Identifying Capital Misallocation*

Zheng (Michael) Song†       Guiying Laura Wu‡

This version: January 2015

Abstract

Resource misallocation lowers aggregate productive efficiency. The existing literature often infers the magnitude of misallocation from the dispersion of average revenue products. However, the methodology is subject to several identification issues. In particular, the estimator would be upward biased by the presence of unobserved heterogeneities in output and demand elasticities; adjustment costs; and measurement errors in the data. This paper develops a new method of identifying capital misallocation in environments where all the above factors can be present. Applying the method to firm-level datasets from China’s industrial survey and Compustat, we find that capital misallocation implies aggregate revenue losses of 20 percent for Chinese firms but virtually zero losses for large Compustat firms.

JEL Classification: E22, O16, O47

Keywords: capital misallocation, generalized ARP approach, identification, structural estimation, unobserved heterogeneities

---

*We would like to thank Steve Bond, Qu Feng, Chang-Tai Hsieh, Måns Söderbom, Kjetil Storesletten, Shang-Jin Wei, Daniel Xu, Xiaodong Zhu, Fabrizio Zilibotti, seminar and conference participants at many institutions for helpful comments. We also thank Chao Duan and Jie Luo for excellent research assistance.

†University of Chicago, Booth School of Business. Email: zheng.song@chicagobooth.edu.

‡Nanyang Technological University, Division of Economics. Email: guiying.wu@ntu.edu.sg.
1 Introduction

Resource allocative efficiency differs across countries. The differences have recently been found important in accounting for the large cross-country difference in aggregate productive efficiency. Hsieh and Klenow (2009) infer the magnitude of resource misallocation by matching the dispersions of average revenue products (henceforth referred to as the ARP approach).\(^1\) The validity of the inference hinges on two conditions: (1) average and marginal revenue products have the same dispersion; and (2) the dispersion of marginal revenue products, a mirror image of price heterogeneity, reflects the magnitude of misallocation. Both conditions are strict. Condition (1) applies only to environments with homogeneous output and demand elasticities. Condition (2) will not necessarily hold in a dynamic environment with frictions such as adjustment costs. When it takes to the data, the ARP approach needs another condition that measurement errors do not add to the dispersions. Violation of any of the conditions would lead to biased estimation.

This paper develops a new method of identifying capital misallocation in a more general environment, where none of the conditions has to hold. The new method has a distinctive feature by matching a set of first and second moments of both the revenue-capital ratio (i.e., the average revenue product of capital) and the profit-revenue ratio. The profit-revenue ratio, which has not yet been explored in the misallocation literature, plays an important role in identification. Specifically, we match the variance of the revenue-capital and profit-revenue ratios and the cross correlation between the two ratios. The three empirical moments allow us to back out the three parameters governing the magnitude of the misallocation and unobserved heterogeneities in output and demand elasticities. In addition, while the ARP approach uses cross-sectional data, the new method explores between-group variations in panel data, which can effectively mitigate the bias caused by capital adjustment costs and measurement errors. We refer to the new method as the generalized ARP approach.

For illustrative purposes, we first present a simple model, where closed-form solutions make the identification of the unobserved heterogeneities highly transparent. Simulations show that the bias of the ARP approach caused by heterogeneities in output and demand elasticities appears to be severe under reasonable parameterization. By contrast, the generalized ARP approach manages to eradicate most of the bias.

\(^1\)This is also called “the indirect approach” by Restuccia and Rogerson (2013). See their paper for a review of the literature that adopts the approach to assess misallocation.
We next extend the model by incorporating a rich structure of capital adjustment costs and transitory measurement errors. The extent to which the generalized ARP approach is biased in a panel depends on the magnitude of capital adjustment costs and measurement errors, which, in turn, needs to be estimated. To this end, we adopt the simulated method of moments to estimate structurally all the key parameters in the full-blown model. As an extension to the identification condition in the generalized ARP approach, we illustrate how the structural estimation can separately identify the much larger set of parameters.

Our main empirical exercise is to apply the generalized ARP approach and the structural estimation to a firm-level panel data from the industrial survey conducted by China’s National Bureau of Statistics. The generalized ARP approach finds that correcting capital misallocation would increase China’s manufacturing output by 20 percent. In contrast, the ARP approach implies a much larger efficiency gain of 35 percent.

To control for other factors that may potentially bias the generalized ARP approach, we back out the magnitude of capital misallocation for U.S. manufacturing firms in Compustat. Improving capital allocation efficiency to the level among the Compustat firms would increase China’s manufacturing output by 16 percent. We then estimate a subsample in Compustat consisting of large firms only as in Bloom (2009). It has been well documented by the literature, for example, Fazzari et al. (1988), that large Compustat firms are less likely to be financially constrained. Capital misallocation would, thus, be less severe among the large firms. Interestingly, the generalized ARP approach finds much weaker evidence for capital misallocation in the subsample. The heterogeneities in output and demand elasticities can essentially account for all the dispersion of the revenue-capital ratio among large Compustat firms, even if the magnitude of the dispersion is similar across the three samples.

We also apply the structural estimation to the three samples. The structural estimation finds capital misallocation to be statistically significant and quantitatively similar to the magnitude backed out by the generalized ARP approach throughout the three samples. Moreover, the misallocation has no significance in the sample with large Compustat firms. In other words, the generalized ARP approach provides a first-order approximation to the structural estimation. We regard it as an important finding since the generalized ARP approach, which preserves some tractability from the ARP approach, is much easier to implement than the structural estimation.

A few extensions are conducted based on the generalized ARP approach. We provide a rough estimate of labor misallocation without resorting to a full specification on labor adjust-
ment costs and measurement errors.\textsuperscript{2} The magnitude of labor misallocation turns out to be much smaller than that of capital misallocation. A complete removal of labor misallocation would increase China’s manufacturing output by less than 5 percent.

Another interesting exercise is to understand the policies or institutional arrangements lying hidden behind the veil of misallocation. Although such distortions are not directly observed, the generalized ARP approach suggests that once heterogeneities in output and demand elasticities are properly controlled, the between-group variation of the revenue-capital ratio would play a key role in identification. Motivated by the insight, we regress the time-series mean of the revenue-capital ratio of each firm on a set of firm characteristics. We find that small, young and non-state Chinese firms are associated with significantly higher revenue-capital ratios than their counterparts that are large, mature, and state-owned. These results are broadly consistent with the findings from a growing literature on financial market imperfections in China.\textsuperscript{3}

Within the growing literature studying the role of particular distortions, Midrigan and Xu (2014) evaluate the importance of a particular collateral constraint on aggregate productive efficiency. They find a quantitatively small effect on surviving firms through the misallocation channel. The main insight is that self-financing can undo the losses caused by the collateral constraint. Using firm-specific borrowing costs for U.S. manufacturing firms directly from the interest rate spreads on their outstanding publicly-traded debt, Gilchrist et al. (2013) also find very modest losses in aggregate TFP.\textsuperscript{4} We estimate the magnitude of capital misallocation on surviving firms caused by all kinds of financial frictions, policy distortions and institutional arrangements. But our exercise is completely silent on entry and exit.

Asker et al. (2014) show that capital adjustment costs can be an important contributing factor to the observed misallocation. Our approach differs from theirs in two aspects. First, methodologically, like the ARP approach, we aim to back out the magnitude of misallocation by matching empirical moments. Asker et al. (2014), instead, explores the extent to which capital adjustment costs alone can explain the correlations between the dispersion and the time-series volatility of productivity across industry and country. Second, empirically, our structural approach, which can estimate capital misallocation and adjustment costs simultaneously, finds the misallocation to be significant and quantitatively important in China’s manufacturing sector.

In terms of structural estimation, Cooper and Haltiwanger (2006) and Bloom (2009) first

\textsuperscript{2}See Cooper et al. (2010) for a structural estimation of labor adjustment costs in Chinese manufacturing.

\textsuperscript{3}See, e.g., Dollar and Wei (2007), Brandt et al. (2013), Hsieh and Song (2014).

\textsuperscript{4}The importance of credit market imperfections on aggregate productive efficiency is far from being settled, however. See Caselli and Gennaioli (2013), Buera et al. (2011), Buera and Shin (2013) and Moll (2014) for different results.
adopt the simulated method of moments to recover structural parameters of capital adjustment costs. We contribute to the empirical investment literature by estimating unobserved heterogeneities and measurement errors.

The rest of the paper is organized as follows. Section 2 outlines the simple model economy with unobserved heterogeneities in production technology and market power. We then develop the generalized ARP approach in the simple economy. Section 3 introduces the full-blown model and the structural estimation. We apply both approaches to the China and U.S. data in Section 4. Section 5 discusses some applications of the generalized ARP approach. Section 6 concludes.

2 A Simple Model

To illustrate the main idea of this paper, we begin with a simple model with two basic features. First, firms face heterogeneous capital goods prices due to capital market distortions. Second, capital output elasticity and markups differ across firms. The full-blown model with capital adjustment costs and measurement errors will be presented in Section 3.

2.1 Production and Demand

Firm \(i\) in period \(t\) uses productive capital, labor and intermediate input, denoted by \(K_{i,t}, L_{i,t}\) and \(M_{i,t}\), respectively, to produce \(Q_{i,t}\) units of good \(i\). The production technology exhibits constant returns to scale and takes a Cobb-Douglas form:

\[
Q_{i,t} = A_{i,t} K_{i,t}^{\alpha_i} L_{i,t}^{\beta_i} M_{i,t}^{1-\alpha_i-\beta_i},
\]

where \(A_{i,t}\) is stochastic, representing randomness in productivity; \(\alpha_i > 0\) and \(\beta_i > 0\) denote firm-specific capital and labor output elasticities, respectively, \(\alpha_i + \beta_i < 1\).

The firm sells its goods in a monopolistic product market, subject to an isoelastic downward-sloping demand curve,

\[
Q_{i,t} = X_{i,t} P_{i,t}^{-\frac{1}{\eta_i}}.
\]

Here, \(X_{i,t}\) is stochastic, representing randomness in demand; \(P_{i,t}\) denotes the price of good \(i\) in period \(t\), and \(\eta_i \in (0, 1)\) is the inverse of firm-specific demand elasticity. Alternatively, one may interpret the heterogeneity in \(\eta_i\) as product market distortions. But our estimation of capital market distortions is independent of the interpretation.

For notational convenience, we define \(Y_{i,t} = P_{i,t} Q_{i,t}\) as sales revenue. A combination of (1) and (2) leads to

\[
Y_{i,t} = X_{i,t}^\eta_i A_{i,t}^{1-\eta_i} \left( K_{i,t}^{\alpha_i} L_{i,t}^{\beta_i} M_{i,t}^{1-\alpha_i-\beta_i} \right)^{1-\eta_i}.
\]
Denote \( w_{i,t} \) and \( m_{i,t} \) as wage rate and intermediate input price, respectively. For a given productive capital stock \( \bar{K}_{i,t} \), firm \( i \) chooses \( L_{i,t} \) and \( M_{i,t} \) to maximize its gross profits, denoted by \( \pi_{i,t} \):
\[
\pi_{i,t} = \max_{L_{i,t},M_{i,t}} \{Y_{i,t} - w_{i,t}L_{i,t} - m_{i,t}M_{i,t}\},
\]
where \( Y_{i,t} \) follows (3). Both capital income and markups are in the gross profits, \( \pi_{i,t} \).

The first-order conditions imply constant labor, intermediate input and profit shares:
\[
\frac{w_{i,t}L_{i,t}}{Y_{i,t}} = \beta_i (1 - \eta_i),
\]
\[
\frac{m_{i,t}M_{i,t}}{Y_{i,t}} = (1 - \alpha_i - \beta_i) (1 - \eta_i),
\]
\[
\frac{\pi_{i,t}}{Y_{i,t}} = \eta_i + \alpha_i (1 - \eta_i).
\]
The labor and intermediate input shares would reduce to \( \beta_i \) and \( 1 - \alpha_i - \beta_i \) in the limiting case where the demand elasticity goes to infinity (i.e., \( \eta_i = 0 \)). Accordingly, the profit-revenue ratio would be identical to \( \alpha_i \) as profits are just capital income. (7) will be a key equation for identifying the heterogeneities of \( \alpha_i \) and \( \eta_i \).

The optimization also establishes a profit function:
\[
\pi_{i,t} = Z_{i,t}^{\gamma_i} \bar{K}_{i,t}^{1-\gamma_i},
\]
where
\[
\gamma_i = 1 - \frac{\alpha_i (1 - \eta_i)}{\eta_i + \alpha_i (1 - \eta_i)},
\]
and \( Z_{i,t} \) encompasses productivity, demand and input prices.\(^5\) One may consider \( Z_{i,t} \) “profitability” (Cooper and Haltiwanger, 2006) or “business environment” (Bloom, 2009). Since the marginal revenue product of capital (MRPK henceforth), \( \partial Y_{i,t}/\partial \bar{K}_{i,t} \), is identical to \( \partial \pi_{i,t}/\partial \bar{K}_{i,t} \), we have the following representation of MRPK:
\[
MRPK_{i,t} = \alpha_i (1 - \eta_i) \frac{Y_{i,t}}{\bar{K}_{i,t}} = (1 - \gamma_i) \left( \frac{Z_{i,t}}{\bar{K}_{i,t}} \right)^{\gamma_i}.
\]
The first and second equalities come from (3) and (8), respectively.

2.2 Distortions and Misallocation

There is a long list of distortions that would cause capital misallocation. We do not need to specify each of the distortions since the goal is to back out the magnitude of their overall effect.

\(^5\) \( Z_{i,t} \equiv \left( \frac{w_{i,t}}{\bar{w}_i} \right)^{\gamma_i} \left[ (1 - \eta_i)^{1-\alpha_i} \left( \frac{\beta_i}{\bar{w}_i} \right)^{\beta_i} \left( \frac{1-\alpha_i-\beta_i}{w_{i,t}} \right)^{1-\alpha_i-\beta_i} \right]^{1-\gamma_i} X_{i,t} A_{i,t}^{\frac{1}{\gamma_i} - 1} \).
Use $\tau_i$ to summarize the effects of various capital market distortions on the capital goods price that firm $i$ faces:

$$P_{i,t}^K = (1 + \tau_i) P_{t}^K,$$

(11)

where $P_{t}^K$ denotes the average capital goods price. A positive value of $\tau_i$ may correspond to the case that firm $i$ has limited access to external financing and, hence, is subject to a higher than average capital goods price. A negative value of $\tau_i$, on the other hand, may represent an investment tax credit.

Denote $I_{i,t}$ and $K_{i,t}$ the new investment and capital at the beginning of each period $t$, respectively. $I_{i,t}$ contributes to the productive capital, $\hat{K}_{i,t}$, immediately within period $t$. $K_{i,t}$ depreciates at the end of that period. The law of motion for capital is given by

$$K_{i,t+1} = (1 - \delta) \hat{K}_{i,t}$$

$$= (1 - \delta) (K_{i,t} + I_{i,t}),$$

(12)

where $\delta$ is the depreciation rate.

Optimal investment is chosen to maximize the discounted present value of dividends, which is the profit net of investment expenditure. Risk-neutral investors allocate capital to maximize the sum of future dividends, which are discounted at the required rate of return, $r$.$^6$

Following Bloom (2009), our timing assumption on investment allows for a closed-form solution in the simple model, which provides a convenient analytical benchmark. Denote $J_t$ the Jorgensonian user cost of capital:

$$J_t \equiv P_{t}^K - \frac{1 - \delta}{1 + r} E_t \left[ P_{t+1}^K \right].$$

(13)

Appendix 7.2 shows that

$$\hat{K}_{i,t} = \left[ \frac{1 - \gamma_i}{(1 + \tau_i) J_t} \right]^{\frac{1}{\gamma_i}} Z_{i,t}.$$

(14)

Intuitively, a firm facing unfavorable capital market distortions ($\tau_i > 0$) ends up with less capital than a firm that is facing favorable distortions ($\tau_i < 0$) but otherwise identical.

Substituting (14) into (10) yields

$$MRPK_{i,t} = \alpha_i (1 - \eta_i) \frac{Y_{i,t}}{K_{i,t}} = (1 + \tau_i) J_t.$$

(15)

$^6$The risk-neutrality assumption is equivalent to having a complete market without aggregate shocks in which risk-averse investors diversify all idiosyncratic risks. A relaxation of the assumption may cause $r$ to vary across firms in a number of ways. Appendix 7.5 will discuss some of the possibilities and how the estimation results would be affected accordingly.
The left- and right-hand sides of (15) represent MRPK and the firm-specific user cost of capital, respectively. In the absence of distortions, MRPK would be identical across firms.\(^7\) In the presence of distortions, \(\log(MRPK_{i,t})\) is proportional to \(\log(1 + \tau_i)\). Denote \(\sigma_\tau\) the standard deviation of \(\log(1 + \tau_i)\) across firms. Appendix 7.1 shows that \(\sigma_\tau\) is a summary statistics of the magnitude of capital misallocation: The aggregate output gain of removing capital misallocation is proportional to \(\sigma^2_\tau\). We will, thus, focus on the identification and estimation of \(\sigma_\tau\), the parameter of our primary interest.

2.3 The ARP Approach

We are now ready to demonstrate the potential bias of the ARP approach. Assume that each firm has a firm-specific \(\tau_i\), where \(\log(1 + \tau_i)\) is drawn independently from an identical normal distribution with mean zero and standard deviation \(\sigma_\tau\):

\[
\log(1 + \tau_i) \overset{i.i.d.}{\sim} N(0, \sigma^2_\tau). \tag{16}
\]

We also allow \(\alpha_i\) and \(\eta_i\) to be firm-specific. They are drawn independently from the following distributions:

\[
\log \alpha_i \overset{i.i.d.}{\sim} N(\mu_{\log \alpha}, \sigma^2_{\log \alpha}), \tag{17}
\]

\[
\log \eta_i \overset{i.i.d.}{\sim} N(\mu_{\log \eta}, \sigma^2_{\log \eta}). \tag{18}
\]

\(\alpha_i\) and \(\eta_i\) are truncated to exclude unrealistic values. Two remarks are in order. First, the log-normality assumptions can well capture the skewness of the profit-revenue and revenue-capital ratios in the data (see the structural estimation results in Section 4.4). Second, \(\alpha_i\) and \(\eta_i\) are exogenous. We will relax this assumption in Section 5.2, which allows \(\tau_i\) to affect \(\alpha_i\).

We assume that \(Z_{i,t}\) follows a trend stationary AR(1) process:

\[
\log Z_{i,t} = \mu t + z_{i,t}, \tag{19}
\]

\[
z_{i,t} = \rho z_{i,t-1} + e_{i,t},
\]

where \(0 < \rho < 1\), \(e_{i,t} \overset{i.i.d.}{\sim} N(0, \sigma^2)\), and \(z_{i,0} = 0.\(^8\) The standard deviation of the shocks, \(\sigma\), is the parameter characterizing the level of uncertainty. We assume homogeneous \(\mu\) and \(\sigma\) in the

---

\(^7\)This is because of the timing assumption on \(\bar{K}_{i,t}\). It is reassuring that the average revenue-capital ratio, a key variable for estimating \(\sigma_\tau\), has very similar empirical distributions in our samples regardless of whether the denominator is \(\bar{K}_{i,t}\) or \(K_{i,t}\). Therefore, the timing assumption should have little effect on our results.

\(^8\)The stochastic process of \(Z_{i,t}\) can be endogenously obtained from its definition, if we assume that each of \(A_{i,t}, X_{i,t}, w_{i,t}\) and \(m_{i,t}\) follow a similar trend stationary AR(1) process. For (19) to hold, a sufficient condition is that these four random variables share a common level of persistence, \(\rho\), and the shocks to each of these random variables are independent.
benchmark case. Appendix 7.7 shows that a relaxation of the assumption will not cause any substantial changes to our main results.

Rearrange (15):

$$\log \left( \frac{Y_{i,t}}{K_{i,t}} \right) = \log J_t + \log (1 + \tau_i) - \log [\alpha_i (1 - \eta_i)].$$

(20) is a cornerstone of the ARP approach in the misallocation literature. It shows how to infer $\sigma_\tau$ from the dispersion of the revenue-capital ratio. However, one challenge in the indirect inference is that, besides capital market distortions, unobserved heterogeneities in $\alpha_i$ and $\eta_i$ also cause the revenue-capital ratio to differ across firms. Under the assumption that $\tau_i$, $\alpha_i$ and $\eta_i$ are independent of each other, heterogeneities in $\alpha_i$ and $\eta_i$ would bias upwards the estimated $\sigma_\tau$.

### 2.4 The Generalized ARP Approach

We are ready to propose a new method, which generalizes the ARP approach along two dimensions. First, the standard ARP approach infers $\sigma_\tau$ by matching the variance of the revenue-capital ratio only. The generalized ARP approach will, instead, explore the second moments of both the revenue-capital and profit-revenue ratios, in order to control for the unobserved heterogeneities in $\alpha_i$ and $\eta_i$. Second, while cross-sectional data are enough for the ARP approach, the generalized ARP approach will explore panel data. For reasons that will be discussed in Sections 3, using the between-group dispersions and correlations allows us to eliminate some of the potential bias caused by capital adjustment costs and measurement errors.

The main challenge of identifying $\sigma_\tau$ in the simple model is how to deal with heterogeneities in $\alpha_i$ and $\eta_i$. (7) suggests that the dispersion of the profit-revenue ratio is informative for $\sigma_{\log \alpha}$ and $\sigma_{\log \eta}$. This, however, is not enough. We have two empirical moments – i.e., the variances of $\log \left( \frac{Y_{i,t}}{K_{i,t}} \right)$ and $\pi_{i,t}/Y_{i,t}$, while there are three parameters governing unobserved heterogeneities: $\sigma_\tau$, $\sigma_{\log \alpha}$ and $\sigma_{\log \eta}$. To resolve the under-identification issue, we introduce the cross correlation between $\log \left( \frac{Y_{i,t}}{K_{i,t}} \right)$ and $\pi_{i,t}/Y_{i,t}$, which follows:

$$\text{corr} \left[ \frac{\pi_{i,t}}{Y_{i,t}}, \log \left( \frac{Y_{i,t}}{K_{i,t}} \right) \right] \begin{cases} < 0, & \text{if } \sigma_{\log \alpha} > 0 \text{ and } \sigma_{\log \eta} = 0 \\ > 0, & \text{if } \sigma_{\log \alpha} = 0 \text{ and } \sigma_{\log \eta} > 0 \end{cases}.$$  

(21)

Intuitively, higher markups ($\eta_i$) increase both the profit-revenue and revenue-capital ratios, while a larger $\alpha_i$ increases the profit-revenue ratio but decreases the revenue-capital ratio. In extreme cases, if there is no heterogeneity in $\eta_i$ ($\alpha_i$), the profit-revenue ratio would be negatively
(positively) correlated with the revenue-capital ratio. Therefore, the sign and magnitude of
the correlation help to pin down the relative importance of $\sigma_{\log \alpha}$ and $\sigma_{\log \eta}$.

Based upon the above identification condition, the generalized ARP approach uses five core
moments to back out the five parameters governing the distributions in (16) to (18): $\sigma\tau$, $\mu_{\log \alpha}$,
$\mu_{\log \eta}$, $\sigma_{\log \alpha}$ and $\sigma_{\log \eta}$. The five moments are means of $\pi_{i,t}/Y_{i,t}$ and $\log \left(Y_{i,t}/\bar{K}_{i,t}\right)$, between-
group standard deviations of $\pi_{i,t}/Y_{i,t}$ and $\log \left(Y_{i,t}/\bar{K}_{i,t}\right)$, and the between-group correlation of $\pi_{i,t}/Y_{i,t}$ and $\log \left(Y_{i,t}/\bar{K}_{i,t}\right)$, denoted as $\text{mean} (\pi/Y)$, $\text{mean} (\log (Y/\bar{K}))$, $\text{bsd} (\pi/Y)$, $\text{bsd} \left(\log \left(Y/\bar{K}\right)\right)$ and $\text{bcorr} \left(\pi/Y, \log \left(Y/\bar{K}\right)\right)$, respectively.$^9$ This constitutes five equations
with five unknown variables:

$$
\begin{bmatrix}
\text{mean} \left(\frac{\pi}{Y}\right) \\
\text{mean} \left(\log \left(\frac{Y}{\bar{K}}\right)\right) \\
\text{bsd} \left(\frac{\pi}{Y}\right) \\
\text{bsd} \left(\log \left(\frac{Y}{\bar{K}}\right)\right) \\
\text{bcorr} \left(\frac{\pi}{Y}, \log \left(\frac{Y}{\bar{K}}\right)\right)
\end{bmatrix}
= 
\begin{bmatrix}
E [\eta_i] + E [\alpha_i (1 - \eta_i)] \\
\log J - E [\log \alpha_i] - E [\log (1 - \eta_i)] \\
\sqrt{\text{var} [\eta_i + \alpha_i (1 - \eta_i)]} \\
\sqrt{\sigma^2 + \text{var} [\log \alpha_i (1 - \eta_i)]} \\
\text{corr} [\eta_i + \alpha_i (1 - \eta_i), \log (1 + \tau_i) - \log [\alpha_i (1 - \eta_i)]]
\end{bmatrix}.
$$

(22)

The first and third equations are about $\text{mean} (\pi/Y)$ and $\text{bsd} (\pi/Y)$, based on (7). The second
and fourth equations are about $\text{mean} \left(\log \left(\frac{Y}{\bar{K}}\right)\right)$ and $\text{bsd} \left(\log \left(\frac{Y}{\bar{K}}\right)\right)$, based on (20). The
last equation is about $\text{bcorr} \left(\pi/Y, \log \left(\frac{Y}{\bar{K}}\right)\right)$, which follows (21). The whole approach boils
down to solving a non-linear equation system. The system can easily be solved numerically.
The convergence of numerical solution turns out to be very fast and independent of initial
guess in all exercises that will be conducted below. Therefore, the generalized ARP approach
preserves some tractability of the ARP approach.

We examine numerically the identification condition of the generalized ARP approach in
a simulated panel of 100,000 firms and 24 years, where moments are calculated by data from
the last four years. The construction is consistent with the size of a balanced panel from
China’s industrial survey involving about 100,000 firms over 2004-2007 which will be used
in the following sections. All simulations assume constant $P^K_t$, normalized to unity, and set
$r = 0.15$, $\delta = 0.05$, $\rho = 0.9$ and the steady state growth rate of $Z_{i,t}$ to 0.05. We set the means
of $\log \alpha$ and $\log \eta$ to $-2.5$ and let the standard deviation of the shock to $Z_{i,t}$ equal 0.4, which

$^9$The between-group correlation is defined as follows:

$$
\text{bcorr} \left(\pi/Y, \log \left(\frac{Y}{\bar{K}}\right)\right) \equiv \text{corr} \left[\frac{1}{T} \cdot \sum_{t=1}^{T} \pi_{i,t}/Y_{i,t}, \frac{1}{T} \cdot \sum_{t=1}^{T} \log \left(Y_{i,t}/\bar{K}_{i,t}\right)\right]
$$

9
fall in the range of the values estimated from the China and U.S. data. The results reported below turn out to be very robust with various parameter values.

Panel A of Table 1 reports the five moments of the simulated data: Column (1) starts with a model with no unobserved heterogeneities. We then add positive $\sigma_\tau$, $\sigma_{\log \alpha}$ and $\sigma_{\log \eta}$, respectively, in Column (2) to (4). Column (2) shows that only $bsd \left( \log \left( \frac{Y}{\bar{K}} \right) \right)$ responds to $\sigma_\tau$. In Column (3), $\sigma_{\log \alpha} > 0$ increases both $bsd \left( \frac{\pi}{Y} \right)$ and $bsd \left( \log \left( \frac{Y}{\bar{K}} \right) \right)$. As predicted by (21), Column (3) and (4) show that $\sigma_{\log \alpha} > 0$ and $\sigma_{\log \eta} > 0$ lead to negative and positive $bcorr \left( \frac{\pi}{Y}, \log \left( \frac{Y}{\bar{K}} \right) \right)$, respectively. The effect of $\sigma_{\log \eta}$ on $bsd \left( \log \left( \frac{Y}{\bar{K}} \right) \right)$ appears to be much smaller than the effect of $\sigma_{\log \alpha}$. The last column lists the moments in the model where all the unobserved heterogeneities are present. The various responses of the second moments to changes in $\sigma_\tau$, $\sigma_{\log \alpha}$ and $\sigma_{\log \eta}$ illustrate how these heterogeneities can be identified by the generalized ARP approach.

We then apply the ARP and generalized ARP approaches to the simulated data. Panel B of Table 1 presents the inferred values of $\sigma_\tau$. As expected, the ARP approach is unbiased only in the simple model with no heterogeneities in $\sigma_{\log \alpha}$ and $\sigma_{\log \eta}$. By contrast, the generalized ARP approach delivers unbiased estimates in all cases.

3 The Full-Blown Model

The simple model shows that the generalized ARP approach isolates capital market distortions from unobserved heterogeneities in $\alpha_i$ and $\eta_i$. There are other factors that are missing in the simple model but may potentially contaminate the inference of $\sigma_\tau$. For instance, capital adjustment costs would cause MRPK to vary across firms even if no distortions are present. Measurement errors, which tend to be more significant in firm-level data from developing economies, are another important issue. To address these concerns, we now turn to a full-blown model that incorporates not only the unobserved heterogeneities, but also capital adjustment costs and measurement errors.

3.1 Capital Adjustment Costs

We first introduce capital adjustment costs as a representation of frictions that reduce, delay or protract investment (Khan and Thomas, 2006). The ARP approach would be biased with the presence of such frictions. A simple way of illustrating the bias is to rewrite $J_t$ in (15) as $J_{i,t}$, which denotes the firm-specific user cost of capital. A combination of idiosyncratic shocks
and capital adjustment costs would cause user cost of capital to vary across firms, adding to the dispersion of the revenue-capital ratio.

Following the literature,\(^{10}\) we consider three forms of capital adjustment costs:

\[
G(K_{i,t}; I_{i,t}) = \frac{b q}{2} \left( \frac{I_{i,t}}{K_{i,t}} \right)^2 K_{i,t} - b^f P_{i,t} K_{i,t} 1_{[I_{i,t} < 0]} + b^f 1_{[I_{i,t} \neq 0]} \pi_{i,t},
\]

where \(G(K_{i,t}; I_{i,t})\) represents the function of capital adjustment costs, with \(1_{[I_{i} < 0]}\) and \(1_{[I_{i} \neq 0]}\) being indicators for negative and non-zero investment; \(b q\) measures the magnitude of quadratic adjustment costs; \(b^f\) can be interpreted as the difference between the purchase price, \(P_{i,t} K\), and the resale price expressed as a percentage of the purchase price of capital goods; finally, \(b^f\) stands for the fraction of gross profit loss due to any non-zero investment.

The model is disciplined by restricting the capital adjustment cost function, \(G\), to be homogenous across firms. If \(G\) were heterogeneous, a firm facing larger capital adjustment costs, holding all else equal, would manifest such costs as a high \(\tau_i\). A caveat is that \(G\) may vary across industries. An auto production line, for instance, is more irreversible than office furniture. Allowing industry-specific \(G\), however, gives essentially the same estimated \(\sigma_{\tau}.\)

### 3.2 Measurement Errors

We next introduce measurement errors. The benchmark specification assumes that

\[
K_{i,t} = K_{i,t}^{true} \exp(e_{i,t}^K), \quad e_{i,t}^K \sim i.i.d. N(0, \sigma_{meK}^2),
\]

\[
Y_{i,t} = Y_{i,t}^{true} \exp(e_{i,t}^Y), \quad e_{i,t}^Y \sim i.i.d. N(0, \sigma_{meY}^2),
\]

\[
\pi_{i,t} = \pi_{i,t}^{true} (1 + e_{i,t}^\pi), \quad e_{i,t}^\pi \sim i.i.d. U[-\sigma_{me\pi}, \sigma_{me\pi}].
\]

Here, variables with and without the “true” superscript denote the true states and their observed counterparts in the data, respectively. \(e_{i,t}^K\) and \(e_{i,t}^Y\) are measurement errors in capital and revenue, respectively. They are drawn independently from an identical normal distribution with mean zero and standard deviation \(\sigma_{meK}\) and \(\sigma_{meY}\), respectively. \(e_{i,t}^\pi\) stands for measurement errors in profit. It follows a uniform distribution \(U[-\sigma_{me\pi}, \sigma_{me\pi}].\) The multiplicative structure and the log-normality assumption guarantee positive values of capital stock and sales revenue, while the reported profits are allowed to be negative.

We consider transitory measurement errors only. This is because persistent measurement errors are by nature indistinguishable from unobserved firm characteristics. Still, abstracting

\(^{10}\)See, for example, Abel and Eberly (1994), Cooper and Haltiwanger (2006) and Bloom (2009).

\(^{11}\)Specifically, we estimate the model using two subsamples that consist of firms in the ten least and most capital-intensive industries. The manufacturing capital intensity rank follows Song et al. (2011). The results are available upon request.
such errors may bias the estimate of $\sigma_r$. To address this concern, we will model transitory measurement errors in investment $I_{i,t}$ and allow $K_{i,t}$ to accumulate the measurement errors according to the law of motion of capital (12). Introducing persistent measure errors in capital via this form has little effect on our main findings (see Appendix 7.7).

3.3 Identification of $\sigma_r$ in the Full-Blown Model

Once again, we simulate a panel of 100,000 firms and 24 years and use the last four years only to compute the five moments. The benchmark economy is parameterized as those in Column (5) of Table 1.

We first introduce quadratic capital adjustment costs. Panel A of Figure 1 plots $bsd\left(\log \left(\frac{Y}{K}\right)\right)$ and $sd\left(\log \left(\frac{Y}{K}\right)\right)$ with respect to $b^q$. On the one hand, both $bsd\left(\log \left(\frac{Y}{K}\right)\right)$ and $sd\left(\log \left(\frac{Y}{K}\right)\right)$ remain essentially flat for modest values of $b^q$. Under the benchmark parameterization, the variance of the revenue-capital ratio caused by unobserved heterogeneities predominates that caused by modest capital adjustment costs. This explains the flat part of the standard deviations, which, in turn, suggests that the generalized ARP approach would not be biased much by modest quadratic capital costs. Column (1) of Table 2 shows that the inferred value of $\sigma_r$ is only 6 percent below its true value if we set $b^q$ to 1, close to the maximum value estimated from our China and U.S. samples. Notably, the generalized ARP approach underestimates $\sigma_r$.\textsuperscript{12} In other words, in contrast to the upward bias of the ARP approach caused by the presence of capital adjustment costs, the generalized ARP approach delivers a lower bound estimate.

\textsuperscript{12}A higher $b^q$ increases the mean of the revenue-capital ratio. The generalized ARP approach would (incorrectly) adjust upwards the inferred values of both $\mu_{\log \alpha}$ and $\sigma_{\log \alpha}$, due to the lognormal distributive assumption (17). A higher $\sigma_{\log \alpha}$, in turn, would account for a larger share of the dispersion of the revenue-capital ratio and, hence, infer a lower value of $\sigma_r$.\textsuperscript{12}
from the China and U.S. data. The details can be found in Column (3) to (5) of Table A.1 in the appendix. The generalized ARP approach is, therefore, not sensitive to the presence of investment irreversibility and fixed capital adjustment cost.

We next introduce measurement errors in the economy parameterized by those in Column (5) of Table 1. Panel B of Figure 1 plots $bsd \left( \log \left( \frac{Y}{\hat{K}} \right) \right)$ and $sd \left( \log \left( \frac{Y}{\hat{K}} \right) \right)$ with respect to $\sigma_{meK}$. Not surprisingly, measurement errors on capital can easily blow up $sd \left( \log \left( \frac{Y}{\hat{K}} \right) \right)$ by increasing the observed volatility of the revenue-capital ratio. In contrast, $bsd \left( \log \left( \frac{Y}{\hat{K}} \right) \right)$ remains largely flat for $\sigma_{meK}$ below 0.5. $\sigma_{meK}$ starts to have a significant effect on $bsd \left( \log \left( \frac{Y}{\hat{K}} \right) \right)$ for $\sigma_{meK}$ above 0.5. The inferred values of $\sigma_\tau$ are reported in Column (3) to (4) of Table 2.

The generalized ARP approach overestimates $\sigma_\tau$ by 12 percent when $\sigma_{meK} = 0.5$, representing large measurement errors on capital and higher than the value estimated from the China data as will be shown below. Yet, the bias is much smaller than that of 74 percent by the ARP approach, which matches $sd \left( \log \left( \frac{Y}{\hat{K}} \right) \right)$. Although $\sigma_{meK} = 1$ increases the bias of the generalized ARP approach to 42 percent, the bias is still less than a third of that of the ARP approach. Column (5) reports the results when both capital adjustment costs and measurement errors are present. Since $b^\theta$ and $\sigma_{meK}$ bias $\sigma_\tau$ in opposite direction, the inferred value of $\sigma_\tau$ is actually closer to the true value than its counterparts with $b^\theta$ or $\sigma_{meK}$ only.

The effect of $\sigma_{meY}$ on the dispersions of the revenue-capital ratio is identical to that of $\sigma_{meK}$. $\sigma_{me\pi}$ has no effect on the dispersions since measurement errors on profits do not affect the revenue-capital ratio. Finally, $\sigma_\tau$ and $\sigma_{log\alpha}$ continue to have first-order effects on $bsd \left( \log \left( \frac{Y}{\hat{K}} \right) \right)$. In summary, the above properties imply that the extension of the simple model does not invalidate the conditions for the generalized ARP approach to identify $\sigma_\tau$ if capital adjustment costs or measurement errors are sufficiently small.

### 3.4 Structural Estimation

We now propose a structural econometric approach to estimate all the relevant parameters in the full-blown model. This is particularly useful for a sample with serious measurement error issues or with firms that are subject to large capital adjustment costs. The full-blown model will be estimated by the simulated method of moments (SMM). The SMM estimator is defined in Appendix 7.6. The upper panel of Table 3 lists $\Theta$, the set of parameters to estimate. There are a total of 13 parameters, including the key parameter $\sigma_\tau$; mean and standard deviation of $\log \alpha$, $\mu_{log\alpha}$ and $\sigma_{log\alpha}$; mean and standard deviation of $\log \eta$, $\mu_{log\eta}$ and $\sigma_{log\eta}$.

$^{13}$We will discuss in Appendix 7.7 how the results would change if measurement errors in capital entailed a persistent component.
\(\sigma_{\log \eta}\); capital adjustment costs parameters, \(b^q\), \(b^i\) and \(b^f\); the trend growth rate, \(\mu\); standard deviation of idiosyncratic shocks, \(\sigma\); and standard deviations of measurement errors in capital, revenue and profit, \(\sigma_{meK}\), \(\sigma_{meY}\), and \(\sigma_{mep}\).

[Insert Table 3]

The lower panel of Table 3 lists \(\hat{\Phi}^D\), the set of moments to match. There are 21 moments. The choice of the moments is guided by two principles. First, \(\hat{\Phi}^D\) is a comprehensive set of moments that characterize the distribution and dynamics of the relevant variables in the model. Second, and more importantly, these moments are informative about the parameters to estimate. Specifically, \(\hat{\Phi}^D\) includes means (mean), between-group standard deviations (bsd), within-group standard deviations (wsd), coefficients of skewness (skew) and serial correlations (scorr) for \(\pi_{i,t}/Y_{i,t}\), \(\log \left(Y_{i,t}/\bar{K}_{i,t}\right)\), \(I_{i,t}/K_{i,t}\) and \(\Delta \log Y_{i,t}\), together with the cross correlation (bcorr) between the between-group \(\pi_{i,t}/Y_{i,t}\) and \(\log \left(Y_{i,t}/\bar{K}_{i,t}\right)\). The following section will establish the identification conditions through which \(\Theta\) can be estimated by matching these moments.

The investment policies, which have to be solved numerically in the presence of capital adjustment costs, differ across firms with various \((\tau_i, \alpha_i, \eta_i)\). To reduce the computational burden, we adopt a standard approach in the literature (e.g., Eckstein and Wolpin, 1999) by considering a finite type of firms. Our benchmark specification assumes \(3 \times 3 \times 3\) types of firms. Each consists of a fixed proportion; i.e., \(1/(3 \times 3 \times 3)\), of the population. The type set is defined as \(F = \{(\tau_u, \alpha_v, \eta_x) : u = 1, 2, 3; v = 1, 2, 3; x = 1, 2, 3\}\). Appendix 7.7 will experiment with increasing the types of firms to \(5 \times 5 \times 5\). The results are essentially the same.

### 3.4.1 Identification of Capital Adjustment Costs and Measurement Errors

One advantage of the structural approach is to estimate capital adjustment costs and measurement errors, which cannot be done by the generalized ARP approach. We first present the identification condition for capital adjustment costs. Following the routine in the literature (e.g., Bloom, 2009), our identification uses information on the investment rate, \(I_{i,t}/K_{i,t}\), and the revenue growth rate, \(\Delta \log Y_{i,t}\), to identify \(b^q\), \(b^i\) and \(b^f\). Column (1) in Table A.1 reports the full set of moments based on the same parameter values from Column (5) in Table 1. As an illustrative example, we add positive \(b^q\), \(b^i\) and \(b^f\), respectively, to Column (2) to (4). Column (5) lists the moments when \(b^q\), \(b^i\) and \(b^f\) are all positive.

Two results are relevant for identification. First, the moments for \(I_{i,t}/K_{i,t}\) are much more sensitive than those for \(\Delta \log Y_{i,t}\) in response to changes in capital adjustment costs. This dif-
ference distinguishes capital adjustment costs from the stochastic process of $\log Z_{i,t}$. Moreover, $b^{q} > 0$ and $b^{i} > 0$ decrease $wsd(I/K)$ and increase $scorr(I/K)$, $b^{i} > 0$ and $b^{f} > 0$ increase $skew(I/K)$, while $b^{f} > 0$ has little effect on $wsd(I/K)$ and $scorr(I/K)$. These properties distinguish different forms of capital adjustment costs from each other.

We now add measurement errors. Once again, let us start with Column (1) in Table A.2, which is replicated from the last column in Table A.1. Columns (2) to (4) in Table A.2 reveal the moments that are informative about measurement errors by adding positive $\sigma_{meK}$, $\sigma_{meY}$ and $\sigma_{mex}$, respectively. Column (5) reports the moments when $\sigma_{meK}$, $\sigma_{meY}$ and $\sigma_{mex}$ are all positive. We find that $\sigma_{meK}$ only affects moments on $\log \left( \frac{Y_{i,t}}{\bar{K}_{i,t}} \right)$ and $I_{i,t}/K_{i,t}$; $\sigma_{meY}$ only affects moments on $\log \left( \frac{Y_{i,t}}{\bar{K}_{i,t}} \right)$, $\pi_{i,t}/Y_{i,t}$ and $\Delta \log Y_{i,t}$; and $\sigma_{mex}$ only affects moments on $\pi_{i,t}/Y_{i,t}$. The three types of measurement errors can, thus, be distinguished from each other.

The remaining challenge is to separate measurement errors from capital adjustment costs. Although capital adjustment costs and measurement errors have qualitatively similar effects on $wsd \left( \log \left( \frac{Y}{\bar{K}} \right) \right)$, their effects differ on other moments. In particular, both $\sigma_{meK} > 0$ and $\sigma_{meY} > 0$ increase $wsd(I/K)$ and $wsd(\Delta \log Y)$ and reduce $scorr(I/K)$ and $scorr(\Delta \log Y)$, while capital adjustment costs have the opposite or no effect on these moments. These properties guarantee the identification of measurement errors.

4 Data and Results

4.1 Data

We first use the firm-level data from China’s Annual Survey of Industry conducted by the National Bureau of Statistics. The dataset (henceforth, the NBS dataset) includes all industrial firms that are identified as state-owned or as non-state firms with sales revenue above RMB 5 million.\footnote{These firms account for about 90 percent of the total industrial output.} Since the model is entirely silent on entry and exit, we will focus on a balanced panel from 2004 to 2007, covering 107,579 firms. We take 2004 as the beginning year, when the number of firms increases by a third due to an economic census conducted in that year. Balanced panels with years earlier than 2004 will, hence, involve substantially fewer firms.

Appendix 7.3 provides detailed information on how to clean the data and to construct some of the key variables in the model. In particular, we measure $\pi$ by the difference between sales and the cost of goods sold. Ideally, $\pi$ should correspond to the difference between sales and the cost of labor and intermediate inputs. Since the cost of labor is known to be poorly measured in the NBS dataset, we use instead the cost of goods sold, which covers material, labor and
overhead for production.\footnote{The cost of goods sold includes: (i) parts, raw materials and supplies used; (ii) labor, including associated costs such as payroll taxes and benefits; and (iii) overhead of the business allocable to production.}

We also use two Compustat samples over 2002-2005. The first one, referred to as Compustat I henceforth, covers U.S. manufacturing firms with sales revenue above USD 0.6 million in 2004 prices. The threshold is chosen to match its counterpart in the NBS sample, where most non-state firms have sales above RMB 5 million. The second sample, referred to as Compustat II henceforth, follows Bloom (2009) by including U.S. manufacturing firms with sales above USD 10 million in 2000 prices and more than 500 employees. This is, therefore, a more homogeneous sample composed of large firms only. See Appendix 7.4 for more details.

4.2 Predetermined Parameters

In addition to the five parameters, $\delta$, $r$ and $P^K_t$ also affect the revenue-capital ratio through $J_t$. There is no obvious time trend in the revenue-capital ratio in any of the samples. So, we assume $P^K_t$ to be constant and normalize it to unity. (13) implies that $J_t = J$, where $J \equiv (r + \delta) / (1 + r)$. $\delta$ is set equal to its value used in constructing real capital stock (see Appendix 7.3 and 7.4 for details). Bai, Hsieh and Qian (2006) find a high and fairly stable aggregate rate of return to capital in China over the period 1978-2004, ranging from 20 to 25 percent in most years. The rate of return is even higher for the secondary sector, which includes mining, manufacturing and construction. We impose a conservative value, $r = 0.20$, for manufacturing firms in the NBS sample. We set $r = 0.10$ for Compustat firms.\footnote{This is consistent with the fact that Compustat samples have much lower revenue-capital ratios than the NBS sample (see Table 1 below for details).}

Our main findings are robust to alternative values of $r$.\footnote{Table A.5 shows the results from robustness tests in a full-fledged model. The results in the simple model are available upon request.}

4.3 The Generalized ARP Approach

The first column of Table 4 reports the results from the NBS sample. The generalized ARP approach finds $\sigma_r = 0.684$. The value of $\sigma_r$ is quantitatively large. According to (16), it implies that a firm with $\tau_i$ at the 75th percentile would face a capital goods price 2.5 times higher than the price for a firm with $\tau_i$ at the 25th percentile.

[Insert Table 4]

We also find large values of $\sigma_{\log \alpha}$ and $\sigma_{\log \eta}$, suggesting the quantitative importance of heterogeneities in $\alpha_i$ and $\eta_i$. Under the log-normality specification (17), the estimated $\mu_{\log \alpha}$
and $\sigma_{\log \alpha}$ imply that $\alpha_i$ has a mean of 0.086 and a standard deviation of 0.051. Both the mean and standard deviation are close to those in the literature that estimates the capital output elasticity in a three-factor model.\(^\text{18}\) By (18), the estimated $\mu_{\log \eta}$ and $\sigma_{\log \eta}$ imply that $\eta_i$ has a mean of 0.082 and a standard deviation of 0.076. This translates into markups of 1.090.

The value of $\sigma_\tau$ inferred by the ARP approach would be identical to any of the cross-sectional standard deviation of the log revenue-capital ratio, which is very stable and around 0.89 during the sample period. Using $bsd \left( \log \left( \frac{Y}{K} \right) \right)$, rather than a cross-sectional standard deviation, would reduce $\sigma_\tau$ to 0.867, which is still considerably higher than the value obtained by the generalized ARP approach. (20) suggests that the difference between the variance of $\log \left( \frac{Y}{K} \right)$, 0.867\(^2\), and the calibrated $\sigma_\tau^2$, 0.684\(^2\), be driven by heterogeneities in $\alpha_i$ and $\eta_i$. So, the unobserved heterogeneities would account for 38 percent of $bsd \left( \log \left( \frac{Y}{K} \right) \right)$.

### 4.3.1 Unobserved Heterogeneities within and across Industries

The heterogeneities in $\alpha_i$ and $\eta_i$ could arguably be much smaller within an industry. To mitigate the bias, the literature often applies the ARP approach to each industry. To evaluate the quantitative importance of the within-industry heterogeneities in $\alpha_i$ and $\eta_i$, we calibrate $\sigma_\tau$ in each of the 30 two-digit industries by the generalized ARP approach. The calibrated value of $\sigma_\tau$ has a mean of 0.645 over the 30 industries, similar to the value from the whole sample. The x- and y-axis in Panel A of Figure 2 plot $bsd \left( \log \left( \frac{Y}{K} \right) \right)$ and the calibrated $\sigma_\tau$ in each industry, respectively. In absence of heterogeneities in $\alpha_i$ and $\eta_i$, the calibrated $\sigma_\tau$ should be located on the 45 degree. The figure shows a big difference between $bsd \left( \log \left( \frac{Y}{K} \right) \right)$ and the calibrated $\sigma_\tau$. The mean ratio of the calibrated $\sigma_\tau^2$ to the variance of $\log \left( \frac{Y}{K} \right)$ is 61 percent, very close to the ratio of 62 percent when we pool all industries together.

[Insert Figure 2]

We can further go down to the four-digit level, which has a total of 482 industries in the NBS sample. It deserves attention that the generalized ARP approach fails to find solution in 42 four-digit industries. The reason is two-fold. On the one hand, the empirical moments could easily be influenced by outliers in industries with few firms. In fact, 28 out of the 42 unsuccessful cases involve industries with less than 100 firms. On the other hand, it reminds us of the limitation of the method: It is based on an extremely simple model that could potentially be mis-specified for some industries. We will enrich the model by incorporating a few other

\(^{18}\)For example, Pavcnik (2002) estimated production function using consistent Olley and Pakes (1996) structural estimates for a large sample of Chilean manufacturing plants. The estimated capital shares vary substantially across industries with an average around 0.085.
important elements in the next section. That said, the main findings are essentially the same as before. The ratio of $\sigma_\tau^2$ to the variance of $\log \left( \frac{Y}{K} \right)$ have an average of 62 percent, close to the results using the whole sample or two-digit industries.

To be sure, within-industry dispersions of $\alpha_i$ and $\eta_i$ are indeed smaller than the overall dispersions. The mean of the calibrated $\sigma_\alpha$ and $\sigma_\eta$ across the four-digit industries equals 0.50 and 0.71, respectively, as opposed to 0.55 and 0.79 using the whole sample. But the within-industry heterogeneities appear to overwhelm the heterogeneities across industries.

4.3.2 Heterogeneities in $J$

One caveat with our empirical strategy is that any heterogeneity in $J$ would show up as capital misallocation. Therefore, heterogeneities in $\delta$ or $r$, if they exist, tend to bias the estimate of $\sigma_\tau$ upwards. Heterogeneous $\delta$ would arise if firms have different combinations of plant and equipment that depreciate at different rates. $r$ would also be firm-specific through the market beta channel if we relax the assumption of risk-neutrality. Appendix 7.5 provides some back-of-the-envelope calculations, suggesting a limited role of market beta in accounting for the inferred misallocation. Market incompleteness may also generate heterogeneous $r$ across firms with different levels of uncertainty (e.g., Angeletos and Panousi, 2011). This can be considered one type of financial market imperfection. A back-of-the-envelope calculation in Appendix 7.5 suggests that market incompleteness account for a small fraction of the inferred misallocation from the NBS sample.

As an alternative approach, we calibrate $\sigma_\tau$ for Compustat firms and take it as the benchmark. If the heterogeneity in $J$ has similar magnitude across economies, the difference of the calibrated $\sigma_\tau$ would isolate the difference in the magnitude of capital misallocation for Chinese and U.S. firms.

The generalized ARP approach finds $\sigma_\tau = 0.31$ in Compustat I, much smaller than its counterpart of 0.68 in China. $\sigma_\tau$ becomes much smaller in Compustat II. These are consistent with the findings that China has a highly distorted capital market (see, e.g., Dollar and Wei, 2007; Hsieh and Song, 2014). The small $\sigma_\tau$ for firms in Compustat II is also in line with the well-established fact that larger firms are less likely to be affected by financial market imperfections. In fact, it suggests that the overall magnitude of capital misallocation plus heterogeneity in $J$ appears to be small for large U.S. manufacturing firms, consistent with the finding in Gilchrist et al. (2013).

It should also be emphasized that $bsd \left( \log \left( \frac{Y}{K} \right) \right)$ has the same order of magnitude across the three samples, while the calibrated $\sigma_\tau$ differs a lot. In particular, despite the relatively
small 30-percent difference in $bsd \left( \log \left( \frac{Y}{\hat{K}} \right) \right)$ between NBS and Compustat II, the calibrated of $\sigma_\tau$ essentially goes down to zero in Compustat II. The unobserved heterogeneities in $\alpha_i$ and $\eta_i$ account for a major share of the dispersion of the revenue-capital ratio for the U.S. firms. The ARP approach would, thus, lead to more biased results.

Although we do not model product market distortions, the heterogeneity in $\eta_i$ is isomorphic to the heterogeneity in $\tau^Y_i$ – the measure of product market distortions in Hsieh and Klenow (2009).\(^1\) One may want to interpret heterogeneous markups as product market distortions. Interestingly, the order of the calibrated value of $\sigma_{\log \eta}$ is exactly the same as that of $\sigma_\tau$ across the three samples. The alternative interpretation would suggest the product market to be most distorted for NBS firms and least distorted for firms in Compustat II.

### 4.4 Structural Estimation

We impose the same values as those in the generalized ARP approach for the predetermined parameters. We set $\rho = 0.90$ in (19) for NBS and Compustat firms.\(^2\) We do not estimate $\rho$ structurally since the model cannot distinguish between a stationary process with heterogeneous $\mu$ and $\sigma$ and a unit root process with homogeneous $\mu$ and $\sigma$. Appendix 7.7 will investigate the sensitivity of our estimates to different values of $r$ and $\rho$ in the full-blown model.

Table 5 presents the structural estimation results for NBS firms. The first and second columns of the left panel report the optimal estimates and the corresponding numerical standard errors. Simulated moments at the optimal estimates are listed in the right panel. We also report the corresponding empirical moments, for which the standard errors are obtained by bootstrapping.

\[ \text{[Insert Table 5]} \]

$\sigma_\tau$ has an estimated value of 0.71. The small simulated standard error suggests that the estimate significantly differs from zero. The structurally estimated $\sigma_\tau$ is very close to the one backed out by the generalized ARP approach. The difference is less than 5 percent. The estimates of $\sigma_{\log \alpha}$ and $\sigma_{\log \eta}$ are also highly significant and close to those in Table 4. Overall, the simulated moments provide a close fit to the five core moments, which are key to identifying the unobserved heterogeneities.

\(^1\)See also Peters (2011) where dispersion in markups leads to inefficient TFP losses.

\(^2\) $\rho$ can be calibrated by applying system GMM (Blundell and Bond, 1998) to estimate a dynamic panel data model of $\log \pi_{i,t}$ (e.g., Cooper and Haltiwanger, 2006). The regressors include $\log \pi_{i,t-1}$, $\log K_{i,t}$, $\log K_{i,t-1}$ and year dummies. The estimated autoregressive coefficient is 0.41 for NBS firms, in contrast to 0.89 found by Cooper and Haltiwanger (2006). The substantially lower estimate may reflect the attenuation bias due to large measurement errors in China’s profit data, which will be confirmed by our structural estimation.
The structural estimation finds statistical evidence for quadratic and fixed adjustment costs. As discussed in Section 3.4.1, a positive $b^q$ reflects positive serial correlation of the investment rate and the revenue growth rate, while a positive $b^f$ comes from a larger skewness of the investment rate than that of the revenue growth rate. The point estimates imply that quadratic adjustment costs increase the user cost of capital by 4.5 percent and any investment or disinvestment would cause a loss of 3.4 percent of gross profits in that period.

The estimated $\mu$ is 0.08, in line with Brandt et al.’s (2012) estimate of TFP growth in Chinese manufacturing over 1998-2007. Chinese firms face considerably higher risks. $\sigma$ has an estimated value of 0.42, 60 percent higher than its counterpart of firms in Compustat I that will be reported below.

Two of the three measurement errors we consider turn out to be statistically significant. Consistent with the usual concern about the accuracy of capital and profit data at the firm level, both $\sigma_{meK}$ and $\sigma_{mey}$ are significant and quantitatively large. By contrast, the model finds $\sigma_{mey}$ to be virtually zero, implying a much better measurement of revenue in the NBS sample.

Robustness of the results is reported in Appendix 7.7. We experiment alternative values of $r$ and $\rho$; allow the long-run growth rate of $Z_{i,t}$, $\mu$, and the level of uncertainty, $\sigma$, to be firm-specific; and replace measurement errors in capital with measurement errors in investment. We also increase the number of firm type from $3 \times 3 \times 3$ to $5 \times 5 \times 5$. The estimated value of $\sigma_\tau$ seem very robust with respect to these variations.

We further conduct specification tests in Appendix 7.8. Three different models are estimated in order. The first one assumes homogenous $\alpha_i$ and $\eta_i$, while the second and third take out capital adjustment costs and measurement errors, respectively. In line with the finding that capital adjustment costs and measurement errors in capital with measurement errors in investment. The estimated value of $\sigma_\tau$ in the second and third models are close to that in the benchmark model. Assuming no heterogeneities in $\alpha_i$ and $\eta_i$, instead, leads to a considerably higher estimate of $\sigma_\tau$.

Table 6 presents results for Compustat firms. The estimated value of $\sigma_\tau$ for firms in Compustat I is significant and almost identical to that obtained by the generalized ARP approach. The structural estimation also finds $\sigma_\tau$ to be close to that inferred by the generalized ARP approach for firms in Compustat II. Moreover, the structurally estimated $\sigma_\tau$ is statistically

---

21 Similar to Cooper and Haltiwanger (2006) and Bloom (2009), we also find that only one form of the non-convex adjustment costs is necessary to fit the data. To be specific, Cooper and Haltiwanger (2006) find $b^q > 0$ and $b^f > 0$ for plants in the Longitudinal Research Database; Bloom (2009) finds $b^q > 0$ and $b^f > 0$ for large firms in Compustat. Most firms in the NBS sample are single-plant enterprises. Our finding that a combination of $b^q > 0$ and $b^f > 0$ fits the data best is, hence, in line with Cooper and Haltiwanger (2006).
insignificant. The parameters governing $\alpha_i$ and $\eta_i$ are all highly significant. They are also close to those backed out by the generalized ARP approach.

[Insert Table 6]

A somewhat surprising result is that the estimated quadratic capital adjustment costs are much larger for Compustat firms. This reflects the smaller dispersion, less skewness and more persistence in the data on Compustat firms' investment rate (see the empirical moments in Tables 5 and 6). One possible explanation is that Compustat firms are much larger and have more plants. Firms in Compustat I, for instance, have mean (median) employees of 10.4 (0.8) thousand; in contrast, the corresponding numbers are only 0.33 (0.13) for Chinese firms in the NBS sample. Therefore, the investment of Compustat firms, consolidated across several plants within firm, is more condensed, symmetric and persistent (Bloom, 2009).

In summary, the structurally estimated values of $\sigma_\tau$ in the three samples are all close to those by the generalized ARP approach. The reason is straightforward by the findings in Section 3.3 – i.e., neither capital adjustment costs nor measurement errors are found to be large enough to have a significant effect on $bsd \left( \log \left( \frac{Y}{\hat{K}} \right) \right)$. This is practically useful since the generalized ARP approach maintains some tractability of the ARP approach. In terms of computational costs, the structural estimation typically takes hours to converge, while the generalized ARP approach delivers results within seconds.

4.5 Welfare Implications

The simple model provides a useful framework to quantify the effect of capital misallocation on aggregate output. Take the values inferred by the generalized ARP approach. Reducing $\sigma_\tau$ from 0.68 to zero, holding aggregate capital and labor constant, would increase China's manufacturing output by 20.4 percent. One may also want to cut $\sigma_\tau$ from 0.68 to 0.31, the value for Compustat I firms, as a more conservative experiment. The aggregate output would increase by 16.2 percent. The fall is modest since the aggregate output gain is proportional to the variance of $\tau_i$ (see Appendix 7.1).

The ARP approach naturally implies larger welfare gains. For instance, the ARP approach infers $\sigma_\tau = 0.89$ from the average $sd \left( \log \left( \frac{Y}{\hat{K}} \right) \right)$ for NBS firms. This translates into aggregate output losses of 34.5 percent, which are more than two-thirds larger than that implied by the generalized ARP approach.
5 Extensions

5.1 Labor Misallocation

The generalized ARP approach motivates a simple way of backing out the magnitude of labor misallocation, which we have not addressed. Analogous to (11), we assume that the actual wage rate paid by firm $i$ is

$$w_{i,t} = (1 + \tau_{i}^L) w_t,$$

where $w_t$ is the average wage rate and $\tau_{i}^L$ is a firm-specific component caused by labor market distortions. $\log (1 + \tau_{i}^L)$ follows a normal distribution with mean zero and standard deviation $\sigma_{\tau^L}$. Notice that in the simple model, the presence of $\tau_{i}^L$ does not affect any of the five core moments that are relevant for the inference of $\sigma_{\tau}$.

Substituting (26) back into (5) yields

$$\log \frac{Y_{i,t}}{L_{i,t}} = \log w_t + \log (1 + \tau_{i}^L) - \log [\beta_i (1 - \eta_i)].$$

(27) is akin to (20). (5) allows us to obtain $bsd (\log (\beta (1 - \eta)))$ by computing the between-group standard deviation of labor income share. Then, a combination of (27) and (5) implies a simple way of backing out $\sigma_{\tau^L}$. For the following two reasons, we refer to this method as an extension of the generalized ARP approach for backing out labor misallocation. First, the main challenge, analogous to that for the inference of capital misallocation, is to control for unobservable heterogeneities in output elasticity and markups. Here, the identification is more straightforward because there is a one-to-one mapping between labor income share and $\beta_i (1 - \eta_i)$ in (27). By contrast, we cannot infer $\alpha_i (1 - \eta_i)$ from one minus the sum of the intermediate input and labor shares, which involves both capital income and markups. Second, it explores the between-group dispersion of the revenue-labor ratio in panel data, which mitigates the biases caused by potential labor adjustment costs and measurement errors for exactly the same reasons discussed in Section 3.3.

To address the concern that labor quality differs across firms, we construct efficiency units of labor, $L_{i,t}^e$, as follows:

$$L_{i,t}^e = \sum_h \exp (b \cdot s (h)) L_{i,t} (h),$$

22 The ratio of labor compensations (including benefits) to sales is unusually low (around 0.08) in the NBS sample. One important reason is the severely under-reported labor compensations (Qian and Zhu, 2011). We use the difference between the costs of goods sold and material costs to back out the actual labor compensations. This leads to a labor share of 0.125, averaged over the sample period. Recall that the average capital share is 0.158 for this sample. Therefore, the labor share is roughly equal to the capital share in Chinese manufacturing, which is in line with the aggregate statistics and firm-level evidence in Qian and Zhu (2011).
where $b$ is the Mincerian rate of return to education, $s(h)$ denotes years of schooling for education group $h$ and $L_{i,t}(h)$ is the number of workers in education group $h$. We set $b = 0.10$ and let $s(h)$ be 6, 9, 12 and 16 for employees with primary school education and below, middle school education, high school education and college education and above, respectively.\(^\text{23}\) The firm-level data on educational composition of employees are available only in the 2004 NBS sample. We assume the educational composition in each firm to be constant over time – i.e., $L_{i,t}(h)/L_{i,t} = L_{i,2004}(h)/L_{i,2004}$ for $t > 2004$.

The raw labor data find $\sigma_{\tau,L} = 0.289$. Using efficiency units of labor reduces the value to 0.230. The difference suggests a non-trivial heterogeneity in labor quality across Chinese firms. Both values of $\sigma_{\tau,L}$ are substantially smaller than the calibrated value of $\sigma_{\tau}$. Accordingly, a removal of labor misallocation by reducing $\sigma_{\tau,L}$ from 0.230 to zero would increase aggregate output by 4.9 percent. The formula of calculating aggregate output losses caused by labor misallocation is provided in Appendix 7.1.

### 5.2 Endogenous Capital Output Elasticity

Our empirical specification assumes $\alpha_i$ and $\tau_i$ to be exogenous. This assumption does not seem innocuous if a firm can choose $\alpha_i$. Song et al. (2011), for instance, show that in a two-sector model, financially constrained firms would penetrate the labor-intensive industry, while financially integrated firms would stay in the capital-intensive industry. They also find evidence from Chinese manufacturing that is consistent with the theory. In the context of our model economy, such mechanism would imply a negative correlation between $\alpha_i$ and $\tau_i$, which biases the estimate of $\sigma_{\tau}$. This section develops a model that incorporates a technological choice on $\alpha_i$. Then, we show a simple way of back out $\sigma_{\tau}$ by adapting the generalized ARP approach to the new environment.

Assume that, before entering the market, each firm must make an irreversible choice on $\alpha_i \in \{\alpha_{l}, \alpha_{h}\}$, with $\alpha_{l} < \alpha_{h}$. Formally,

$$\alpha_i^* = \arg \max_{\alpha_i \in \{\alpha_{l}, \alpha_{h}\}} E[V(\alpha_i)],$$

where $E[V(\alpha_i)]$ stands for the ex-ante expected firm value conditional on the technological choice of $\alpha_i$. In the simple model without capital adjustment costs, $E[V(\alpha_i)]$ solves a static optimization problem according to (37). Dropping the irrelevant time subscript $t$ and using

\(^{23}\)Zhang et al. (2005) estimated returns to education in China’s urban areas. The averaged returns are 10.3 percent in 2001.
(8), (38) and the facts that $I_i = \delta K_i$, we have

$$E[V(\alpha_i)] = \frac{1 + r}{r} \left[ 1 - \frac{\delta (1 - \gamma_i (\alpha_i))}{J} \right] \left[ \frac{1 - \gamma_i (\alpha_i)}{(1 + \tau_i) J} \right]^{\frac{1}{\gamma_i (\alpha_i)} - 1} E[Z_i(\alpha_i)], \quad (29)$$

where

$$\gamma_i (\alpha_i) = 1 - \frac{\alpha_i (1 - \eta_i)}{\eta_i + \alpha_i (1 - \eta_i)},$$

$$E[Z_i(\alpha_i)] = \left[ \frac{\eta_i}{\gamma_i (\alpha_i)} \right]^{\frac{1}{\gamma_i (\alpha_i)}} \left[ (1 - \eta_i)^{1-\alpha_i} \left( \frac{1 - \alpha_i - \beta_i}{m_i} \right)^{1-\alpha_i-\beta_i} \right]^{\frac{1}{m_i} - 1} Z,$$

and $Z$, independent of $\alpha_i$, is an unimportant constant.\(^{24}\) Since $\gamma_i (\alpha_i) > \gamma_i (\alpha_h)$, substituting (29) back into (28) yields

$$\alpha_i^* = \begin{cases} \alpha_i & \text{if } \tau_i > \Pi (\alpha_l, \alpha_h) \\ \alpha_h & \text{if } \tau_i < \Pi (\alpha_l, \alpha_h) \end{cases}, \quad (30)$$

where

$$\Pi (\alpha_l, \alpha_h) = \frac{1}{J} \left( \frac{J - \delta (1 - \gamma_i (\alpha_h)) [1 - \gamma_i (\alpha_h)]^{\frac{1}{\gamma_i (\alpha_h)} - 1} E[Z_i(\alpha_h)]}{J - \delta (1 - \gamma_i (\alpha_l)) [1 - \gamma_i (\alpha_l)]^{\frac{1}{\gamma_i (\alpha_l)} - 1} E[Z_i(\alpha_l)]} \right)^{\frac{\gamma_i (\alpha_h) \gamma_i (\alpha_l)}{\gamma_i (\alpha_l) - \gamma_i (\alpha_h)}} - 1.$$

When $\tau_i = \Pi (\alpha_l, \alpha_h)$ – i.e., firms are indifferent between $\alpha_l$ and $\alpha_h$ – we assume that $\alpha_i$ will be chosen in a random way. (30) shows that a firm would optimally choose the more capital-intensive technology if $\tau_i$ is sufficiently low.

We use the NBS sample to calibrate the model. We maintain the assumptions that $r = 0.20$, $\log \eta_i \sim N (\mu_{\log \eta}, \sigma_{\log \eta}^2)$ and $\log (1 + \tau_i) \sim N (0, \sigma^2_\tau)$. For simplicity, we assume no heterogeneity in $\beta_i$ and $m_i$ such that $\beta_i = \beta$ and $m_i = m$. So, there are seven parameters, $\alpha_l$, $\alpha_h$, $\beta$, $m$, $\mu_{\log \eta}$, $\sigma_{\log \eta}$ and $\sigma_\tau$, left to be estimated.

For any given $\beta$ and $m$, the generalized ARP approach can directly be applied to back out $\alpha_l$, $\alpha_h$, $\mu_{\log \eta}$, $\sigma_{\log \eta}$ and $\sigma_\tau$ by matching the five core moments. The only difference is that $\mu_{\log \alpha}$ and $\sigma_{\log \alpha}$ are now replaced with $\alpha_l$ and $\alpha_h$. We calibrate $\beta$ to match the aggregate labor income share in the NBS data (see footnote 20), which yields a value of $\beta = 0.136$. To pin down $m$, we assume that half of the firms would choose $\alpha_h$ in the absence of distortions (i.e., $\tau_i = 0$ for all firms). As will be shown below, the estimation turns out to be insensitive to the distortionless distribution of $\alpha_i$.

\(^{24}\)Here, we let $\beta_i$ be independent of $\alpha_i$. Alternatively, under the assumption of constant returns to scale, we may let $1 - \alpha_i - \beta_i$ be independent of $\alpha_i$. Then, $\left( (1 - \alpha_i - \beta_i)/m_i \right)^{1-\alpha_i-\beta_i}$ in $E[Z_i(\alpha_i)]$ should be replaced with $(\beta_i/w_i)^{\beta_i}$. Accordingly, the following estimation would involve $1 - \alpha_i - \beta_i$ and $w_i$, rather than $\beta_i$ and $m_i$. The alternation has no effect on the estimation of $\sigma_\tau$. 

24
The benchmark results are reported in the first row of Table 7. Compared with those in Table 4, the technological choice model finds a smaller $\sigma_t$ (0.68 vs. 0.58). The second and third rows report the estimates when $m$ is calibrated to 20 percent and 80 percent of firms choosing $\alpha_h$ in the distortionless environment, respectively. The parameter of key interest, $\sigma_t$, is basically unaffected.

[Insert Table 7]

5.3 Regressions on Firm Characteristics

In the generalized ARP approach, $bsd \left( \log \left( \frac{Y}{\hat{K}} \right) \right)$ is the only empirical moment informative for $\sigma_t$. The other moments are deployed to tease out the effects of heterogeneous $\alpha_i$ and $\eta_i$ on $bsd \left( \log \left( \frac{Y}{\hat{K}} \right) \right)$. In other words, the average revenue-capital ratio, $\frac{1}{T} \sum_{t=1}^{T} \log \left( \frac{Y_{i,t}}{\hat{K}_{i,t}} \right)$, may serve as a proxy for $\tau_i$ once we control for heterogeneities in $\alpha_i$ and $\eta_i$. This motivates a reduced-form regression that allows us to check the correlation between $\tau_i$ and firm characteristics, which is hard to get by structural approach.

Specifically, we run simple regression of $\frac{1}{T} \sum_{t=1}^{T} \log \left( \frac{Y_{i,t}}{\hat{K}_{i,t}} \right)$ on firm characteristics. We add four-digit industry dummies and the average profit-revenue ratio, $\frac{1}{T} \sum_{t=1}^{T} \log \left( \frac{\pi_{i,t}}{Y_{i,t}} \right)$, to the control variables. The idea is that industry dummies control the heterogeneities across industries, while the average profit-revenue ratio takes care of the within-industry heterogeneities. Admittedly, the average profit-revenue ratio alone cannot control both of the within-industry heterogeneities in $\alpha_i$ and $\eta_i$ since it is co-determined by $\alpha_i$ and $\eta_i$. Nevertheless, our simulations suggest that $bsd \left( \log \left( \frac{Y}{\hat{K}} \right) \right)$ is not sensitive to $\sigma_{\log \eta}$ (see Table 1). This provides a justification for omitting the heterogeneity in $\eta_i$ in the regression.

The regression equation is:

$$\frac{1}{T} \sum_{t=1}^{T} \log \left( \frac{Y_{i,t}/\hat{K}_{i,t}}{\hat{K}_{i,t}} \right) = b_0 + b_1 \cdot \frac{1}{T} \sum_{t=1}^{T} \log \left( \frac{\pi_{i,t}}{Y_{i,t}} \right) + b_2 \cdot D_i + b_3 \cdot X_i + \xi_i,$$

(31)

where $D_i$ represents a vector of industry dummies and $X_i$ is a vector of firm characteristics. The parameter of interest is $b_3$, which isolates the channel linking the revenue-capital ratio to firm characteristics via capital market distortions.

Table 8 presents the regression results for NBS firms. The baseline model uses firm age and size as $X_i$. All else being equal, it predicts that the capital goods price of a firm is 3 percent lower if a firm is one year older, and 4 percent lower if a firm has 1000 more employees. This is consistent with the common finding in the large literature on capital market imperfections.\textsuperscript{25}

\textsuperscript{25}For instance, Hadlock and Pierce (2010) examine many commonly-used measures or sorting criteria for the severity of financial constraints and find that firm age and size are the most reliable and useful predictors of financial constraint levels.
The regional disparity of capital returns documented by Brandt et al. (2013) indicates a role of firm location. NBS classifies all 31 provinces into four regions: eastern, central, western and northeastern. We take the eastern region as the reference group. Dummy variables for the central (CENTRAL), western (WESTERN) and northeastern (NORTHEASTERN) regions are added to \( X_i \) in the second regression. Consistent with Brandt et al. (2013), we find significant and large coefficients for the western and northeastern provinces, suggesting that the state government has been heavily subsidizing firms in these regions. The capital goods price for firms in northeastern provinces, for instance, appears to be a quarter lower than that for firms in the eastern provinces.

There is a growing literature on heterogeneous financing costs across firms with different ownership in China. State firms often have much better access to external financing than non-state firms (e.g., Dollar and Wei, 2007; Song et al., 2011; Hsieh and Song, 2014). To check the role of ownership, we take state-owned enterprises (SOE) as the reference group. Dummy variables for collective-owned enterprises (COE), domestic private enterprises (DPE), foreign-invested enterprises (FIE) and other ownership types (OTHERS) are included in the third regression. All the ownership dummies turn out to be positive and highly significant, suggesting lower capital goods prices for state firms. Specifically, the capital goods prices of COE, DPE, FIE and “others” are 50, 45, 10 and 26 percent higher, respectively, than that of SOE.

6 Conclusion

Misallocation has been viewed as a promising candidate for explaining the large TFP differences across countries. To evaluate the importance of misallocation, we need to estimate its magnitude at disaggregate levels. The ARP approach, which has been widely adopted in the literature, relies on the dispersion of revenue productivity to back out misallocation. However, such dispersion may also be generated by unobserved heterogeneities other than the factors that cause misallocation. This paper contributes to the literature in two aspects. The first is methodological. To address the identification issue, we present models that incorporate (i) unobserved heterogeneities in the capital output elasticity and markups; (ii) capital adjustment costs; and (iii) measurement errors. All three factors contribute to the dispersion of the revenue-capital ratio and, hence, bias the estimate upwards. We then develop identification

\[ \text{The category of “OTHERS” includes cooperative units, joint ownership units, limited liability corporations and share-holding corporations Ltd.} \]
conditions that isolate capital misallocation. In particular, we propose the generalized ARP approach to calibrate a simple model and a structural approach to estimate a full-blown model. Secondly, when applying the methods to a firm-level panel dataset from Chinese manufacturing, we find the magnitude of capital misallocation to be quantitatively sizable. In contrast, for large Compustat firms, capital misallocation turns out to be negligible. If capital adjustment costs and measurement errors are found to be modest, the generalized ARP approach would become our favorite due to its tractability and good approximation to the structural estimation.

To be sure, there are potentially other factors that may bias the results upwards. Our estimate can, thus, be taken as a more reasonable upper bound than those from the ARP approach. It is also worth emphasizing that our model does not distinguish the channels through which capital is misallocated. Midrigan and Xu (2014), for instance, find a large effect of credit constraint on misallocation via entry. A future research direction would be to identify the underlying mechanism of misallocation. Moreover, like most others in the literature, our paper accommodates static misallocation only.\(^\text{27}\) It would be interesting in exploring dynamic welfare implications of misallocation.

References


\(^\text{27}\)One exception is Michael Peters (2011).


Figure 1: This figure plots the sensitivity of the standard deviation (sd) and the between-group standard deviation (bsd) of log(Y/Khat) with respect to quadratic capital adjustment cost (Panel A) and measurement error on capital (Panel B), respectively. The benchmark parameterization is set equal to that in Column (5) of Table 1.
Figure 2: Results from Two- and Four-Digit Industries

Panel A: Two-Digit Industries

Panel B: Four-Digit Industries

Figure 2: Panel A plots the between-group standard deviation of $\log(Y/Khat)$ (x-axis) and the calibrated $\sigma_r$ (y-axis) by the generalized ARP approach in each two-digit industry. The solid line is the 45 degree line. Panel B plots the results from each four-digit industry.
Table 1 Illustration for Identification in the Simple Model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_r$ = 0.0</td>
<td>$\sigma_r$ = 0.5</td>
<td>$\sigma_r$ = 0.0</td>
<td>$\sigma_r$ = 0.0</td>
<td>$\sigma_r$ = 0.5</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\log\alpha}$ = 0.0</td>
<td>$\sigma_{\log\alpha}$ = 0.0</td>
<td>$\sigma_{\log\alpha}$ = 0.5</td>
<td>$\sigma_{\log\alpha}$ = 0.0</td>
<td>$\sigma_{\log\alpha}$ = 0.5</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\log\eta}$ = 0.0</td>
<td>$\sigma_{\log\eta}$ = 0.0</td>
<td>$\sigma_{\log\eta}$ = 0.0</td>
<td>$\sigma_{\log\eta}$ = 0.5</td>
<td>$\sigma_{\log\eta}$ = 0.5</td>
<td></td>
</tr>
</tbody>
</table>

**Set of Moments**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean($\pi/Y$)</td>
<td>0.157</td>
<td>0.157</td>
<td>0.167</td>
<td>0.167</td>
<td>0.170</td>
</tr>
<tr>
<td>mean($\log(Y/Khat)$)</td>
<td>0.834</td>
<td>0.834</td>
<td>0.834</td>
<td>0.847</td>
<td>0.840</td>
</tr>
<tr>
<td>bsd($\pi/Y$)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.044</td>
<td>0.044</td>
<td>0.061</td>
</tr>
<tr>
<td>bsd($\log(Y/Khat)$)</td>
<td>0.000</td>
<td>0.496</td>
<td>0.495</td>
<td>0.054</td>
<td>0.682</td>
</tr>
<tr>
<td>bcorr($\pi/Y$, $\log(Y/Khat)$)</td>
<td>N.A.</td>
<td>N.A.</td>
<td>-0.954</td>
<td>0.991</td>
<td>-0.378</td>
</tr>
</tbody>
</table>

**Panel B**

<table>
<thead>
<tr>
<th>Inferred $\sigma_r$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The APR Approach</td>
<td>0.000</td>
<td>0.496</td>
<td>0.495</td>
<td>0.054</td>
<td>0.682</td>
</tr>
<tr>
<td>The Generalized ARP Approach</td>
<td>N.A.</td>
<td>N.A.</td>
<td>0.000</td>
<td>0.000</td>
<td>0.489</td>
</tr>
</tbody>
</table>

**Panel C**

<table>
<thead>
<tr>
<th>Estimates for other parameters</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{\log\alpha}$</td>
<td>N.A.</td>
<td>N.A.</td>
<td>-2.498</td>
<td>-2.499</td>
<td>-2.498</td>
</tr>
<tr>
<td>$\sigma_{\log\alpha}$</td>
<td>N.A.</td>
<td>N.A.</td>
<td>0.505</td>
<td>0.000</td>
<td>0.491</td>
</tr>
<tr>
<td>$\mu_{\log\eta}$</td>
<td>N.A.</td>
<td>N.A.</td>
<td>-2.372</td>
<td>-2.391</td>
<td>-2.493</td>
</tr>
<tr>
<td>$\sigma_{\log\eta}$</td>
<td>N.A.</td>
<td>N.A.</td>
<td>0.098</td>
<td>0.423</td>
<td>0.490</td>
</tr>
</tbody>
</table>

Note: The imposed parameter values are $\delta = 0.05$, $r = 0.15$, $\mu_{\log\alpha} = \mu_{\log\eta} = -2.50$, $\sigma_{\log\alpha} = \sigma_{\log\eta} = 0.50$, $\rho = 0.90$, $\mu = 0.05$, $\sigma = 0.40$, and $b^g = b^i = b^f = \sigma_{mcK} = \sigma_{mcY} = \sigma_{mcX} = 0$. 
Table 2 Illustration for Identification in the Full-Blown Model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(b^q = 1)</td>
<td>(b^q = 10)</td>
<td>(\sigma_{mck} = 0.5)</td>
<td>(\sigma_{mck} = 1)</td>
<td>(b^q = 1, \sigma_{mck} = 0.5)</td>
</tr>
<tr>
<td><strong>Set of Moments</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean((\pi/Y))</td>
<td>0.170</td>
<td>0.170</td>
<td>0.170</td>
<td>0.170</td>
<td>0.170</td>
</tr>
<tr>
<td>mean(log(Y/Khat))</td>
<td>0.865</td>
<td>1.209</td>
<td>0.840</td>
<td>0.765</td>
<td>0.854</td>
</tr>
<tr>
<td>bsd((\pi/Y))</td>
<td>0.061</td>
<td>0.061</td>
<td>0.061</td>
<td>0.061</td>
<td>0.061</td>
</tr>
<tr>
<td>bsd(log(Y/Khat))</td>
<td>0.676</td>
<td>0.716</td>
<td>0.733</td>
<td>0.835</td>
<td>0.711</td>
</tr>
<tr>
<td>bcorr((\pi/Y, \log(Y/Khat)))</td>
<td>-0.392</td>
<td>-0.350</td>
<td>-0.350</td>
<td>-0.304</td>
<td>-0.371</td>
</tr>
</tbody>
</table>

| **Panel B**         |     |     |     |     |     |
| Inferred \(\sigma_r\) |     |     |     |     |     |
| The APR Approach    | 0.688 | 0.727 | 0.872 | 1.188 | 0.822 |
| The Generalized ARP Approach | 0.469 | 0.432 | 0.561 | 0.708 | 0.523 |

| **Panel C**         |     |     |     |     |     |
| Estimates for other parameters |     |     |     |     |     |
| \(\mu_{\log\alpha}\) | -2.521 | -2.840 | -2.497 | -2.429 | -2.511 |
| \(\sigma_{\log\alpha}\) | 0.503 | 0.585 | 0.490 | 0.470 | 0.499 |
| \(\mu_{\log\eta}\) | -2.467 | -2.210 | -2.493 | -2.582 | -2.478 |
| \(\sigma_{\log\eta}\) | 0.478 | 0.402 | 0.491 | 0.524 | 0.483 |

Note: The imposed parameters are \(\delta = 0.05\), \(r = 0.15\), \(\mu_{\log\alpha} = -2.50\), \(\sigma_{\log\alpha} = \sigma_{\log\eta} = 0.50\), \(\rho = 0.90\), \(\mu = 0.05\) and \(\sigma = 0.40\).

Column (1) and (2): \(b^i = b^f = \sigma_{mck} = \sigma_{mey} = \sigma_{mex} = 0\)

Column (3) and (4): \(b^i = b^f = \sigma_{mey} = \sigma_{mex} = 0\)

Column (5): \(b^i = b^f = \sigma_{mey} = \sigma_{mex} = 0\)
### Table 3: Parameters and Moments in the Structural Estimation

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_z$</td>
<td>standard deviation of heterogeneities in capital goods price</td>
</tr>
<tr>
<td>$\mu_{\log a}$</td>
<td>mean of log capital output elasticity in production function</td>
</tr>
<tr>
<td>$\sigma_{\log a}$</td>
<td>standard deviation of log capital output elasticity</td>
</tr>
<tr>
<td>$\mu_{\log \eta}$</td>
<td>mean of log inverse of demand elasticity</td>
</tr>
<tr>
<td>$\sigma_{\log \eta}$</td>
<td>standard deviation of log inverse of demand elasticity</td>
</tr>
<tr>
<td>$b^q$</td>
<td>quadratic adjustment costs</td>
</tr>
<tr>
<td>$b^i$</td>
<td>partial irreversibility</td>
</tr>
<tr>
<td>$b^f$</td>
<td>fixed adjustment costs</td>
</tr>
<tr>
<td>$\mu$</td>
<td>mean of growth rate in $Z_{it}$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>standard deviation of shocks to $Z_{it}$</td>
</tr>
<tr>
<td>$\sigma_{meK}$</td>
<td>standard deviation of measurement errors in capital stock</td>
</tr>
<tr>
<td>$\sigma_{meY}$</td>
<td>standard deviation of measurement errors in sales revenue</td>
</tr>
<tr>
<td>$\sigma_{me\pi}$</td>
<td>standard deviation of measurement errors in gross profit</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Moments</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean($\pi/Y$)</td>
<td>mean of profit-revenue ratio</td>
</tr>
<tr>
<td>mean(log($Y/Khat$))</td>
<td>mean of log revenue-capital ratio</td>
</tr>
<tr>
<td>mean($I/K$)</td>
<td>mean of investment rate</td>
</tr>
<tr>
<td>mean($\Delta \log Y$)</td>
<td>mean of revenue growth rate</td>
</tr>
<tr>
<td>bsd($\pi/Y$)</td>
<td>between-group standard deviation of profit-revenue ratio</td>
</tr>
<tr>
<td>wsd($\pi/Y$)</td>
<td>within-group standard deviation of profit-revenue ratio</td>
</tr>
<tr>
<td>bsd(log($Y/Khat$))</td>
<td>between-group standard deviation of log revenue-capital ratio</td>
</tr>
<tr>
<td>wsd(log($Y/Khat$))</td>
<td>within-group standard deviation of log revenue-capital ratio</td>
</tr>
<tr>
<td>bsd($I/K$)</td>
<td>between-group standard deviation of investment rate</td>
</tr>
<tr>
<td>wsd($I/K$)</td>
<td>within-group standard deviation of investment rate</td>
</tr>
<tr>
<td>bsd($\Delta \log Y$)</td>
<td>between-group standard deviation of revenue growth rate</td>
</tr>
<tr>
<td>wsd($\Delta \log Y$)</td>
<td>within-group standard deviation of revenue growth rate</td>
</tr>
<tr>
<td>skew($\pi/Y$)</td>
<td>skewness of profit-revenue ratio</td>
</tr>
<tr>
<td>skew(log($Y/Khat$))</td>
<td>skewness of log revenue-capital ratio</td>
</tr>
<tr>
<td>skew($I/K$)</td>
<td>skewness of investment rate</td>
</tr>
<tr>
<td>skew($\Delta \log Y$)</td>
<td>skewness of revenue growth rate</td>
</tr>
<tr>
<td>scorr($\pi/Y$)</td>
<td>serial correlation of profit-revenue ratio</td>
</tr>
<tr>
<td>scorr(log($Y/Khat$))</td>
<td>serial correlation of log revenue-capital ratio</td>
</tr>
<tr>
<td>scorr($I/K$)</td>
<td>serial correlation of investment rate</td>
</tr>
<tr>
<td>scorr($\Delta \log Y$)</td>
<td>serial correlation of revenue growth rate</td>
</tr>
<tr>
<td>bcorr($\pi/Y, \ log(Y/Khat)$)</td>
<td>cross correlation between between-group profit-revenue ratio and log revenue-capital ratio</td>
</tr>
</tbody>
</table>
Table 4: Results by the Generalized ARP Approach

<table>
<thead>
<tr>
<th>Parameters</th>
<th>NBS</th>
<th>Compustat I</th>
<th>Compustat II</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_r$</td>
<td>0.6843 (0.7143)</td>
<td>0.3095 (0.3020)</td>
<td>0.1313 (0.1113)</td>
</tr>
<tr>
<td>$\mu_{loga}$</td>
<td>-2.6189 (-2.6058)</td>
<td>-1.9320 (-1.8745)</td>
<td>-1.9788 (-2.1797)</td>
</tr>
<tr>
<td>$\sigma_{loga}$</td>
<td>0.5539 (0.5568)</td>
<td>0.6259 (0.6271)</td>
<td>0.5643 (0.5338)</td>
</tr>
<tr>
<td>$\mu_{logq}$</td>
<td>-2.8086 (-2.8084)</td>
<td>-1.4527 (-1.5831)</td>
<td>-1.6241 (-1.6313)</td>
</tr>
<tr>
<td>$\sigma_{logq}$</td>
<td>0.7887 (0.7253)</td>
<td>0.5246 (0.3949)</td>
<td>0.5587 (0.5235)</td>
</tr>
</tbody>
</table>

Moments

<table>
<thead>
<tr>
<th></th>
<th>NBS</th>
<th>Compustat I</th>
<th>Compustat II</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean($\pi/Y$)</td>
<td>0.1578</td>
<td>0.3929</td>
<td>0.3514</td>
</tr>
<tr>
<td>mean(log($Y/Khat$))</td>
<td>1.1377</td>
<td>0.5659</td>
<td>0.5542</td>
</tr>
<tr>
<td>bsd($\pi/Y$)</td>
<td>0.0763</td>
<td>0.1591</td>
<td>0.1448</td>
</tr>
<tr>
<td>bsd(log($Y/Khat$))</td>
<td>0.8666</td>
<td>0.7288</td>
<td>0.6064</td>
</tr>
<tr>
<td>bcorr($\pi/Y$, log($Y/Khat$))</td>
<td>-0.2422</td>
<td>-0.0738</td>
<td>-0.0879</td>
</tr>
</tbody>
</table>

Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>NBS</th>
<th>Compustat I</th>
<th>Compustat II</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of firms</td>
<td>107579</td>
<td>1431</td>
<td>970</td>
</tr>
<tr>
<td>mean(sales)</td>
<td>97</td>
<td>2066</td>
<td>3132</td>
</tr>
<tr>
<td>med(sales)</td>
<td>20</td>
<td>121</td>
<td>461</td>
</tr>
<tr>
<td>mean(employees)</td>
<td>0.33</td>
<td>10.4</td>
<td>20.8</td>
</tr>
<tr>
<td>med(employees)</td>
<td>0.13</td>
<td>0.8</td>
<td>4.3</td>
</tr>
</tbody>
</table>

Note: The numbers in parentheses are the results of structural estimation of the full-blown model (see Table 5 for more details). The imposed parameters are $\delta = 0.05$ (0.10), $r = 0.20$ (0.10) for Chinese (Compustat) firms, respectively. Unit of sales is millions of RMB (USD) in 2004 prices for Chinese (Compustat) firms. Unit of employees is thousand.
Table 5: Structural Estimation Results for Chinese Firms

<table>
<thead>
<tr>
<th>Parameters</th>
<th>estimate</th>
<th>s.e.</th>
<th>Moments</th>
<th>empirical</th>
<th>s.e.</th>
<th>simulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_z$</td>
<td>0.7143</td>
<td>0.0033</td>
<td>mean($\pi/Y$)</td>
<td>0.1578</td>
<td>0.0002</td>
<td>0.1542</td>
</tr>
<tr>
<td>$\mu_{\log\alpha}$</td>
<td>-2.6058</td>
<td>0.0019</td>
<td>mean(log($Y/Khat$))</td>
<td>1.1377</td>
<td>0.0025</td>
<td>1.1456</td>
</tr>
<tr>
<td>$\sigma_{\log\alpha}$</td>
<td>0.5568</td>
<td>0.0043</td>
<td>mean(I/K)</td>
<td>0.1640</td>
<td>0.0005</td>
<td>0.1729</td>
</tr>
<tr>
<td>$\mu_{\log\eta}$</td>
<td>-2.8084</td>
<td>0.0051</td>
<td>mean($\Delta\log Y$)</td>
<td>0.0963</td>
<td>0.0005</td>
<td>0.0803</td>
</tr>
<tr>
<td>$\sigma_{\log\eta}$</td>
<td>0.7253</td>
<td>0.0061</td>
<td>bsd($\pi/Y$)</td>
<td>0.0763</td>
<td>0.0001</td>
<td>0.0745</td>
</tr>
<tr>
<td>$b^f$</td>
<td>0.2777</td>
<td>0.0038</td>
<td>wsd($\pi/Y$)</td>
<td>0.0506</td>
<td>0.0001</td>
<td>0.0488</td>
</tr>
<tr>
<td>$b^i$</td>
<td>0.0001</td>
<td>0.0395</td>
<td>bsd(log($Y/Khat$))</td>
<td>0.8666</td>
<td>0.0011</td>
<td>0.8781</td>
</tr>
<tr>
<td>$b^f$</td>
<td>0.0335</td>
<td>0.0006</td>
<td>wsd(log($Y/Khat$))</td>
<td>0.3470</td>
<td>0.0009</td>
<td>0.3321</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.0802</td>
<td>0.0004</td>
<td>bsd(I/K)</td>
<td>0.1991</td>
<td>0.0006</td>
<td>0.1642</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.4253</td>
<td>0.0016</td>
<td>wsd(I/K)</td>
<td>0.2027</td>
<td>0.0006</td>
<td>0.2149</td>
</tr>
<tr>
<td>$\sigma_{meK}$</td>
<td>0.4010</td>
<td>0.0013</td>
<td>bsd($\Delta\log Y$)</td>
<td>0.1876</td>
<td>0.0004</td>
<td>0.1632</td>
</tr>
<tr>
<td>$\sigma_{meY}$</td>
<td>0.0007</td>
<td>0.1255</td>
<td>wsd($\Delta\log Y$)</td>
<td>0.2042</td>
<td>0.0004</td>
<td>0.2187</td>
</tr>
<tr>
<td>$\sigma_{mez}$</td>
<td>0.5777</td>
<td>0.0020</td>
<td>skew($\pi/Y$)</td>
<td>0.7760</td>
<td>0.0039</td>
<td>0.8539</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>skew(log($Y/Khat$))</td>
<td>0.0570</td>
<td>0.0038</td>
<td>0.0037</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>skew(I/K)</td>
<td>2.2307</td>
<td>0.0075</td>
<td>2.2510</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>skew(dlogY)</td>
<td>0.1567</td>
<td>0.0037</td>
<td>0.1760</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>scorr($\pi/Y$)</td>
<td>0.5703</td>
<td>0.0021</td>
<td>0.5993</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>scorr(log($Y/Khat$))</td>
<td>0.8403</td>
<td>0.0009</td>
<td>0.8377</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>scorr(I/K)</td>
<td>0.1188</td>
<td>0.0030</td>
<td>0.2430</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>scorr($\Delta\log Y$)</td>
<td>0.0685</td>
<td>0.0028</td>
<td>0.0526</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>bcorr($\pi/Y$, log($Y/Khat$))</td>
<td>-0.2422</td>
<td>0.0034</td>
<td>-0.2707</td>
</tr>
<tr>
<td>OI/100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>183</td>
</tr>
</tbody>
</table>

Note: The imposed parameter values are $\delta = 0.05$, $r = 0.20$, and $\rho = 0.90$. 
Table 6: Structural Estimation Results for Compustat Firms

<table>
<thead>
<tr>
<th>Sample</th>
<th>Compustat I</th>
<th>Compustat II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimate</td>
<td>s.e.</td>
</tr>
<tr>
<td>$\sigma_t$</td>
<td>0.3020</td>
<td>0.0676</td>
</tr>
<tr>
<td>$\mu_{\log \alpha}$</td>
<td>-1.8745</td>
<td>0.0193</td>
</tr>
<tr>
<td>$\sigma_{\log \alpha}$</td>
<td>0.6271</td>
<td>0.0272</td>
</tr>
<tr>
<td>$\mu_{\log \eta}$</td>
<td>-1.5831</td>
<td>0.0305</td>
</tr>
<tr>
<td>$\sigma_{\log \eta}$</td>
<td>0.5949</td>
<td>0.0146</td>
</tr>
<tr>
<td>$b^q$</td>
<td>1.1355</td>
<td>0.1489</td>
</tr>
<tr>
<td>$b^i$</td>
<td>0.0073</td>
<td>0.1836</td>
</tr>
<tr>
<td>$b^f$</td>
<td>0.0017</td>
<td>0.0117</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.0263</td>
<td>0.0017</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.2636</td>
<td>0.0096</td>
</tr>
<tr>
<td>$\sigma_{\text{me}K}$</td>
<td>0.2014</td>
<td>0.0117</td>
</tr>
<tr>
<td>$\sigma_{\text{me}Y}$</td>
<td>0.0023</td>
<td>0.2542</td>
</tr>
<tr>
<td>$\sigma_{\text{me}H}$</td>
<td>0.2286</td>
<td>0.0112</td>
</tr>
</tbody>
</table>

Note: See the text for the definition of Compustat I and II. The imposed parameter values are $\delta = 0.10$, $r = 0.10$, $\rho = 0.90$. 
Table 7: Estimation with Endogenous Capital Output Elasticity

<table>
<thead>
<tr>
<th></th>
<th>$\sigma_t$</th>
<th>$\mu_{\log y}$</th>
<th>$\sigma_{\log y}$</th>
<th>$\alpha_l$</th>
<th>$\alpha_h$</th>
<th>$\beta$</th>
<th>$m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.5784</td>
<td>-2.7460</td>
<td>0.7841</td>
<td>0.0551</td>
<td>0.1090</td>
<td>0.1369</td>
<td>1.2024</td>
</tr>
<tr>
<td>(2)</td>
<td>0.5800</td>
<td>-2.7456</td>
<td>0.7838</td>
<td>0.0555</td>
<td>0.1097</td>
<td>0.1369</td>
<td>1.1769</td>
</tr>
<tr>
<td>(3)</td>
<td>0.5887</td>
<td>-2.7434</td>
<td>0.7817</td>
<td>0.0576</td>
<td>0.1142</td>
<td>0.1369</td>
<td>1.0396</td>
</tr>
</tbody>
</table>

Note: $m$ and $\beta$ denote the material cost relative to the unit price of capital and labor output elasticity, respectively. (1) is the benchmark case where 50% of firms would choose $\alpha_h$ if $\tau_i = 0$. (2) and (3) refer to the cases where 20% and 80% of firms would choose $\alpha_h$ if $\tau_i = 0$, respectively.
Table 8: Regressions on Firm Characteristics

<table>
<thead>
<tr>
<th>Dep. Variable</th>
<th>The average log revenue-capital ratio</th>
<th>(1) age and size</th>
<th>(2) region</th>
<th>(3) ownership</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>log(\pi/Y)</strong></td>
<td></td>
<td>-0.3929</td>
<td>-0.3921</td>
<td>-0.3590</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0027)</td>
<td>(0.0027)</td>
<td>(0.0026)</td>
</tr>
<tr>
<td>age</td>
<td></td>
<td>-0.0304</td>
<td>-0.0303</td>
<td>-0.0291</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>emp</td>
<td></td>
<td>-0.0399</td>
<td>-0.0394</td>
<td>-0.0244</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0032)</td>
<td>(0.0032)</td>
<td>(0.0021)</td>
</tr>
<tr>
<td>CENTRAL</td>
<td></td>
<td></td>
<td>0.0610</td>
<td>0.0319</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0040)</td>
<td>(0.0039)</td>
<td></td>
</tr>
<tr>
<td>WESTERN</td>
<td></td>
<td>-0.1673</td>
<td>-0.1721</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0045)</td>
<td>(0.0044)</td>
<td></td>
</tr>
<tr>
<td>NORTHEASTERN</td>
<td></td>
<td>-0.2495</td>
<td>-0.2551</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0056)</td>
<td>(0.0055)</td>
<td></td>
</tr>
<tr>
<td>COE</td>
<td></td>
<td></td>
<td></td>
<td>0.4973</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0075)</td>
<td></td>
</tr>
<tr>
<td>DPE</td>
<td></td>
<td></td>
<td></td>
<td>0.4481</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0064)</td>
<td></td>
</tr>
<tr>
<td>FIE</td>
<td></td>
<td></td>
<td></td>
<td>0.0986</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0067)</td>
<td></td>
</tr>
<tr>
<td>OTHERs</td>
<td></td>
<td></td>
<td></td>
<td>0.2614</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0063)</td>
<td></td>
</tr>
<tr>
<td>Obs.</td>
<td></td>
<td>107579</td>
<td>107579</td>
<td>107579</td>
</tr>
<tr>
<td>R-sq</td>
<td></td>
<td>0.2420</td>
<td>0.2493</td>
<td>0.2806</td>
</tr>
</tbody>
</table>

Note: 1. The parentheses report robust standard errors.
2. 4-digit industry dummies are included in all regressions.
3. Age is the difference between 2004 and the year of firm foundation.
4. Emp is the number of total employees normalized by 1000.
5. Baseline group is EASTERN--dummy = 1 if province is Beijing, Tianjin, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, or Hainan.
   CENTRAL--dummy = 1 if province is Shanxi, Anhui, Jiangxi, Henan, Hubei or Hunan.
   WESTERN--dummy =1 if province is Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia or Xinjiang.
   NORTHWESTERN-- dummy = 1 if province is Liaoning, Jilin or Heilongjiang.
6. Baseline group is SOE--dummy = 1 if state-owned; defined as registration type = 110, 141 and 151.
   COE--dummy = 1 if collective owned; defined as registration type = 120 and 142.
   DPE--dummy = 1 if domestic private-owned; defined as registration type from 170 to 174.
   FIE--dummy = 1 if foreign-owned; defined as registration type from 200 to 240 or from 300 to 340.
   OTHERS--dummy = 1 if other types, including cooperative units, joint ownership units, limited liability corporations and share-holding corporation Ltd.