All	0.042	0.041	0.926	0.968	1.009	-

[a]: All input coefficients of industries' production function are statistically significant estimated at 1%. t statistics are not reported limited to space. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.
[b]: Observations used in ACF estimation.

In Table 2 we report point estimates of production function parameters and returns to scale (RTS) for each of the 15 industries. The elasticities of the inputs are all precisely estimated at 1% level of statistical significance. In all 15 industries, the coefficients on capital and labor are between 0.027 - 0.057; the materials coefficient lay between 0.892-0.941. Short-run RTS ranges from 0.936 to 0.988; long-term RTS stays between 0.991-1.021. These results are consistent with the literature estimations.[12]

5.2. Distribution of CU

We now use our production function estimates to construct measures of firm level CU and describe its distribution and changes between 1998 and 2008. Figure 4 depicts firm's CU distribution density of all firms in the 15 heavy industries. Table 3 shows CU median value in more details.[13] CU median value of all these manufacturing firms is 90.36% and the ratio of firms with CU less than 80%[14] is 45.86%. According to 80% thresh-

[12]See Doraszelski and Jaumandreu (2013), for example.
[13]If we don't specially point out, all numbers are the median of CU in the related cell in the rest of this paper.
[14]We choose 80% as a thumb rule of the threshold of excess capacity.

TABLE 3.

CU estimation during 1998-2008[a]

Industry	1998	2000	2004	2008	98-03	04-08	Growth[b]	All year	Ratio[c]
20	0.599	0.699	0.933	1.169	0.821	1.151	40.20	1.016	41.96
25	0.915	1.236	1.915	3.197	1.476	2.604	76.38	1.968	19.96
26	0.527	0.682	1.111	1.399	0.809	1.323	63.54	1.062	40.25
29	0.300	0.397	0.709	0.983	0.489	0.880	79.87	0.673	56.02
30	0.412	0.534	0.880	1.193	0.639	1.086	69.85	0.868	46.89
31	0.354	0.395	0.643	1.013	0.490	0.884	80.40	0.677	55.61
32	0.611	0.693	1.376	1.742	0.921	1.523	65.37	1.217	34.69
33	0.307	0.324	0.771	0.661	0.495	0.645	30.36	0.528	63.08
34	0.372	0.451	0.849	1.017	0.593	0.926	56.20	0.773	51.22
35	0.351	0.453	0.895	1.110	0.613	1.003	63.56	0.828	48.72
36	0.777	0.932	1.777	2.213	1.230	2.013	63.67	1.646	26.21
37	0.492	0.569	1.196	1.506	0.788	1.369	73.58	1.089	39.53
39	0.199	0.292	0.604	0.781	0.377	0.691	83.49	0.527	63.75
40	0.539	0.678	1.092	1.465	0.801	1.303	62.64	1.069	40.92
41	0.913	1.198	2.276	3.149	1.533	2.736	78.50	2.165	20.38
All	0.424	0.527	0.948	1.247	0.670	1.130	68.71	0.904	45.95

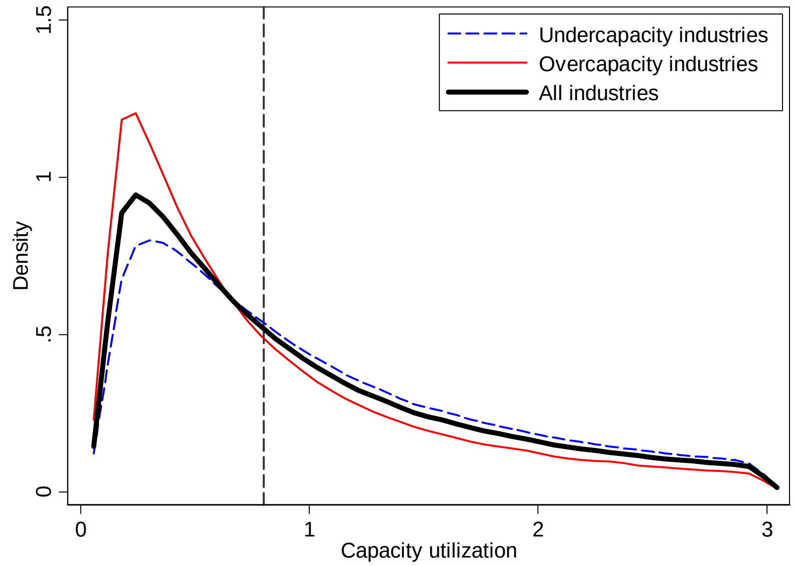
[a] Numbers are median of CU except specially mentioned.
[b] The % change of median between 1998–2003 and 2004–2008.
[c] The proportion (%) of firms with capacity utilization lower than 0.8.

old, we identify 5 excess capacity industries, i.e. Electronic Machinery and Equipment (39, 52.7%), Non-Ferrous Metal Rolling Processing Industry (33, 52.8%), Balata Product Industry (29, 67.3%), Nonmetallic Minerals Products (31, 67.7%) and Metal Product (34, 77.3%). Very interesting, these 5 industries are also widely considered excess capacity industries by policy makers and practitioners.[15] We further divide all firms into under-capacity group and over-capacity group, the latter including five most excess capacity industries mentioned above. Figure 4 shows the density distribution function for over-capacity industry (solid line) and under-capacity industry (dash line). The distribution for over-capacity firms is obviously to the left, the spike phenomenon being more significant.

Overall 46.56% firms' CU is greater than 1. Some industries' CU median even reach 1.5, especially in booming years (for example, 2004). Of course, from the view of technology approach (introduced in section 2), this is unacceptable. By definition, CU estimated by technology approach should always less than 1. In fact, potential output from economic approach is

[15]For example, all these industries are mentioned as serious excess by "the Sate Council's guidelines for solving serious excess capacity" (Chinese Sate Council, NO.2013-41, October 6th, 2013).

FIG. 4. Overall distribution of CU

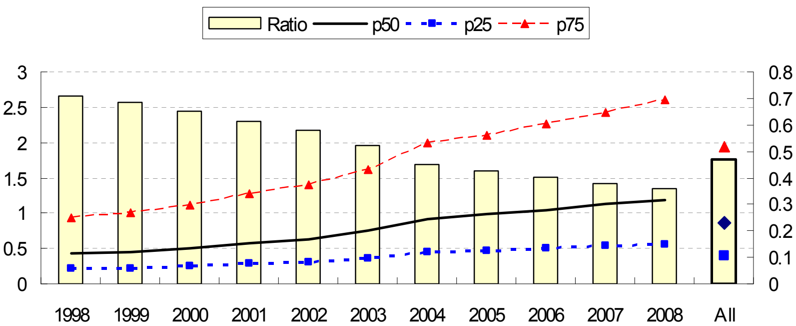


usually smaller than technology approach, which defines potential output the greatest capacity output or peak output. Accordingly, CU estimation of economic approach is systematically bigger, and exceeding 1 is rather common. We believe this is just what the advantage of economic approach lies in. As we have known, firm's capacity is limited by its fixed factors (capital), which can't be adjusted freely in the short run. If the market is booming, firms will produce more through balancing marginal revenue and marginal cost, producing even at rather high cost region. Apparently it's unreasonable to define this as firm's potential output. As the time horizon prolongs, firms will try to increase their capacity by investment and sooner or later their production will return to the position with the lowest cost.

5.3. CU trends and industrial differences

Figure 5 and Table 3 present the time trends of CU during 1998-2008. We also calculate the ratio of firms whose CU are lower than 80%. They both show that CU increases year by year, with obvious stage characteristics. First, CU increases from 42.4% in 1998 to 124.7% in 2008. Meanwhile, the ratio of CU lower than 0.8 decreases steadily from 70.15% in 1998 to 35.06% in 2008. Second, the whole interval can be further separated into two subperiods with salient difference. In the first subperiod (1998-2003), all industry's CU remains in comparatively low level (66.01%). Asian financial crisis occurs during this period. At the same time Chinese economy also suffers from serious deflation and structural adjustment. Chinese government keeps stimulating the economy using both expansionary fiscal and monetary policy, and gradually it works. At the end of this period the excess capacity has been alleviated obviously. In the second subperiod (2004-2008), CU maintains at high level and the growth trend of CU slows down. As the global financial crisis takes place in 2008, CU in some industries begins to decline. There appears obvious symptom that a new round of excess capacity is coming.

FIG. 5. Time variations of CU



According to Table 3, CU also shows obvious across-industry differences. First, as mentioned above, five industries are identified as excess capacity. Ratios of whose firms CU lower than 80% in these five industries are 63.75%, 63.08%, 56.02%, 55.61% and 51.22% respectively. Non-Ferrous Metal Rolling Processing Industry (33), for example, CU is only 32.4% before 2000, then begins to increase, falling again after 2004. Second, some industries' CU remains at relatively high level, even greater than 1.5 for some year. For Instrument, Meter, Stationery and Office Machine Manufacturing (41), Petroleum Processing and Coking Plant Industry (25) and Special Machinery Manufacturing (36), more than half of firms' CU is significantly higher than 90% in these three sectors even during the first subperiod. Table 2 also presents that capital coefficient on these three sectors are 0.056, 0.054 and 0.057, respectively, higher than the average level of the entire industry (0.042).[16] Third, some industries' CU hover around 0.8-1.2, with significant cyclical variation. Take Ferrous Metal Smelting and Rolling Processing (32) for example, CU in 1998-1999 is slightly less than 70%, then increases steadily to 137.6% in 2004 and slightly rises in the rest four years.

5.4. Regional and ownership differences

[16]This result is similar to Han at al. (2011). They argue that the main reason why these industries maintain a high CU level is that their capital-output ratio keeps in low level, while large number of orders make firms overload their equipments.

TABLE 4.

Pairwise comparison of CU on regional and ownership differences

	Panel A: Region					Panel B: Ownership			
	East	Middle	West	East-North		SOE	Collective	Private	HMT & Foreign
All[a]	0.961	0.863	0.675	0.765	All	0.366	0.730	1.045	0.725
East[b]		32.397	70.861	38.836	SOE		−92.657	−158.546	−91.676
Mid.[b]			37.866	14.151	Coll.			−85.680	−0.939
West[b]				−18.057	Pri.				86.624

[a] Numbers in this row are median of CU.
[b] Numbers in this row are z-statistics of pairwise median test.

TABLE 5.

Industry, region and ownership differences in CU[a]

Industry	East	Middle	West	East-North	SOE	Collective	Private	HMT & foreign
20	1.148	0.924	0.834	0.697	0.294	0.930	1.108	0.701
25	2.405	1.392	1.755	2.413	1.028	1.738	2.207	1.621
26	1.191	1.013	0.755	0.825	0.387	1.004	1.224	0.828
29	0.708	0.622	0.497	0.428	0.255	0.517	0.812	0.493
30	0.901	0.812	0.767	0.645	0.379	0.682	1.028	0.569
31	0.763	0.753	0.386	0.615	0.270	0.588	0.781	0.549
32	1.405	0.967	1.068	1.211	0.515	0.854	1.389	0.937
33	0.562	0.552	0.333	0.476	0.165	0.456	0.630	0.297
34	0.807	0.677	0.621	0.597	0.311	0.610	0.893	0.515
35	0.859	0.766	0.693	0.738	0.286	0.685	0.927	0.669
36	1.673	1.694	1.286	1.567	0.652	1.408	1.905	1.325
37	1.157	1.029	0.983	0.835	0.451	0.960	1.334	0.784
39	0.556	0.468	0.456	0.397	0.199	0.396	0.619	0.449
40	1.107	0.835	0.742	0.767	0.459	0.819	1.395	0.910
41	2.206	2.070	1.931	2.129	0.891	1.813	2.533	1.978

[a] Numbers are median of CU.

We also examine the cross-region variations and ownership differences of CU in Table 4 and Table 5. Generally China can be divided into four broad regions.[17] Results of cross-region variations are presented in the left panel of Table 4. Overall CU of the eastern is 96.1%, the central and the northeast being 86.3% and 76.5% respectively, and the western staying at the lowest level (67.5%). CU in the eastern region is 28 percentage points higher

[17] Eastern region includes Beijing, Tianjin, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, and Hainan; Middle region includes Shanxi, Anhui, Jiangxi, Henan, Hubei, Hunan; Western region includes Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang; Northeast region includes Liaoning, Heilongjiang, Jilin.

than the western region. Panel B of Table 4 shows ownership differences of CU.[18] Private firms' CU is significantly higher than the other three groups, with HMT & foreign and collective firms locating in the middle, while SOE being the lowest. Median CU of private and HMT & foreign firms is 1.86 and 0.98 times higher than SOE's.

Table 4 also shows median test based on industry and region dimensions. We present z statistics of pairwise comparison. For example, median test between eastern and central regions gives the z statistics of 32.40, a very significant difference. As this table displays, all differences at median level between subgroups of region and ownership are very significant except one case, i.e. collective and HMT & foreign firms. Table 5 further shows subgroups' CU of region and ownership at 2-digit industry level. Eastern firms' CU are among the top in 14 of these 15 industries, except Special Machinery Manufacturing (36), which being a little lower than Middle region. And in all the 15 industries, private sector's CU is saliently higher than the rest ownership groups, while SOE's CU always at the bottom.

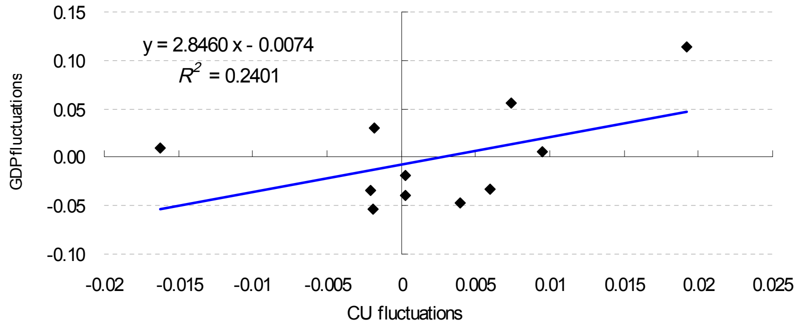
5.5. Business cycle and CU

In general, CU is expected to fluctuate with the rhythm of business cycle. Corrado and Mattey (1997) takes CU as the main indicator of economic fluctuations, pointing out that an 82% deviation in overall CU reveals highly volatility of the economy. Comin and Gertler (2006) finds closely positive correlation between U.S. CU and output volatility.[19] To test whether our estimates contain such correlation or not, Figure 6 depicts the scatter plot of CU fluctuations and real short-run GDP fluctuations, using the deviation from their time trend as measurement of fluctuations. The first impression it conveys is there indeed is positive relationship. The fitted R^2 of regression is 0.24, with the regression coefficient being 2.85. The correlation coefficient between two fluctuations is 0.49. All these exercises show that there indeed is strong positive correlation and CU is highly pro-cyclical during business cycle. We also conduct correlation analysis within each industry. Correlation coefficients between CU volatilities and real GDP fluctuations for the five over-capacity industries mentioned above are among the highest, indicating that these industries are more intensely affected by economic fluctuations. For example, correlation coefficients in Balata Product Industry (29), Nonmetallic Minerals Products (31) and Metal Product (34) are 0.878, 0.733 and 0.522, respectively.

[18] According to the biggest share in paid-in-capital, firms are classified into 4 groups: SOEs, collective, private, and HMT & foreign. HMT & foreign include those from Hong-Kong, Macau, and Taiwan (HMT) and those from foreign countries.

[19] The correlation coefficient between CU and annual output is 0.67 in the medium-term cycle. The number is 0.93 for high-frequency measured cycle. See their Table 4.

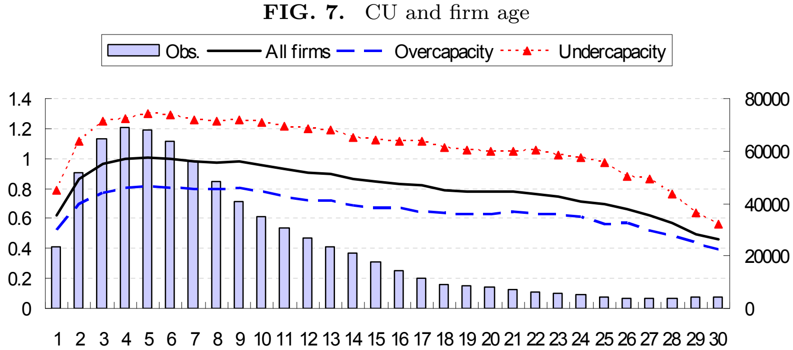
FIG. 6. CU and economic fluctuation



5.6. Firm age and CU

Figure 7 illustrates the relationship between CU and firms' age. In our sample, most firms are under 30 years old, so we take 30 as the horizon. CU presents an inverted U-shape relationship with firms' age. Within the first 5 years most firms' CU increase rapidly to the highest point and then decline gradually. A possible explanation for this pattern maybe as follows. Consumers' recognition rapidly increases as new firms enters and launches powerful promotion for their products. Firms' markets enlarge quickly and they accordingly make full use of their production capacity. As time passing by, their products tend to mature and most incumbents enter a stable period. Then their product advantages gradually go away and they switch into the last part of life cycle.

FIG. 7. CU and firm age



6. SENSITIVITY ANALYSIS

In this section, we do three exercises to test whether our CU estimations are robust to alternative specifications or not.

6.1. The definition of potential output

The first sensitivity analysis focuses on the definition of potential output. (8) provides the second method to estimate potential output and CU. Table 6 compares the results with the benchmark estimation.

TABLE 6.

Sensitivity analysis on definition of potential output

Industry	Median		Capital coef.		s.d.		CV
	$SRATC^a$	$LRTC^b$	Estimate	Quasi[c]	$SRATC^a$	$LRTC^b$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
20	1.016	1.383	0.035	0.025	5.032	6.846	1.000
25	1.968	1.790	0.054	0.060	16.724	15.210	0.669
26	1.062	1.347	0.044	0.034	6.582	8.343	0.932
29	0.673	1.184	0.043	0.024	4.131	7.269	1.294
30	0.869	1.179	0.038	0.028	4.178	5.671	0.998
31	0.677	0.745	0.040	0.036	4.546	4.999	0.809
32	1.217	1.316	0.029	0.027	10.197	11.025	0.795
33	0.528	1.371	0.032	0.012	5.269	13.684	1.910
34	0.773	1.382	0.038	0.021	4.556	8.152	1.316
35	0.828	1.465	0.047	0.026	4.835	8.551	1.301
36	1.647	1.597	0.057	0.059	13.225	12.823	0.713
37	1.089	1.223	0.046	0.041	14.228	15.978	0.826
39	0.527	1.129	0.031	0.014	6.018	12.889	1.575
40	1.069	1.000	0.036	0.039	10.447	9.777	0.688
41	2.165	1.893	0.056	0.064	17.495	15.295	0.643
All	0.904	1.229	0.042	0.032	8.272	9.984	1.360

[a] the first economic definition of potential output, i.e. the minimum point of $SRATC$ curve.
[b] the second economic definition of potential output, i.e. the tangent of $LRAC$ and $SRATC$ curves
[c] 1- short-run returns to scale.

Overall, CUs estimated by the second definition of potential output are systematically higher than the benchmark results. Column (2) of Table 6 shows median level of all 15 heavy industries' CU reaches to 122.9%. Recall the comparison of two definitions in Figure 3. In fact, Table 2 shows on average returns to scale of all 15 industries is 1.009. There are 11 industries' returns to scale are greater than 1. Table 6 clearly shows that in all these 11 increasing-return industries CU estimations by the second definition are higher than benchmark estimations, while in the rest 4 somewhat decreasing-return industries CU estimations being lower. These results are completely consistent with our expectation in section 2.

As Table 6 shows, the sharpest contrasts appear in 5 industries: Balata Product Industry (29, 0.673 to 1.184); Non-Ferrous Metal Rolling Processing Industry (33, 0.528 to 1.371); Metal Product (34, 0.773 to 1.382); Electronic Machinery and Equipment (39, 0.527 to 1.129); General Machinery Manufacturing (35, 0.828 to 1.465). To uncover in depth the comparison, we first calculate each industry's ratio of two standard deviations of CU under $LRTC$ and $SRATC$ estimation, then divide this ratio by 1.36 (i.e. 9.984 under $LRTC$ divided by 8.272 under $SRATC$) to convert it into what we call coefficient of variation (CV). Column (7) of table 6 shows that above-mentioned 5 industries' coefficient are greater than 1. This further demonstrates the huge differences between two definitions. As discussed in Section 5.2, these 5 industries are also widely considered excess capacity industries.[20] However, estimations from the second definition show that the capacity have been sufficiently utilized in these 5 industries.

In fact, the extent of increasing-return that leads to over-estimation of the second method is rather small. According to Table 2 and Table 6, these 5 industries are indeed among the top increasing-return, but they only exceed the constant return by 0.019; 0.020; 0.017; 0.021 and 0.017 respectively. These results show that estimation of CU by the definition of tangent of $LRAC$ and $SRATC$ curve is very sensitive to changes in return of scale. Slightly deviating constant return of scale can make the estimation fluctuate greatly, even obtaining totally unreasonable results. Therefore, what we discussed in Section 2 about the second definition is relevant according to above empirical test.

6.2. Alternative estimation approach

As above section shows that CU estimation may be sensitive to the coefficients on inputs, and productivity estimation may also matter, we now try another way to do these exercise. Specifically, we use LP method to reestimate all 15 heavy industries. The results are shown in the left side of Table 7.[21] Overall, LP estimates tend to somewhat underestimate CU. CU median for all firms is 80.16%, lower than the benchmark by 10 percentage points. Industries estimations further show that two methods obtain basically consistent results and CU share similar trends. To name a few, median CU of three industries such as General Machinery Manufacturing (35) , Electronic Machinery and Equipment (39) and Nonmetallic Minerals Products (31) only differ in about 2 percentage points. Among the 5 over

[20] General Machinery Manufacturing is at the border of 80% according to the first definition. Only Nonmetallic Minerals Products isn't in the list. From Table 2, return of scale is only 1.004, and the difference between two estimations is rather small.

[21] To save space, we don't report production parameters.

TABLE 7.

Sensitivity analysis on estimation method and factor price effect[a]

Industry	LP		Factor price effect				
	CU	Diff[b]	Bench[c]	0.2[d]	0.5[d]	2[d]	5[d]
20	0.781	0.235	1.016	0.959	0.991	1.042	1.077
25	0.833	1.135	1.968	1.866	1.923	2.014	2.076
26	0.728	0.334	1.062	1.001	1.035	1.090	1.128
29	0.877	-0.204	0.673	0.634	0.656	0.690	0.713
30	0.781	0.087	0.868	0.812	0.844	0.894	0.929
31	0.653	0.024	0.677	0.636	0.659	0.695	0.721
32	0.855	0.362	1.217	1.157	1.191	1.244	1.280
33	0.722	-0.194	0.528	0.482	0.508	0.549	0.578
34	0.821	-0.048	0.773	0.712	0.746	0.800	0.838
35	0.821	0.007	0.828	0.774	0.805	0.852	0.886
36	1.319	0.327	1.646	1.577	1.616	1.678	1.719
37	0.923	0.166	1.089	1.024	1.060	1.118	1.158
39	0.508	0.019	0.527	0.481	0.507	0.548	0.578
40	0.908	0.161	1.069	0.985	1.032	1.107	1.159
41	1.652	0.513	2.165	2.016	2.100	2.233	2.326
All	0.802	0.102	0.904	0.846	0.878	0.930	0.966

[a] Numbers are median of CU except specially mentioned.
[b] Median difference of CU between LP and bench model.
[c] Estimetion results of bench model.
[d] Times of real wage.

capacity industries identified by benchmark estimation, 3 remain there: Nonmetallic Minerals Products (31); Non-Ferrous Metal Rolling Processing Industry (33); Electronic Machinery and Equipment (39). The other 2 near the 80% threshold: Metal Product (34, 82.1%) and Balata Product Industry (29, 87.7%).

6.3. Changes in input prices

(7) shows that input prices can also be important determinant of CU estimation. Considering large wage variations even within industries; we choose wage distribution as representative to test the effects of input price changes. Specifically, we take four wage distributions, i.e. 0.2 times, 0.5 times, 2 times and 5 times of the benchmark distribution. As the right hand of Table 7 presents, in all these cases, CU estimation's changes are almost negligible for all industries. For example, when wage distribution enlarges five times, the corresponding change in overall CU is only 6.2 percentage points. This shows that CU measurement is rather robust to factor price changes.

In short, above sensitivity analyses together indicate that our benchmark framework for the CU estimation is quite robust and reliable.

7. EMPIRICAL APPLICATION

It's widely believed that over investment and government intervention are two main thrusts behind over capacity in China. For example, Lin et al. (2010) put forward a concept "tide-swell phenomenon" in the firms' investments to explain the formation of excess capacity in developing countries like China, whose economy may suffer from investment expansion owing to firms' periodic co-action based on the similar expectation. Han et al. (2011) also point out that over investment is the direct reason for over capacity. Meanwhile, government intervention especially from local governments may also mislead firms' behavior. Most Chinese local governments rely on business projects to boosting GDP. They tend to provide firm low-cost resources, such as cheap land, subsidies, etc. to attract their investment. For example, Dong et al. (2015) argue that Chinese fiscal decentralization system gives local governments strong incentive to intervene in firm's investment, which causes excess capacity.

As another test for our CU estimation framework, we run simple panel data fixed effects regression to check above relationship. Firm's investment is measured by the ratio of investment to capital stock. We apply subsidy dummy to proxy government intervention. Other controlling variables are as follows. Variable cost ratio, defining as the ratio of wage bill and intermediate inputs to total costs. Firms are willing to spend more on labor and materials to increase production when market booming, so variable cost rate may be positively related to CU. Fixed cost rate, defining as the ratio of firms' management expenses to revenue. Higher fixed cost rate will reduce firms' profits and restrain their ability to expand. Firm size, proxied by firm's total assets, may reflect their management efficiency on resource allocation.[22] As Figure 7 shows, we also use age and age square to control its non-linear effect.

Table 8 presents the empirical results.[23] Column (1) and (2) separately regress investment and government intervention on CU, column (3) puts two key variables together. We also divide the sample into two sub-samples: over-capacity industries (5 industries identified above) and under-capacity industries (the rest 10 industries). Results are reported in column (4) and

[22] See Hsieh and Klenow (2014), and Bloom et al. (2013) for example.

[23] In order to deal with endogenous problems caused by missing unobservables, we apply fixed effect model. To save space, table 8 only reports fixed effects regression results. Pool OLS regression results also support our conclusion. In addition, when we use the two main variables' one lag value as dependent variables, regression results do not change.

TABLE 8.

Determinants of CU, firm investment and government intervention[a]

	(1)	(2)	(3)	(4)	(5)
Investment Rate	-0.577^{***}		-0.577^{***}	-0.413^{***}	-0.662^{***}
	(-116.692)		(-116.750)	(-68.932)	(-95.378)
Subsidy		-0.056^{***}	-0.047^{***}	-0.043^{***}	-0.049^{***}
		(-8.318)	(-6.583)	(-4.872)	(-4.894)
Variable Cost Ratio	14.898***	18.426***	14.901***	11.170***	17.825***
	(165.897)	(263.642)	(165.943)	(111.023)	(134.192)
Fixed Cost Rate	-2.935^{***}	-2.052^{***}	-2.925^{***}	-1.361^{***}	-3.541^{***}
	(-40.084)	(-32.998)	(-39.944)	(-15.272)	(-34.447)
Firm Size	-0.093^{***}	-0.184^{***}	-0.091^{***}	-0.105^{***}	-0.093^{***}
	(-16.400)	(-39.375)	(-15.985)	(-15.432)	(-11.453)
Age	-0.009	0.136^{***}	-0.007	-0.125^{**}	0.016
	(-0.198)	(8.238)	(-0.154)	(-2.294)	(0.253)
Age^2	-0.007	-0.009	-0.007	0.107^{***}	-0.048^{*}
	(-0.338)	(-0.951)	(-0.388)	(4.542)	(-1.797)
Constant	-11.787^{***}	-14.599^{***}	-11.808^{***}	-8.767^{***}	-14.251^{***}
	(-106.849)	(-169.733)	(-107.001)	(-69.362)	(-88.567)
R^2	0.198	0.176	0.198	0.220	0.203
Obs.	536415	815698	536415	196318	340097

[a] Table reports results of FE regressions where individual and year fixed effects are controlled. t statistics in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

(5). We can see the estimated coefficients on the two key variables are strongly significant at 1% in all the regressions. Their coefficients are all negative. Therefore, as we have expected, all regression results consistently show that the more investment and the more government intervention are, the lower CU. High investment rate is indeed related to lower capacity utilization. And government intervention tends to exacerbate over-capacity problem. The coefficients of other control variables are all in line with that the literature have expected.

In summary, findings on relationships between firm's investment and government intervention and CU are consistent with the literature and empirical observations, which providing further evidences for the robustness of our CU estimation framework.

8. CONCLUDING REMARKS

Due to take firm heterogeneity into account, CU estimation at firm level are more suitable to analyze determinants and derive related policy implications. This paper presents a structural framework to estimate firm-level

CU. The biggest advantage is the ability to control productivity heterogeneity. Using 15 Chinese heavy industries data during the period the 1998-2008, we obtain a lot of reasonable estimation results, which provide strong evidences for the reliability of our estimation framework.

CU is an important indicator of macroeconomic performance. Accurate and timely CU information can improve the quality of macroeconomic analysis and the relevance and effectiveness of macroeconomic policies. CU is also the cornerstone of industrial policy. To develop an appropriate industrial policy and implement it effectively at proper time, we certainly need to grasp comprehensive and accurate information on firm-level CU in this industry. We believe that with more convenient information gathering technology and government's increasing awareness of the importance of CU information, there will be more abundant firm-level annual database and more high-frequency survey data targeted for CU measurement and analysis. We hope that the CU estimation framework in this paper can add a reliable instrument to dig CU information from these massive data.

APPENDIX: A: HEAVY INDUSTRY CLASSIFICATION

Two-digit Industry Name	Two-digit Code	Sub-industry Code Included
Wood Processing, and Other Wood Products	20	201, 202; 2031, 2032
Petroleum Processing and Coking Plant Industry	25	
Chemical Materials & Products Manufacturing	26	261, 262, 263, 265, 266
Balata Product Industry	29	
Plastic Product Industry	30	301, 302, 303, 304 , 305, 307
Nonmetallic Minerals Products	31	311, 312, 313, 316, 319, 3141, 3142, 3143, 3144, 3147, 3148, 3149, 3151, 3152
Ferrous Metal Smelting and Rolling Processing	32	
Non-Ferrous Metal Rolling Processing Industry	33	
Metal Product	34	341, 343, 344, 345, 346, 349; 3421, 3422, 3423, 3429, 3471
General Machinery Manufacturing	35	
Special Machinery Manufacturing	36	361, 362, 363, 364, 365, 366, 367, 369
Transportation Equipment Manufacturing	37	371, 372, 375, 376, 379
Electronic Machinery and Equipment	39	391, 392, 393, 399
Electronic Communication Equipment and Computer Manufacturing	40	401, 402, 403, 404, 405, 406, 409
Instrument, Meter, Stationery and Office Machine Manufacturing	41	411, 419; 412, 4141

APPENDIX: B: VARIABLES

Age. Current year minus the year in which the firm was born.

Subsidy. State aid received by the firm as proportion of sales.

Revenue. Revenue after taxes, at current prices, as reported by the firm.

Price of output. Output price index of the 2-digit industry the firm belongs to, taken from China Statistical Yearbook.

Capital. Estimate of the real stock constructed as follows. Firms report the value of their capital stock at original purchase prices and their capital stock at original purchase prices less accumulated depreciation. From these nominal values we estimate a sequence of real investments and a real capital stock at the starting year. Capital is then constructed by applying the perpetual inventory method assuming a yearly depreciation of 9%.

For firms founded after 1997, it is straightforward to get the starting nominal capital stock and the sequence of nominal investments by difference between the gross capital book values of two years. For those founded before 1998, we apply a method similar to Brandt, Van Biesebroeck, and Zhang (2012). We first estimate a yearly nominal rate of investment in fixed assets at 2-digit industry level using 1998-2003 firms' data. We assume that the nominal gross capital observed for the firm comes from the growth at this rate of the capital with which the firm was born. We then estimate the capital stock at birth, deflate it, and compute the real stock in the first year of observation by applying the perpetual inventory method with the series of real investments implied by our calculation. The investment deflator before 2006 is taken from Brandt, Rawski and Sutton (2008). We have updated it using the Fixed Asset Investment price index from China Statistical Yearbook.

Cost of materials. Estimate of the intermediate consumption in production as follows. The survey definition of intermediate inputs includes direct materials, intermediate inputs used in production, intermediate input in management, intermediate input in business operation (sales cost) and financial expenses. As we want to use a measure of variable cost, the inclusion of general management expenses, sales cost and financial costs is problematic. We have alternatively started by the manufacturing costs, which include materials, labor cost and depreciation of capital during the process of production. From these manufacturing costs we have then deduced the imputed wage bill and imputed depreciation of capital. From 2004 to 2007, we can do this using the detailed information on the structure of intermediate inputs. For the rest of years we assume the same proportions.

Price of materials. Estimate of a price index for the intermediate consumption of the industry the firm belongs to as follows. As Brandt, Van Biesebroeck and Zhang (2012) we compute a weighted average of the output prices of the industries to which the concerned industry purchases its inputs. For the weights we use the Input-Output table corresponding to 2002, which includes 42 sectors. The 2-digit manufacture price indices are from China Statistical Yearbook. The prices of agriculture, construction, transportation, retail and wholesale and some service sectors are calculated by comparing GDP at current prices and constant prices, which are included in the Collection of Statistical Material from 1949 to 2009.

Wage bill. We add up as wage bill several components of the yearly employees compensation. These components are wages, unemployment insurance premium, pension and medical insurance premium, housing mutual fund and total welfare fees. It should be taken into account that firms only began to report retirement and health insurance in 2003, and housing benefits in 2004.

Employment. Total number of employees, which includes all the full-time production and nonproduction workers, as reported by the firm. It excludes part-time and casual workers.

Wage. Wage bill divided by employment.

Variable Cost. Sum of the cost of materials and wage bill.

REFERENCES

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer, 2015. Identification properties of recent production function estimators. *Econometrica* **83**, 2411-2451.

Auernheimer, Leonardo, and Danilo R. Trupkin, 2014. The role of inventories and capacity utilization as shock absorbers. *Review of Economic Dynamics* **17**, 70-85.

Aw, Bee Yan, Mark J. Roberts, and Daniel Yi Xu, 2011. R&D Investment, Exporting, and Productivity Dynamics. *American Economic Review* **101**, 1312-44.

Baltagi, Badi H., James M. Griffin, and Sharada R. Vadali, 1998. Excess capacity, A permanent characteristic of US airlines? *Journal of Applied Econometrics* **13**, 645-657.

Basu, Susanto, and Miles S. Kimball, 1997. Cyclical productivity with unobserved input variation. No. w5915. National Bureau of Economic Research.

Berndt, Ernst R., and Catherine J. Morrison, 1981. Capacity utilization measures, underlying economic theory and an alternative approach. *American Economic Review* **71**, 48-52.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, 2013. Does Management Matter? Evidence from India. *Quarterly Journal of Economics* **128**, 1-51.

Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang, 2012. Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing. *Journal of Development Economics* **97**, 339-351.

Cassels, John M., 1937. Excess capacity and monopolistic competition. *Quarterly Journal of Economics* **51**, 426-443.

Cette, Gilbert, Nicolas Dromel, Remy Lecat, and Anne-Charlotte Paret, 2015. Production factor returns, the role of factor utilization. *Review of Economics and Statistics* **97**, 134-143

Collard-Wexler, Allan, and Jan De Loecker, 2015. Reallocation and technology, evidence from the US steel industry. *American Economic Review* **105**, 131-171.

Comin, Diego, and Mark Gertler, 2006. Medium-Term Business Cycles. *American Economic Review* **96**, 523-551.

Corrado, Carol, and Joe Mattey, 1997. Capacity utilization. *Journal of Economic Perspectives* **11**, 151-167.

De Loecker, Jan, 2011. Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica* **79**, 1407-1451.

Dong, Minjie, Yongmei Liang, and Qizi Zhang, 2015. Industry capacity utilization of china: industry comparisons, regional gap and affecting factors. *Economic Research Journal* (Jingji Yanjiu) **1**, 84-98 (in Chinese).

Doraszelski, Ulrich, and Jordi Jaumandreu, 2013. R&D and productivity, estimating endogenous productivity. *Review of Economic Studies* **80**, 1338-1383.

Gajanan, Shailendra, and D. Malhotra, 2007. Measures of capacity utilization and its determinants, a study of Indian manufacturing. *Applied Economics* **39**, 765-776.

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W. Huffman, 1988. Investment, capacity utilization, and the real business cycle. *American Economic Review* **78**, 402-417.

Han, Guogao, Tiemei Gao, Liguo Wang, Yingfei Qi, and Xiaoshu Wang, 2011. Research on measurement, volatility and causes of excess production capacity of Chinese manufacturing industries. *Economic Research Journal* (Jingji Yanjiu) **12**, 18-30 (in Chinese).

Hickman, Bert G, 1964. On a new method of capacity estimation. *Journal of the American Statistical Association* **59**, 529-549.

Hsieh, C.T. and P. Klenow, 2009, Misallocation and manufacturing TFP in China and India, *Quarterly Journal of Economics* **124**, 1403-1448.

Hsieh, C.T. and P. Klenow, 2014. The life cycle of plants in india and mexico. *Quarterly Journal of Economics* **129**, 1035-1084.

Jaumandreu, Jordi, and Heng Yin, 2014. Cost and product advantages: A firm-level model for the Chinese exports and industry growth. Boston University mimeo. http://people.bu.edu/jordij/papers/jy_may2014.pdf.

Klein, Lawrence R, 1960. Some theoretical issues in the measurement of capacity. *Econometrica* **28**, 272-286.

Lee, Jong-Kun, 1995. Comparative performance of short-run capacity utilization measures. *Economics Letters* **48**, 293-300.

Levinsohn, James, and Amil Petrin, 2003. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies* 70, 317-341.

Lin, J. Yifu, Hemao Wu, and Yiqing Xing, 2011. “Wave phenmena” and formation of excess capacity. *Economic Research Journal* (Jingji Yanjiu) **10**, 4-19 (in Chinese).

Morrison, Catherine J, 1985. On the economic interpretation and measurement of optimal capacity utilization with anticipatory expectations. *Review of Economic Studies* **52**, 295-309.

Nelson, Randy A, 1989. On the measurement of capacity utilization. *Journal of Industrial Economics* **37**, 273-286.

Olley, G. Steven, and Ariel Pakes, 1996. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica* **64**, 1263-1297.

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti, 2011. Growing like China. *American Economic Review* **101**, 196-233.