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 adjustment costs and measurement errors.[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.[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.[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

[2] See Cooper et al. (2010) for a structural estimation of labor adjustment costs in Chinese manufacturing.

[3] See, e.g., Dollar and Wei (2007), Brandt et al. (2013), Hsieh and Song (2014).

[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 $\hat{K}_{i,t}$, $L_{i,t}$ 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:

$$Q_{i,t} = A_{i,t}\hat{K}_{i,t}^{\alpha_i} L_{i,t}^{\beta_i} M_{i,t}^{1-\alpha_i-\beta_i}, \tag{1}$$

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,

$$Q_{i,t} = X_{i,t} P_{i,t}^{-\frac{1}{\eta_i}}. \tag{2}$$

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 η_i as product market distortions. But our estimation of capital market distortions is independent of the interpretation.

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

$$Y_{i,t} = X_{i,t}^{\eta_i} A_{i,t}^{1-\eta_i} \left(\hat{K}_{i,t}^{\alpha_i} L_{i,t}^{\beta_i} M_{i,t}^{1-\alpha_i-\beta_i} \right)^{1-\eta_i}. \tag{3}$$

Denote $w_{i,t}$ and $m_{i,t}$ as wage rate and intermediate input price, respectively. For a given productive capital stock $\hat{K}_{i,t}$, firm i chooses $L_{i,t}$ and $M_{i,t}$ to maximize its gross profits, denoted by $\pi_{i,t}$:

$$\pi_{i,t} = \max_{L_{i,t}, M_{i,t}} \{Y_{i,t} - w_{i,t}L_{i,t} - m_{i,t}M_{i,t}\}, \tag{4}$$

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:

$$\begin{aligned} \frac{w_{i,t}L_{i,t}}{Y_{i,t}} &= \beta_i(1-\eta_i), & (5)\\ \frac{m_{i,t}M_{i,t}}{Y_{i,t}} &= (1-\alpha_i-\beta_i)(1-\eta_i), & (6)\\ \frac{\pi_{i,t}}{Y_{i,t}} &= \eta_i + \alpha_i(1-\eta_i). & (7) \end{aligned}$$

The labor and intermediate input shares would reduce to β_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 α_i as profits are just capital income. (7) will be a key equation for identifying the heterogeneities of α_i and η_i.

The optimization also establishes a profit function:

$$\pi_{i,t} = Z_{i,t}^{\gamma_i}\hat{K}_{i,t}^{1-\gamma_i}, \tag{8}$$

where

$$\gamma_i \equiv 1 - \frac{\alpha_i(1-\eta_i)}{\eta_i + \alpha_i(1-\eta_i)}, \tag{9}$$

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), $\partial Y_{i,t}/\partial \hat{K}_{i,t}$, is identical to $\partial \pi_{i,t}/\partial \hat{K}_{i,t}$, we have the following representation of MRPK:

$$MRPK_{i,t} = \alpha_i(1-\eta_i)\frac{Y_{i,t}}{\hat{K}_{i,t}} = (1-\gamma_i)\left(\frac{Z_{i,t}}{\hat{K}_{i,t}}\right)^{\gamma_i}. \tag{10}$$

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] $Z_{i,t} \equiv \left(\frac{\eta_i}{\gamma_i}\right)^{\frac{1}{\gamma_i}}\left[(1-\eta_i)^{1-\alpha_i}\left(\frac{\beta_i}{w_{i,t}}\right)^{\beta_i}\left(\frac{1-\alpha_i-\beta_i}{m_{i,t}}\right)^{1-\alpha_i-\beta_i}\right]^{\frac{1}{\eta_i}-1} X_{i,t}A_{i,t}^{\frac{1}{\eta_i}-1}$.

Use τ_i to summarize the effects of various capital market distortions on the capital goods price that firm i faces:

$$P_{i,t}^K = (1 + \tau_i)\, P_t^K, \tag{11}$$

where P_t^K denotes the average capital goods price. A positive value of τ_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 τ_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. $\hat{K}_{i,t}$ depreciates at the end of that period. The law of motion for capital is given by

$$\begin{aligned} K_{i,t+1} &= (1-\delta)\hat{K}_{i,t} \\ &= (1-\delta)\left(K_{i,t} + I_{i,t}\right), \end{aligned} \tag{12}$$

where δ 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:

$$J_t \equiv P_t^K - \frac{1-\delta}{1+r} E_t\left[P_{t+1}^K\right]. \tag{13}$$

Appendix 7.2 shows that

$$\hat{K}_{i,t} = \left[\frac{1-\gamma_i}{(1+\tau_i)\, J_t}\right]^{\frac{1}{\gamma_i}} Z_{i,t}. \tag{14}$$

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

$$MRPK_{i,t} = \alpha_i \left(1-\eta_i\right) \frac{Y_{i,t}}{\hat{K}_{i,t}} = (1+\tau_i)\, J_t. \tag{15}$$

[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 σ_τ the standard deviation of $\log(1+\tau_i)$ across firms. Appendix 7.1 shows that σ_τ is a summary statistics of the magnitude of capital misallocation: The aggregate output gain of removing capital misallocation is proportional to σ_τ^2. We will, thus, focus on the identification and estimation of σ_τ, 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 τ_i, where $\log(1+\tau_i)$ is drawn independently from an identical normal distribution with mean zero and standard deviation σ_τ:

$$\log(1+\tau_i) \overset{i.i.d}{\sim} N\left(0, \sigma_\tau^2\right). \tag{16}$$

We also allow α_i and η_i to be firm-specific. They are drawn independently from the following distributions:

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

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

α_i and η_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, α_i and η_i are exogenous. We will relax this assumption in Section 5.2, which allows τ_i to affect α_i.

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

$$\begin{aligned} \log Z_{i,t} &= \mu t + z_{i,t}, \\ z_{i,t} &= \rho z_{i,t-1} + e_{i,t}, \end{aligned} \tag{19}$$

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, σ, is the parameter characterizing the level of uncertainty. We assume homogeneous μ and σ in the

[7] This is because of the timing assumption on $\hat{K}_{i,t}$. It is reassuring that the average revenue-capital ratio, a key variable for estimating σ_τ, has very similar empirical distributions in our samples regardless of whether the denominator is $\hat{K}_{i,t}$ 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, ρ, 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):

$$\log\left(\frac{Y_{i,t}}{\hat{K}_{i,t}}\right)=\log J_t+\log\left(1+\tau_i\right)-\log\left[\alpha_i\left(1-\eta_i\right)\right]. \tag{20}$$

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

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 σ_τ 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 α_i and η_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 σ_τ in the simple model is how to deal with heterogeneities in α_i and η_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 $\log\left(Y_{i,t}/\hat{K}_{i,t}\right)$ and $\pi_{i,t}/Y_{i,t}$, while there are three parameters governing unobserved heterogeneities: σ_τ, $\sigma_{\log\alpha}$ and $\sigma_{\log\eta}$. To resolve the under-identification issue, we introduce the cross correlation between $\log\left(Y_{i,t}/\hat{K}_{i,t}\right)$ and $\pi_{i,t}/Y_{i,t}$, which follows:

$$corr\left[\frac{\pi_{i,t}}{Y_{i,t}},\log\left(\frac{Y_{i,t}}{\hat{K}_{i,t}}\right)\right]\left\{\begin{array}{l}<0,\text{ if }\sigma_{\log\alpha}>0\text{ and }\sigma_{\log\eta}=0\\ \\ >0,\text{ if }\sigma_{\log\alpha}=0\text{ and }\sigma_{\log\eta}>0\end{array}\right.. \tag{21}$$

Intuitively, higher markups (η_i) increase both the profit-revenue and revenue-capital ratios, while a larger α_i increases the profit-revenue ratio but decreases the revenue-capital ratio. In extreme cases, if there is no heterogeneity in η_i (α_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): σ_τ, $\mu_{\log\alpha}$, $\mu_{\log\eta}$, $\sigma_{\log\alpha}$ and $\sigma_{\log\eta}$. The five moments are means of $\pi_{i,t}/Y_{i,t}$ and $\log\left(Y_{i,t}/\hat{K}_{i,t}\right)$, between-group standard deviations of $\pi_{i,t}/Y_{i,t}$ and $\log\left(Y_{i,t}/\hat{K}_{i,t}\right)$, and the between-group correlation of $\pi_{i,t}/Y_{i,t}$ and $\log\left(Y_{i,t}/\hat{K}_{i,t}\right)$, denoted as $mean\,(\pi/Y)$, $mean\left(\log\left(Y/\hat{K}\right)\right)$, $bsd\,(\pi/Y)$, $bsd\left(\log\left(Y/\hat{K}\right)\right)$ and $bcorr\left(\pi/Y,\ \log\left(Y/\hat{K}\right)\right)$, respectively.[9] This constitutes five equations with five unknown variables:

$$\begin{bmatrix} mean\left(\frac{\pi}{Y}\right) \\ mean\left(\log\left(\frac{Y}{\hat{K}}\right)\right) \\ bsd\left(\frac{\pi}{Y}\right) \\ bsd\left(\log\left(\frac{Y}{\hat{K}}\right)\right) \\ bcorr\left(\frac{\pi}{Y},\ \log\left(\frac{Y}{\hat{K}}\right)\right) \end{bmatrix} = \begin{bmatrix} E\left[\eta_i\right] + E\left[\alpha_i\left(1-\eta_i\right)\right] \\ \log J - E\left[\log\alpha_i\right] - E\left[\log\left(1-\eta_i\right)\right] \\ \sqrt{var\left[\eta_i + \alpha_i\left(1-\eta_i\right)\right]} \\ \sqrt{\sigma_\tau^2 + var\left[\log\alpha_i\left(1-\eta_i\right)\right]} \\ corr\left[\eta_i + \alpha_i\left(1-\eta_i\right), \log\left(1+\tau_i\right) - \log\left[\alpha_i\left(1-\eta_i\right)\right]\right] \end{bmatrix}. \quad (22)$$

The first and third equations are about $mean\,(\pi/Y)$ and $bsd\,(\pi/Y)$, based on (7). The second and fourth equations are about $mean\left(\log\left(Y/\hat{K}\right)\right)$ and $bsd\left(\log\left(Y/\hat{K}\right)\right)$, based on (20). The last equation is about $bcorr\left(\pi/Y,\ \log\left(Y/\hat{K}\right)\right)$, 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_t^K, 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: $bcorr\left(\pi/Y, \log\left(Y/\hat{K}\right)\right) \equiv corr\left[1/T \cdot \sum_{t=1}^{T} \pi_{i,t}/Y_{i,t}, 1/T \cdot \sum_{t=1}^{T} \log\left(Y_{i,t}/\hat{K}_{i,t}\right)\right]$

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 σ_τ, $\sigma_{\log\alpha}$ and $\sigma_{\log\eta}$, respectively, in Column (2) to (4). Column (2) shows that only $bsd\left(\log\left(Y/\hat{K}\right)\right)$ responds to σ_τ. In Column (3), $\sigma_{\log\alpha} > 0$ increases both $bsd\,(\pi/Y)$ and $bsd\left(\log\left(Y/\hat{K}\right)\right)$. As predicted by (21), Column (3) and (4) show that $\sigma_{\log\alpha} > 0$ and $\sigma_{\log\eta} > 0$ lead to negative and positive $bcorr\left(\pi/Y, \log\left(Y/\hat{K}\right)\right)$, respectively. The effect of $\sigma_{\log\eta}$ on $bsd\left(\log\left(Y/\hat{K}\right)\right)$ 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 σ_τ, $\sigma_{\log\alpha}$ and $\sigma_{\log\eta}$ illustrate how these heterogeneties can be identified by the generalized ARP approach.

[Insert Table 1]

We then apply the ARP and generalized ARP approaches to the simulated data. Panel B of Table 1 presents the inferred values of σ_τ. 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 α_i and η_i. There are other factors that are missing in the simple model but may potentially contaminate the inference of σ_τ. 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:

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

where $G(K_{i,t}; I_{i,t})$ represents the function of capital adjustment costs, with $\mathbf{1}_{[I_t<0]}$ and $\mathbf{1}_{[I_t\neq 0]}$ being indicators for negative and non-zero investment; b^q measures the magnitude of quadratic adjustment costs; b^i can be interpreted as the difference between the purchase price, $P^K_{i,t}$, 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 τ_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 σ_τ.[11]

3.2 Measurement Errors

We next introduce measurement errors. The benchmark specification assumes that

$$K_{i,t} = K^{true}_{i,t}\exp(e^K_{i,t}), \quad e^K_{i,t} \overset{i.i.d}{\sim} N(0, \sigma^2_{meK}), \tag{23}$$

$$Y_{i,t} = Y^{true}_{i,t}\exp(e^Y_{i,t}), \quad e^Y_{i,t} \overset{i.i.d}{\sim} N(0, \sigma^2_{meY}), \tag{24}$$

$$\pi_{i,t} = \pi^{true}_{i,t}(1 + e^\pi_{i,t}), \quad e^\pi_{i,t} \overset{i.i.d}{\sim} U[-\sigma_{me\pi}, \sigma_{me\pi}]. \tag{25}$$

Here, variables with and without the "true" superscript denote the true states and their observed counterparts in the data, respectively. $e^K_{i,t}$ and $e^Y_{i,t}$ are measurement errors in capital and revenue, respectively. They are drawn independently from an identical normal distribution with mean zero and standard deviation σ_{meK} and σ_{meY}, respectively. $e^\pi_{i,t}$ 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 σ_τ. 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 σ_τ 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 $bsd\left(\log\left(Y/\hat{K}\right)\right)$ and $sd\left(\log\left(Y/\hat{K}\right)\right)$ with respect to b^q. On the one hand, both $bsd\left(\log\left(Y/\hat{K}\right)\right)$ and $sd\left(\log\left(Y/\hat{K}\right)\right)$ 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 σ_τ 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 σ_τ.[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.

[Insert Figure 1 and Table 2]

On the other hand, the standard deviations of the revenue-capital ratio start to increase in b^q when b^q is sufficiently large. It illustrates a caveat of applying the generalized ARP approach to an economy with large capital adjustment costs. One can find in Column (2) of Table 2 that the inferred value of would be 14 percent lower than its true value if $b^q = 10$, a high-end estimate of bq in the literature. This echos the issue raised by Asker et al. (2014), who show that large capital adjustment costs can be an important contributing factor to the observed misallocation. The potential bias motivates a more sophisticated structural estimation that will be introduced below.

The standard deviations with respect to b^i and b^f turn out to be flat even when both b^i and b^f are increased to 0.1, more than three times larger than the maximum values estimated

[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 disperson of the revenue-capital ratio and, hence, infer a lower value of σ_τ.

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 $bsd\left(\log\left(Y/\hat{K}\right)\right)$ and $sd\left(\log\left(Y/\hat{K}\right)\right)$ with respect to σ_{meK}. Not surprisingly, measurement errors on capital can easily blow up $sd\left(\log\left(Y/\hat{K}\right)\right)$ by increasing the observed volatility of the revenue-capital ratio. In contrast, $bsd\left(\log\left(Y/\hat{K}\right)\right)$ remains largely flat for σ_{meK} below 0.5.[13] σ_{meK} starts to have a significant effect on $bsd\left(\log\left(Y/\hat{K}\right)\right)$ for σ_{meK} above 0.5. The inferred values of σ_τ are reported in Column (3) to (4) of Table 2. The generalized ARP approach overestimates σ_τ 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 $sd\left(\log\left(Y/\hat{K}\right)\right)$. 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^q and σ_{meK} bias σ_τ in opposite direction, the inferred value of σ_τ is actually closer to the true value than its counterparts with b^q or σ_{meK} only.

The effect of σ_{meY} on the dispersions of the revenue-capital ratio is identical to that of σ_{meK}. $\sigma_{me\pi}$ has no effect on the dispersions since measurement errors on profits do not affect the revenue-capital ratio. Finally, σ_τ and $\sigma_{\log\alpha}$ continue to have first-order effects on $bsd\left(\log\left(Y/\hat{K}\right)\right)$. 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 σ_τ 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 Θ, the set of parameters to estimate. There are a total of 13 parameters, including the key parameter σ_τ; 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

[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, μ; standard deviation of idiosyncratic shocks, σ; and standard deviations of measurement errors in capital, revenue and profit, σ_{meK}, σ_{meY}, and $\sigma_{me\pi}$.

[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 $\pi_{i,t}/Y_{i,t}$, $\log\left(Y_{i,t}/\hat{K}_{i,t}\right)$, $I_{i,t}/K_{i,t}$ and $\Delta\log Y_{i,t}$, together with the cross correlation (*bcorr*) between the between-group $\pi_{i,t}/Y_{i,t}$ and $\log\left(Y_{i,t}/\hat{K}_{i,t}\right)$. The following section will establish the identification conditions through which Θ 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 $3\times3\times3$ types of firms. Each consists of a fixed proportion; i.e., $1/(3\times3\times3)$, 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 $5\times5\times5$. 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 difference 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 σ_{meK}, σ_{meY} and $\sigma_{me\pi}$, respectively. Column (5) reports the moments when σ_{meK}, σ_{meY} and $\sigma_{me\pi}$ are all positive. We find that σ_{meK} only affects moments on $\log\left(Y_{i,t}/\hat{K}_{i,t}\right)$ and $I_{i,t}/K_{i,t}$; σ_{meY} only affects moments on $\log\left(Y_{i,t}/\hat{K}_{i,t}\right)$, $\pi_{i,t}/Y_{i,t}$ and $\Delta\log Y_{i,t}$; and $\sigma_{me\pi}$ 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 $wsd\left(\log\left(Y/\hat{K}\right)\right)$, 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.[14] 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 π by the difference between sales and the cost of goods sold. Ideally, π 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

[14] These firms account for about 90 percent of the total industrial output.

overhead for production.[15]

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, δ, r and P_t^K 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_t^K to be constant and normalize it to unity. (13) implies that $J_t = J$, where $J \equiv (r+\delta)/(1+r)$. δ 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.[16] Our main findings are robust to alternative values of r.[17]

4.3 The Generalized ARP Approach

The first column of Table 4 reports the results from the NBS sample. The generalized ARP approach finds $\sigma_\tau = 0.684$. The value of σ_τ is quantitatively large. According to (16), it implies that a firm with τ_i at the 75th percentile would face a capital goods price 2.5 times higher than the price for a firm with τ_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 α_i and η_i. Under the log-normality specification (17), the estimated $\mu_{\log\alpha}$

[15] 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.

[16] 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).

[17] Table A.5 shows the results from robustness tests in a full-fledged model. The results in the simple model are available upon request.

and $\sigma_{\log\alpha}$ imply that α_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.[18] By (18), the estimated $\mu_{\log\eta}$ and $\sigma_{\log\eta}$ imply that η_i has a mean of 0.082 and a standard deviation of 0.076. This translates into markups of 1.090.

The value of σ_τ 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 $bsd\left(\log\left(Y/\hat{K}\right)\right)$, rather than a cross-sectional standard deviation, would reduce σ_τ 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 $\log\left(Y/\hat{K}\right)$, 0.867^2, and the calibrated σ_τ^2, 0.684^2, be driven by heterogeneities in α_i and η_i. So, the unobserved heterogeneities would account for 38 percent of $bsd\left(\log\left(Y/\hat{K}\right)\right)$.

4.3.1 Unobserved Heterogeneities within and across Industries

The heterogeneities in α_i and η_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 α_i and η_i, we calibrate σ_τ in each of the 30 two-digit industries by the generalized ARP approach. The calibrated value of σ_τ 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 $bsd\left(\log\left(Y/\hat{K}\right)\right)$ and the calibrated σ_τ in each industry, respectively. In absence of heterogeneities in α_i and η_i, the calibrated σ_τ should be located on the 45 degree. The figure shows a big difference between $bsd\left(\log\left(Y/\hat{K}\right)\right)$ and the calibrated σ_τ. The mean ratio of the calibrated σ_τ^2 to the variance of $\log\left(Y/\hat{K}\right)$ 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 σ_τ^2 to the variance of $\log\left(Y/\hat{K}\right)$ 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 α_i and η_i are indeed smaller than the overall dispersions. The mean of the calibrated σ_α and σ_η 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 δ or r, if they exist, tend to bias the estimate of σ_τ upwards. Heterogeneous δ 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 σ_τ for Compustat firms and take it as the benchmark. If the heterogeneity in J has similar magnitude across economies, the difference of the calibrated σ_τ 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. σ_τ 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 σ_τ 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 $bsd\left(\log\left(Y/\hat{K}\right)\right)$ has the same order of magnitude across the three samples, while the calibrated σ_τ differs a lot. In particular, despite the relatively small 30-percent difference in $bsd\left(\log\left(Y/\hat{K}\right)\right)$ between NBS and Compustat II, the calibrated of σ_τ essentially goes down to zero in Compustat II. The unobserved heterogeneities in α_i and η_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 η_i is isomorphic to the heterogeneity in τ_i^Y – the measure of product market distortions in Hsieh and Klenow (2009).[19] 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 σ_τ 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.[20] We do not estimate ρ structurally since the model cannot distinguish between a stationary process with heterogeneous μ and σ and a unit root process with homogeneous μ and σ. Appendix 7.7 will investigate the sensitivity of our estimates to different values of r and ρ 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.

[Insert Table 5]

σ_τ has an estimated value of 0.71. The small simulated standard error suggests that the estimate significantly differs from zero. The structurally estimated σ_τ 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.

[19] See also Peters (2011) where dispersion in markups leads to inefficient TFP losses.

[20] ρ 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 $\log\pi_{i,t-1}$, $\log\hat{K}_{i,t}$, $\log\hat{K}_{i,t-1}$ 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.[21] 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 μ 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. σ 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 σ_{meK} and $\sigma_{me\pi}$ are significant and quantitatively large. By contrast, the model finds σ_{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 ρ; allow the long-run growth rate of $Z_{i,t}$, μ, and the level of uncertainty, σ, 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 σ_τ 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 α_i and η_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 are not important for the estimation of the key parameters, the estimated values of σ_τ in the second and third models are close to that in the benchmark model. Assuming no heterogeneities in α_i and η_i, instead, leads to a considerably higher estimate of σ_τ.

Table 6 presents results for Compustat firms. The estimated value of σ_τ for firms in Compustat I is significant and almost identical to that obtained by the generalized ARP approach. The structural estimation also finds σ_τ to be close to that inferred by the generalized ARP approach for firms in Compustat II. Moreover, the structurally estimated σ_τ 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 $b^i > 0$ 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 α_i and η_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 σ_τ 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 $bsd\left(\log\left(Y/\hat{K}\right)\right)$. 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 σ_τ 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 σ_τ 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 τ_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 $sd\left(\log\left(Y/\hat{K}\right)\right)$ 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

$$w_{i,t} = \left(1 + \tau_i^L\right) w_t, \tag{26}$$

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

Substituting (26) back into (5) yields

$$\log \frac{Y_{i,t}}{L_{i,t}} = \log w_t + \log\left(1 + \tau_i^L\right) - \log\left[\beta_i\left(1 - \eta_i\right)\right]. \tag{27}$$

(27) is akin to (20). (5) allows us to obtain $bsd\left(\log\left(\beta\left(1-\eta\right)\right)\right)$ by computing the between-group standard deviation of labor income share.[22] Then, a combination of (27) and (5) implies a simple way of backing out σ_{τ^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 $\beta_i\left(1-\eta_i\right)$ in (27). By contrast, we cannot infer $\alpha_i\left(1-\eta_i\right)$ 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:

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

[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.[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 σ_{τ^L} are substantially smaller than the calibrated value of σ_τ. Accordingly, a removal of labor misallocation by reducing σ_{τ^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 α_i and τ_i to be exogenous. This assumption does not seem innocuous if a firm can choose α_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 α_i and τ_i, which biases the estimate of σ_τ. This section develops a model that incorporates a technological choice on α_i. Then, we show a simple way of back out σ_τ 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,

$$\alpha_i^* = \arg \max_{\alpha_i \in \{\alpha_l, \alpha_h\}} E\left[V\left(\alpha_i\right)\right], \tag{28}$$

where $E\left[V\left(\alpha_i\right)\right]$ stands for the ex-ante expected firm value conditional on the technological choice of α_i. In the simple model without capital adjustment costs, $E\left[V\left(\alpha_i\right)\right]$ 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 $I_i = \delta \hat{K}_i$, we have

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

where

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

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

and Z, independent of α_i, is an unimportant constant.[24] Since $\gamma_i\left(\alpha_l\right) > \gamma_i\left(\alpha_h\right)$, substituting (29) back into (28) yields

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

where

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

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

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

For any given β and m, the generalized ARP approach can directly be applied to back out α_l, α_h, $\mu_{\log \eta}, \sigma_{\log \eta}$ and σ_τ by matching the five core moments. The only difference is that $\mu_{\log \alpha}$ and $\sigma_{\log \alpha}$ are now replaced with α_l and α_h. We calibrate β 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 α_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 α_i.

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

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 σ_τ (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 α_h in the distortionless environment, respectively. The parameter of key interest, σ_τ, is basically unaffected.

[Insert Table 7]

5.3 Regressions on Firm Characteristics

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

Specifically, we run simple regression of $\frac{1}{T}\sum_{t=1}^{T}\log\left(Y_{i,t}/\hat{K}_{i,t}\right)$ on firm characteristics. We add four-digit industry dummies and the average profit-revenue ratio, $\frac{1}{T}\sum_{t=1}^{T}\log\left(\pi_{i,t}/Y_{i,t}\right)$, 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 α_i and η_i since it is co-determined by α_i and η_i. Nevertheless, our simulations suggest that $bsd\left(\log\left(Y/\hat{K}\right)\right)$ is not sensitive to $\sigma_{\log\eta}$ (see Table 1). This provides a justification for omitting the heterogeneity in η_i in the regression.

The regression equation is:

$$\frac{1}{T}\sum\nolimits_{t=1}^{T}\log\left(Y_{i,t}/\hat{K}_{i,t}\right) = b_0 + b_1\cdot\frac{1}{T}\sum\nolimits_{t=1}^{T}\log\left(\pi_{i,t}/Y_{i,t}\right) + b_2\cdot D_i + b_3\cdot X_i + \xi_i, \tag{31}$$

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.[25]

[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.

[Insert Table 8]

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.[26] 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

[26] 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.[27] It would be interesting in exploring dynamic welfare implications of misallocation.

References

[1] Abel, Andrew B. and Janice C. Eberly (1994), "A Unified Model of Investment under Uncertainty," *American Economic Review*, 84, 1369-1384.

[2] Angeletos, George-Marios and Vasia Panousi (2011), "Financial Integration, Entrepreneurial Risk, and Global Dynamics," *Journal of Economic Theory*, 146, 863-896.

[3] Asker John, Allan Collard-Wexler and Jan De Loecker (2011), "Dynamic Inputs and Resource (Mis)Allocation," forthcoming, *Journal of Political Economy.*

[4] Bai, Chong-En, Chang-Tai Hsieh and Yingyi Qian (2006), "The Returns to Capital in China," *Brookings Papers on Economic Activity*, 37, 61-102.

[5] Banerjee, Abhijit V. and Esther Duflo (2005), "Growth Theory through the Lens of Development Economics," in Philippe Aghion & Steven Durlauf (ed.), *Handbook of Economic Growth*, edition 1, volume 1, chapter 7, 473-552.

[6] Bartelsman Eric J., Randy A. Becker, and Wayne B. Gray (2000), "NBER Productivity Database," available at http://www.nber.org/nberces.

[27] One exception is Michael Peters (2011).

[7] Bloom, Nick (2000), "The Dynamic Effects of Real Options and Irreversibility on Investment and Labour Demand," Institute for Fiscal Studies Working Paper No.W00/15.

[8] Bloom, Nick (2009), "The Impact of Uncertainty Shocks," *Econometrica*, 77, 623-685.

[9] Blundell, Richard and Stephen R. Bond (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115-143.

[10] Brandt, Loren, Johannes Van Biesebroeck and Yifan Zhang (2012), "Creative Accounting or Creative Destruction? Firm-level Productivity Growth in Chinese Manufacturing," *Journal of Development Economics*, 97, 339-351.

[11] Brandt, Loren, Trevor Tombe and Xiaodong Zhu (2013), "Factor Market Distortions Across Time, Space and Sectors in China," *Review of Economic Dynamics*, 16, 39-58.

[12] Buera, Francisco J., Joseph P. Kaboski and Yongseok Shin (2011), "Finance and Development: A Tale of Two Sectors," *American Economic Review*, 101, 1964-2002.

[13] Buera, Francisco J. and Yongseok Shin (2013), "Financial Frictions and the Persistence of History," *Journal of Political Economy*, 121, 221-272.

[14] Caselli, Francesco (2005), "Accounting for Cross-Country Differences," in Philippe Aghion & Steven Durlauf (ed.), *Handbook of Economic Growth*, edition 1, volume 1, chapter 9, 679-741.

[15] Caselli, Francesco and James Feyrer (2007), "The Marginal Product of Capital," *Quarterly Journal of Economics*, 122, 535-568.

[16] Caselli, Francesco and Nicola Gennaioli (2013), "Dynastic Management," *Economic Inquiry*, 51, 971-996.

[17] Cooper, Russell W. and John Haltiwanger (2006), "On the Nature of Capital Adjustment Costs," *Review of Economic Studies*, 73, 611-634.

[18] Cooper, Russell W., Guan Gong and Ping Yan (2010), "Dynamic Labor Demand in China: Public and Private Objectives," NBER Working Paper, No. 16498.

[19] Dollar, David and Shang-jin Wei (2007), "Das (Wasted) Kapital: Firm Ownership and Investment Efficiency in China," NBER Working Paper, No. 13103.

[20] Eckstein, Zvi and Kenneth I. Wolpin (1999), "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities," *Econometrica*, 67, 1295-1339.

[21] Fazzari, S.M., R. G. Hubbard, and B. C. Petersen (1988), "Financing Constraints and Corporate Investment," *Brookings Papers on Economic Activity*, 1, 141-195.

[22] Hadlock, Charles J. and Joshua R. Pierce (2010), "New Evidence on Measuring Financial Constraints: Moving Beyond the KZ Index," *Review of Financial Studies*, 23, 1909-1940.

[23] Hsieh, Chang-Tai and Peter J. Klenow (2009), "Misallocation and Manufacturing TFP in China and India," *Quarter Journal of Economics*, 124, 1403-1448.

[24] Hsieh, Chang-Tai and Zheng Song (2014), "Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China," mimeo.

[25] Gouriéroux, Christian and Alain Monfort (1996), *Simulation-Based Econometric Methods*, Oxford University Press.

[26] Gilchrist, Simon, Jae Sim and Egon Zakrajsek (2013), "Misallocation and Financial Frictions: Some Direct Evidence from the Dispersion in Borrowing Costs", *Review of Economic Dynamcis*, 16, 159-176.

[27] Khan, Aubhik and Julia K. Thomas (2006), "Adjustment Costs," *The New Palgrave Dictionary of Economics*, Palgrave Macmillan.

[28] Mankiw, N. Gregory and Matthew D. Shapiro (1986), "Risk and Return: Consumption Beta Versus Market Beta," *Review of Economics and Statistics*, 68, 452-459.

[29] Midrigan, Virgiliu and Daniel Yi Xu (2014), "Finance and Misallocation: Evidence from Plant-Level Data," *American Economic Review*, 104, 422-458.

[30] Morck, Randall, Bernard Yeung and Wayne Yu (2000), "The Information Content of Stock Markets: Why do Emerging Markets have Synchronous Stock Price Movements?" *Journal of Financial Economics*, 58, 215-260.

[31] Moll, Benjamin (2014), "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation," forthcoming, *American Economic Review*.

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

[33] Pavcnik, Nina (2002), "Trade Liberalization, Exit, and Productivity Improvement: Evidence from Chilean Plants," *Review of Economic Studies*, 69, 245-276.

[34] Perkins, Dwight H. and Thomas G. Rawski (2008), "Appendix to Forecasting China's Economic Growth to 2025," mimeo.

[35] Peters, Michael (2011), "Heterogeneous Mark-Ups and Endogenous Misallocation," mimeo.

[36] Qian, Zhenjie and Xiaodong Zhu (2012), "Why Is Labor Income Share So Low in China?" mimeo.

[37] Song, Zheng, Kjetil Storesletten and Fabrizio Zilibotti (2011), "Growing Like China," *American Economic Review*, 101, 202-241.

[38] Restuccia, Diego, and Richard Rogerson (2008), "Policy Distortions and Aggregate Productivity with Heterogeneous Plants," *Review of Economic Dynamics*, 11, 707-720.

[39] Restuccia, Diego, and Richard Rogerson (2013), "Misallocation and Productivity," *Review of Economic Dynamics*, 16, 1-10.

[40] WU, Guiying Laura (2009), *Uncertainty, Investment and Capital Accumulation: A Structural Econometric Approach*, D.Phil. Thesis, University of Oxford.

[41] WU, Guiying Laura (2014), "Investment Frictions and the Aggregate Output Loss in China," forthcoming, *Oxford Bulletin of Economics and Statistics*.

[42] Zhang, Junsen, Yaohui Zhao, Albert Park and Xiaoqing Song (2005), "Economic Returns to Schooling in Urban China, 1988 to 2001," *Journal of Comparative Economics*, 33, 730-752.

Figure 1: Identification in the Full-Blown Model

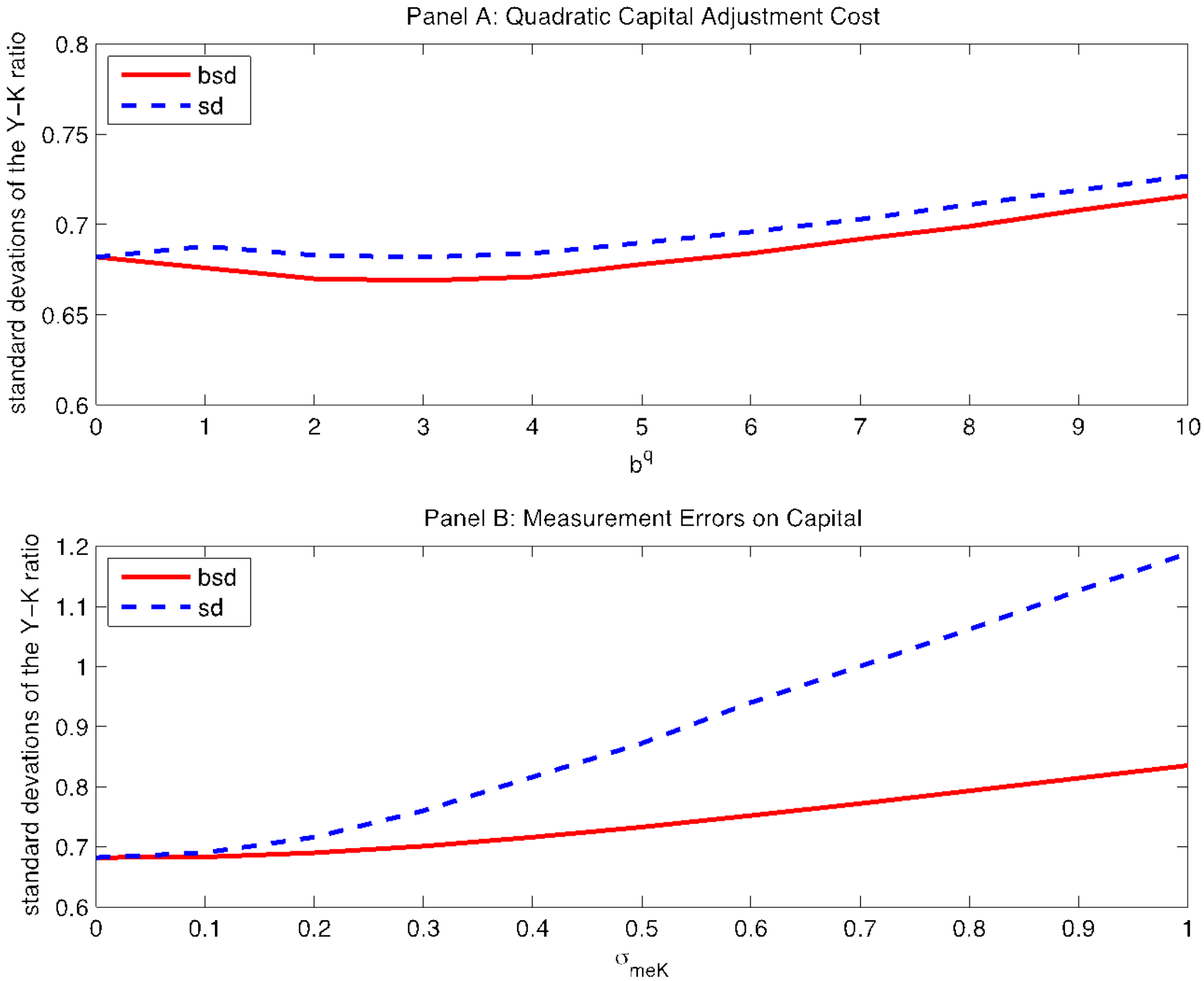

Figure 1: This figure plots the sensitivity of the standard deviation (*sd*) and the between-group standard deviation (*bsd*) of *log*(*Y*/*Khat*) with respect to quadratic capital adjustment cost (Panel A) and measurement error on capital (Panel B), respectively. The benchmark parameterization is set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

Figure 2: Results from Two- and Four-Digit Industries

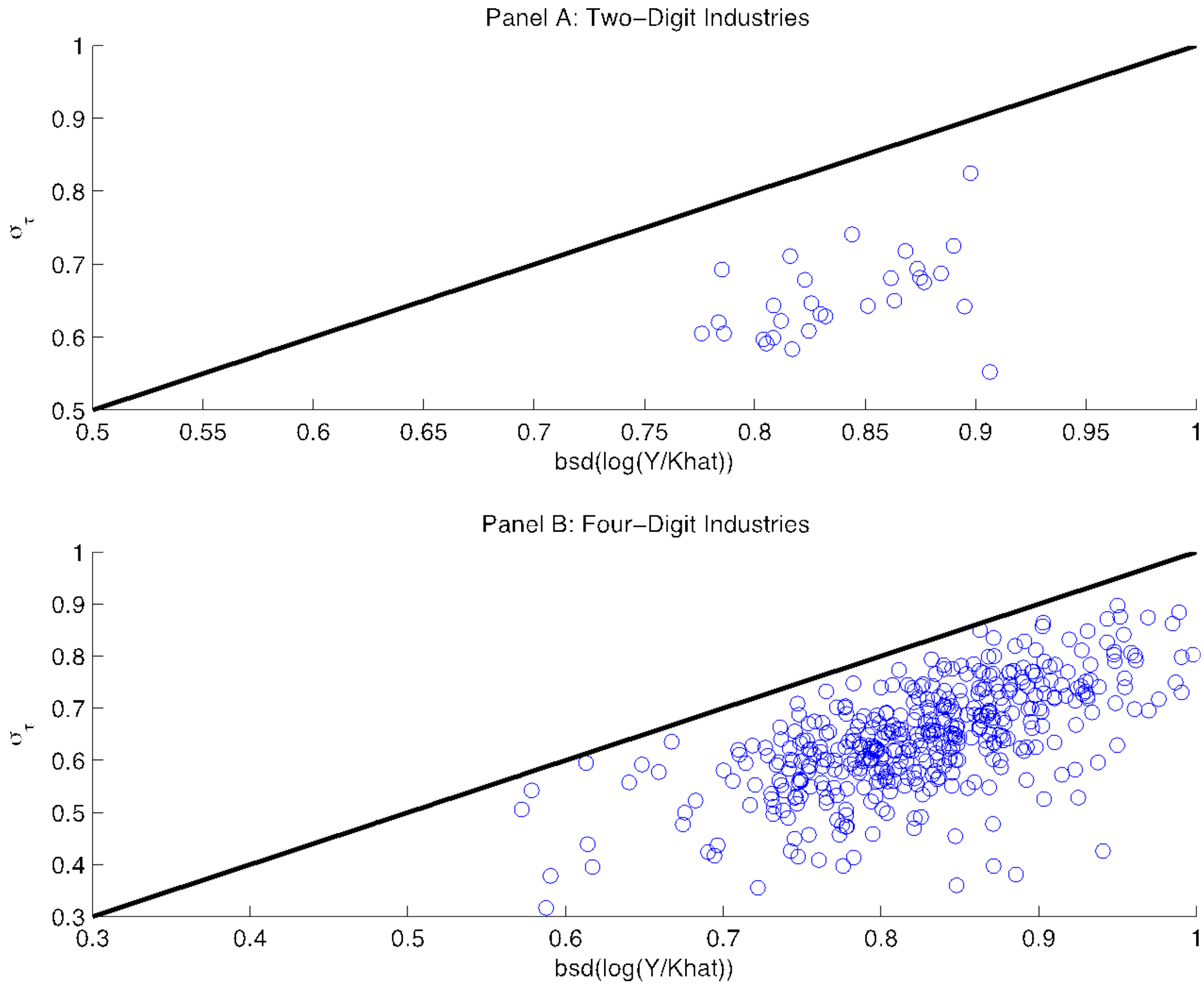

Figure 2: Panel A plots the between-group standard deviation of *log*(*Y*/*Khat*) (x-axis) and the calibrated σ_τ (y-axis) by the generalized ARP approach in each two-digit industry. The solid line is the 45 degree line. Panel B plots the results from each four-digit industry.

Table 1 Illustration for Identification in the Simple Model

	Panel A				
Parameters	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	$\sigma_\tau = 0.0$	$\sigma_\tau = 0.5$	$\sigma_\tau = 0.0$	$\sigma_\tau = 0.0$	$\sigma_\tau = 0.5$
	$\sigma_{log\alpha} = 0.0$	$\sigma_{log\alpha} = 0.0$	$\sigma_{log\alpha} = 0.5$	$\sigma_{log\alpha} = 0.0$	$\sigma_{log\alpha} = 0.5$
	$\sigma_{log\eta} = 0.0$	$\sigma_{log\eta} = 0.0$	$\sigma_{log\eta} = 0.0$	$\sigma_{log\eta} = 0.5$	$\sigma_{log\eta} = 0.5$
Set of Moments					
mean(π/Y)	0.157	0.157	0.167	0.167	0.170
mean(log(Y/Khat))	0.834	0.834	0.834	0.847	0.840
bsd(π/Y)	0.000	0.000	0.044	0.044	0.061
bsd(log(Y/Khat))	0.000	0.496	0.495	0.054	0.682
bcorr(π/Y, log(Y/Khat))	N.A.	N.A.	-0.954	0.991	-0.378
			Panel B		
Inferred σ_τ	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
The APR Approach	0.000	0.496	0.495	0.054	0.682
The Generalized ARP Approach	N.A.	N.A.	0.000	0.000	0.489
			Panel C		
Estimates for other parameters	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
$\mu_{log\alpha}$	N.A.	N.A.	-2.498	-2.499	-2.498
$\sigma_{log\alpha}$	N.A.	N.A.	0.505	0.000	0.491
$\mu_{log\eta}$	N.A.	N.A.	-2.372	-2.391	-2.493
$\sigma_{log\eta}$	N.A.	N.A.	0.098	0.423	0.490

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 $b^q = b^i = b^f = \sigma_{meK} = \sigma_{meY} = \sigma_{me\pi} = 0$.

Table 2 Illustration for Identification in the Full-Blown Model

			Panel A		
Parameters	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	$b^q = 1$	$b^q = 10$	$\sigma_{meK} = 0.5$	$\sigma_{meK} = 1$	$b^q = 1$, $\sigma_{meK} = 0.5$
Set of Moments					
mean(π/Y)	0.170	0.170	0.170	0.170	0.170
mean(log(Y/Khat))	0.865	1.209	0.840	0.765	0.854
bsd(π/Y)	0.061	0.061	0.061	0.061	0.061
bsd(log(Y/Khat))	0.676	0.716	0.733	0.835	0.711
bcorr(π/Y, log(Y/Khat))	-0.392	-0.350	-0.350	-0.304	-0.371
			Panel B		
Inferred σ_τ	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
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	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
$\mu_{log\alpha}$	-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 $\delta = 0.05$, $r = 0.15$, $\mu_{log\alpha} = \mu_{log\eta} = -2.50$, $\sigma_{log\alpha} = \sigma_{log\eta} = 0.50$, $\rho = 0.90$, $\mu = 0.05$ and $\sigma = 0.40$.

Column (1) and (2): $b^i = b^f = \sigma_{meK} = \sigma_{meY} = \sigma_{me\pi} = 0$

Column (3) and (4): $b^q = b^i = b^f = \sigma_{meY} = \sigma_{me\pi} = 0$

Column (5): $b^i = b^f = \sigma_{meY} = \sigma_{me\pi} = 0$

Table 3: Parameters and Moments in the Structural Estimation

Parameters	Definition
σ_τ	standard deviation of heterogeneities in capital goods price
$\mu_{log\alpha}$	mean of log capital output elasticity in production function
$\sigma_{log\alpha}$	standard deviation of log capital output elasticity
$\mu_{log\eta}$	mean of log inverse of demand elasticity
$\sigma_{log\eta}$	standard deviation of log inverse of demand elasticity
b^q	quadratic adjustment costs
b^i	partial irreversibility
b^f	fixed adjustment costs
μ	mean of growth rate in $Z_{i,t}$
σ	standard deviation of shocks to $Z_{i,t}$
σ_{meK}	standard deviation of measurement errors in capital stock
σ_{meY}	standard deviation of measurement errors in sales revenue
$\sigma_{me\pi}$	standard deviation of measurement errors in gross profit
Moments	**Definition**
mean(π/Y)	mean of profit-revenue ratio
mean(log(Y/Khat))	mean of log revenue-capital ratio
mean(I/K)	mean of investment rate
mean(ΔlogY)	mean of revenue growth rate
bsd(π/Y)	between-group standard deviation of profit-revenue ratio
wsd(π/Y)	within-group standard deviation of profit-revenue ratio
bsd(log(Y/Khat))	between-group standard deviation of log revenue-capital ratio
wsd(log(Y/Khat))	within-group standard deviation of log revenue-capital ratio
bsd(I/K)	between-group standard deviation of investment rate
wsd(I/K)	within-group standard deviation of investment rate
bsd(ΔlogY)	between-group standard deviation of revenue growth rate
wsd(ΔlogY)	within-group standard deviation of revenue growth rate
skew(π/Y)	skewness of profit-revenue ratio
skew(log(Y/Khat))	skewness of log revenue-capital ratio
skew(I/K)	skewness of investment rate
skew(dlogY)	skewness of revenue growth rate
scorr(π/Y)	serial correlation of profit-revenue ratio
scorr(log(Y/Khat))	serial correlation of log revenue-capital ratio
scorr(I/K)	serial correlation of investment rate
scorr(ΔlogY)	serial correlation of revenue growth rate
bcorr(π/Y, log(Y/Khat))	cross correlation between between-group profit-revenue ratio and log revenue-capital ratio

Table 4: Results by the Generalized ARP Approach

Parameters	**NBS**	**Compustat I**	**Compustat II**
σ_τ	0.6843	0.3095	0.1313
	(0.7143)	(0.3020)	(0.1113)
$\mu_{log\alpha}$	-2.6189	-1.9320	-1.9788
	(-2.6058)	(-1.8745)	(-2.1797)
$\sigma_{log\alpha}$	0.5539	0.6259	0.5643
	(0.5568)	(0.6271)	(0.5338)
$\mu_{log\eta}$	-2.8086	-1.4527	-1.6241
	(-2.8084)	(-1.5831)	(-1.6313)
$\sigma_{log\eta}$	0.7887	0.5246	0.5587
	(0.7253)	(0.5949)	(0.5235)
Moments			
mean(π/Y)	0.1578	0.3929	0.3514
mean(log(Y/Khat))	1.1377	0.5659	0.5542
bsd(π/Y)	0.0763	0.1591	0.1448
bsd(log(Y/Khat))	0.8666	0.7288	0.6064
bcorr(π/Y, log(Y/Khat))	-0.2422	-0.0738	-0.0879
Descriptive Statistics			
No. of firms	107579	1431	970
span of years	2004-2007	2002-2005	2002-2005
mean(sales)	97	2066	3132
med(sales)	20	121	461
mean(employees)	0.33	10.4	20.8
med(employees)	0.13	0.8	4.3

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 δ = 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

Parameters	estimate	s.e.
σ_τ	0.7143	0.0033
$\mu_{log\alpha}$	-2.6058	0.0019
$\sigma_{log\alpha}$	0.5568	0.0043
$\mu_{log\eta}$	-2.8084	0.0051
$\sigma_{log\eta}$	0.7253	0.0061
b^q	0.2777	0.0038
b^i	0.0001	0.0395
b^f	0.0335	0.0006
μ	0.0802	0.0004
σ	0.4253	0.0016
σ_{meK}	0.4010	0.0013
σ_{meY}	0.0007	0.1255
$\sigma_{me\pi}$	0.5777	0.0020

Moments	empirical	s.e.	simulated
mean(π/Y)	0.1578	0.0002	0.1542
mean(log(Y/Khat))	1.1377	0.0025	1.1456
mean(I/K)	0.1640	0.0005	0.1729
mean(ΔlogY)	0.0963	0.0005	0.0803
bsd(π/Y)	0.0763	0.0001	0.0745
wsd(π/Y)	0.0506	0.0001	0.0488
bsd(log(Y/Khat))	0.8666	0.0011	0.8781
wsd(log(Y/Khat))	0.3470	0.0009	0.3321
bsd(I/K)	0.1991	0.0006	0.1642
wsd(I/K)	0.2027	0.0006	0.2149
bsd(ΔlogY)	0.1876	0.0004	0.1632
wsd(ΔlogY)	0.2042	0.0004	0.2187
skew(π/Y)	0.7760	0.0039	0.8539
skew(log(Y/Khat))	0.0570	0.0038	0.0037
skew(I/K)	2.2307	0.0075	2.2510
skew(dlogY)	0.1567	0.0037	0.1760
scorr(π/Y)	0.5703	0.0021	0.5993
scorr(log(Y/Khat))	0.8403	0.0009	0.8377
scorr(I/K)	0.1188	0.0030	0.2430
scorr(ΔlogY)	0.0685	0.0028	0.0526
bcorr(π/Y, log(Y/Khat))	-0.2422	0.0034	-0.2707
OI/100		183	

Note: The imposed parameter values are δ = 0.05, r = 0.20, and ρ = 0.90.

Table 6: Structural Estimation Results for Compustat Firms

Sample	Compustat I		Compustat II	
Parameters	**estimate**	**s.e.**	**estimate**	**s.e.**
σ_τ	0.3020	0.0676	0.1113	0.1822
$\mu_{log\alpha}$	-1.8745	0.0193	-2.1797	0.0225
$\sigma_{log\alpha}$	0.6271	0.0272	0.5338	0.0308
$\mu_{log\eta}$	-1.5831	0.0305	-1.6313	0.0270
$\sigma_{log\eta}$	0.5949	0.0146	0.5235	0.0186
b^q	1.1355	0.1489	0.9992	0.0741
b^i	0.0073	0.1836	0.0008	0.5095
b^f	0.0017	0.0017	0.0003	0.0136
μ	0.0263	0.0017	0.0524	0.0018
σ	0.2636	0.0096	0.2009	0.0060
σ_{meK}	0.2014	0.0117	0.1875	0.0055
σ_{meY}	0.0023	0.2542	0.0010	0.2222
$\sigma_{me\pi}$	0.2286	0.0112	0.1678	0.0089

Sample	Compustat I		Compustat II	
Moments	**empirical**	**simulated**	**empirical**	**simulated**
mean(π/Y)	0.3929	0.3839	0.3458	0.3206
mean(log(Y/Khat))	0.5659	0.5600	0.8560	0.8194
mean(I/K)	0.1338	0.1463	0.1779	0.1752
mean(ΔlogY)	0.0900	0.0274	0.0696	0.0523
bsd(π/Y)	0.1591	0.1624	0.1251	0.1300
wsd(π/Y)	0.0450	0.0481	0.0456	0.0327
bsd(log(Y/Khat))	0.7288	0.7148	0.5297	0.5544
wsd(log(Y/Khat))	0.2301	0.1829	0.2662	0.1978
bsd(I/K)	0.0729	0.0614	0.0755	0.0217
wsd(I/K)	0.0522	0.0582	0.0667	0.0696
bsd(ΔlogY)	0.1483	0.1020	0.1040	0.0239
wsd(ΔlogY)	0.1297	0.1497	0.1182	0.1433
skew(π/Y)	0.4315	0.3018	0.3549	0.3385
skew(log(Y/Khat))	-0.2917	0.0636	-0.0427	0.0555
skew(I/K)	0.9393	0.7311	1.0737	0.6744
skew(dlogY)	0.3168	0.0435	0.6425	0.0052
scorr(π/Y)	0.9183	0.8926	0.9500	0.9374
scorr(log(Y/Khat))	0.9117	0.9245	0.9265	0.9078
scorr(I/K)	0.5630	0.4926	0.6206	0.4558
scorr(ΔlogY)	0.2057	-0.0253	0.3217	-0.0202
bcorr(π/Y, log(Y/Khat))	-0.0738	-0.1149	-0.0298	-0.0398
OI/100	10		26	

Note: See the text for the definition of Compustat I and II. The imposed parameter values are δ = 0.10, r = 0.10, ρ = 0.90.

Table 7: Estimation with Endogenous Capital Output Elasticity

	σ_{τ}	$\mu_{log\eta}$	$\sigma_{log\eta}$	α_l	α_h	β	m
(1)	**0.5784**	**-2.7460**	**0.7841**	**0.0551**	**0.1090**	**0.1369**	**1.2024**
(2)	0.5800	-2.7456	0.7838	0.0555	0.1097	0.1369	1.1769
(3)	0.5887	-2.7434	0.7817	0.0576	0.1142	0.1369	1.0396

Note: m and β 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 α_h if $\tau_i = 0$. (2) and (3) refer to the cases where 20% and 80% of firms would choose α_h if $\tau_i = 0$, respectively.

Table 8: Regressions on Firm Characteristics

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

Note: 1. The parentheses report robust standard errors.
2. 4-digital 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, Chongqin, 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.

7 Appendix

7.1 Aggregate TFPR Gains

We first transform gross output into value added. Define value added as $\hat{Y}_{i,t} \equiv \max_{M_{i,t}} \{Y_{i,t} - m_{i,t}M_{i,t}\}$. This yields

$$\hat{Y}_{i,t} = (1-(1-\alpha_i-\beta_i)(1-\eta_i))\left[\frac{(1-\alpha_i-\beta_i)(1-\eta_i)}{m_{i,t}}\right]^{\frac{(1-\alpha_i-\beta_i)(1-\eta_i)}{1-(1-\alpha_i-\beta_i)(1-\eta_i)}} \left(\hat{Z}_{i,t}^{\frac{\eta_i}{1-\eta_i}}\hat{K}_{i,t}^{\alpha_i}L_{i,t}^{\beta_i}\right)^{\frac{(1-\eta_i)}{1-(1-\alpha_i-\beta_i)(1-\eta_i)}},$$

where $\hat{Z}_{i,t} \equiv X_{i,t}A_{i,t}^{\frac{1}{\eta_i}-1}$. We then calculate efficiency gain from capital reallocation within each type of firms associated with the same α_i and η_i. The aggregate output gain is obtained by averaging the gain across different types of firms.

For notational convenience, consider an economy in which all firms have the same α, η and, thus, γ. Efficient capital allocation features identical MRPK across firms. Without loss of generality, we drop time subscript. For simplicity, we assume away labor and intermediate input distortions such that $w_i = w$ and $m_i = m$. Value-added would, thus, follow

$$\hat{Y}_i = \left(\hat{Z}_i^{\frac{\eta}{1-\eta}}\hat{K}_i^{\alpha}L_i^{\beta}\right)^{\frac{(1-\eta)}{1-(1-\alpha-\beta)(1-\eta)}}, \tag{32}$$

where irrelevant terms are omitted.

Denote L_i^* and $\hat{K}_i^*$ firm i's labor and productive capital in the efficient allocation. (14) implies $\hat{K}_i^* \propto \hat{Z}_i$. Using (5) to (8), together with the fact that $\hat{K}_i^* \propto \hat{Z}_i$, we have $L_i^* \propto \hat{Z}_i$. Define $\hat{K} \equiv \sum_i \hat{K}_i^*$ and $L \equiv \sum_i L_i^*$ as the total productive capital and labor, respectively. (32) implies that the total value added in the efficient allocation is equal to

$$\hat{Y}^* = \left(\hat{K}^{\alpha}L^{\beta}\right)^{\frac{(1-\eta)}{1-(1-\alpha-\beta)(1-\eta)}} \times \frac{\sum_i \hat{Z}_i}{\left(\sum_i \hat{Z}_i\right)^{\frac{(\alpha+\beta)(1-\eta)}{1-(1-\alpha-\beta)(1-\eta)}}}. \tag{33}$$

We omit irrelevant constant terms.

Now we turn to the actual total value added in the economy with capital distortions. (14) implies $\hat{K}_i \propto \hat{Z}_i/(1+\tau_i)^{\frac{\eta+\alpha(1-\eta)}{\eta}}$. Moreover, (5) to (8) establish that $L_i \propto \hat{Z}_i/(1+\tau_i)^{\frac{\alpha(1-\eta)}{\eta}}$. Then, the actual total value added follows

$$\hat{Y} = \left(\hat{K}^{\alpha}L^{\beta}\right)^{\frac{(1-\eta)}{1-(1-\alpha-\beta)(1-\eta)}} \times \frac{\sum_i \frac{\hat{Z}_i}{(1+\tau_i)^{\frac{\alpha(1-\eta)}{\eta}}}}{\left[\left(\sum_i \frac{\hat{Z}_i}{(1+\tau_i)^{\frac{\eta+\alpha(1-\eta)}{\eta}}}\right)^{\alpha}\left(\sum_i \frac{\hat{Z}_i}{(1+\tau_i)^{\frac{\alpha(1-\eta)}{\eta}}}\right)^{\beta}\right]^{\frac{(1-\eta)}{1-(1-\alpha-\beta)(1-\eta)}}}. \tag{34}$$

The efficiency gain from capital reallocation can, thus, be represented by the difference between $\hat{Y}^*$ and $\hat{Y}$:

$$\log \hat{Y}^* - \log \hat{Y} = \frac{\alpha(1-\eta)}{1-(1-\alpha-\beta)(1-\eta)} \log \frac{\sum_i \frac{\hat{Z}_i}{(1+\tau_i)^{\frac{\eta+\alpha(1-\eta)}{\eta}}}}{\sum_i \hat{Z}_i} - \frac{\eta+\alpha(1-\eta)}{1-(1-\alpha-\beta)(1-\eta)} \log \frac{\sum_i \frac{\hat{Z}_i}{(1+\tau_i)^{\frac{\alpha(1-\eta)}{\eta}}}}{\sum_i \hat{Z}_i}.$$

With a large number of firms, the efficiency gain can be approximated by:

$$\log \hat{Y}^* - \log \hat{Y} \approx \frac{1}{2}\frac{\alpha(1-\eta)}{\eta}\frac{\eta+\alpha(1-\eta)}{1-(1-\alpha-\beta)(1-\eta)} Var\left[\log(1+\tau_i)\right]. \tag{35}$$

(35) shows that the efficiency gain from removing capital misallocation is proportional to the variance of $\log(1+\tau_i)$.

Now we aggregate the efficiency gain across different firm types.

$$\text{Aggregate output gain} = \log \int \hat{Y}^*(j)\,\psi(j) - \log \int \hat{Y}(j)\,\psi(j),$$

where $\psi(j)$ represents the density for the number of firms associated with $\alpha(j)$ and $\eta(j)$.

7.1.1 Labor Market Distortions

We now introduce labor distortions but maintain the assumption that intermediate inputs are efficiently allocated. (14) implies $\hat{K}_i \propto \hat{Z}_i / \left((1+\tau_i)^{\frac{\eta+\alpha(1-\eta)}{\eta}} \left(1+\tau_i^L\right)^{\frac{\beta(1-\eta)}{\eta}}\right)$. Equations (5), (26), together with (7) and (8), imply that $L_i \propto \hat{Z}_i / \left((1+\tau_i)^{\frac{\alpha(1-\eta)}{\eta}} \left(1+\tau_i^L\right)^{\frac{\beta(1-\eta)+\eta}{\eta}}\right)$. Therefore, the total value added with both capital and labor distortions follows

$$\hat{Y}^L = \left(\hat{K}^\alpha L^\beta\right)^{\frac{(1-\eta)}{1-(1-\alpha-\beta)(1-\eta)}} \times \frac{\sum \frac{\hat{Z}_i}{(1+\tau_i)^{\frac{\alpha(1-\eta)}{\eta}}\left(1+\tau_i^L\right)^{\frac{\beta(1-\eta)}{\eta}}}}{\left[\left(\sum_i \frac{\hat{Z}_{i,t}}{(1+\tau_i)^{\frac{\eta+\alpha(1-\eta)}{\eta}}\left(1+\tau_i^L\right)^{\frac{\beta(1-\eta)}{\eta}}}\right)^\alpha \left(\sum_i \frac{\hat{Z}_{i,t}}{(1+\tau_i)^{\frac{\alpha(1-\eta)}{\eta}}\left(1+\tau_i^L\right)^{\frac{\beta(1-\eta)+\eta}{\eta}}}\right)^\beta\right]^{\frac{(1-\eta)}{1-(1-\alpha-\beta)(1-\eta)}}}.$$

Efficient allocation features identical marginal revenue products of both capital and labor across firms. It is immediate that the total value added in the efficient allocation is identical to that in (33). The efficiency gain from reallocation can thus be approximated by

$$\log \hat{Y}^* - \log \hat{Y}^L \approx \frac{1}{2}\frac{\alpha(1-\eta)}{\eta}\frac{\eta+\alpha(1-\eta)}{1-(1-\alpha-\beta)(1-\eta)} Var\left[\log(1+\tau_i)\right] + \frac{1}{2}\frac{\beta(1-\eta)}{\eta}\frac{\eta+\beta(1-\eta)}{1-(1-\alpha-\beta)(1-\eta)} Var\left[\log\left(1+\tau_i^L\right)\right], \tag{36}$$

where we assume that τ_i and τ_i^L are uncorrelated. The second term on the right-hand side of (36) captures the welfare gain from correcting labor misallocation.

7.2 Investment Decision

The investment problem is defined by the stochastic Bellman equation:

$$V(Z_{i,t}, K_{i,t}) = \max_{I_{i,t}} \left\{ \pi(Z_{i,t}, K_{i,t}; I_{i,t}) - P_{i,t}^K I_{i,t} - G(K_{i,t}; I_{i,t}) + \frac{1}{1+r} E_t \left[V(Z_{i,t+1}, K_{i,t+1}) \right] \right\}, \tag{37}$$

where $Z_{i,t+1}$ and $K_{i,t+1}$ follow the law of motion (19) and (12), respectively. When $G(Z_{i,t}, K_{i,t}; I_{i,t}) = 0$, the optimal investment rate is a linear function of $Z_{i,t}/K_{i,t}$:

$$\frac{I_{i,t}}{K_{i,t}} = \left[\frac{1-\gamma_i}{(1+\tau_i) J_t} \right]^{\frac{1}{\gamma_i}} \left(\frac{Z_{i,t}}{K_{i,t}} \right) - 1, \tag{38}$$

which leads to (14). When $G(K_{i,t}; I_{i,t}) > 0$, the investment policy can be solved numerically. Wu (2014) provides further details.

7.3 China Data

Brandt et al. (2012) provide an excellent description of the dataset and implement a series of consistency checks. We strictly follow them in constructing a panel and cleaning the data. A few things deserve attention. The first is how to construct capital data. The survey does not contain information on investment expenditures. However, firms report the book value of their fixed capital stock at original purchase prices. Since these book values are the sum of nominal values from different years, they should not be used directly. To construct the real capital stock series, we use the following formula:

$$K_{i,t} = (1-\delta) K_{i,t-1} + \frac{BK_{i,t} - BK_{i,t-1}}{P_t},$$

where $BK_{i,t}$ is the gross book value of capital stock for firm i in year t; P_t is the price index of investment in fixed assets in year t constructed by Perkins and Rawski (2008). The initial book value of capital stock is taken directly from the dataset for firms founded later than 1998. For firms founded before 1998, we predict it to be

$$BK_{i,t_0} = \frac{BK_{i,t_1}}{(1+g_i)^{t_1 - t_0}},$$

where BK_{i,t_0} is the projected initial book value of capital stock when firm i was born in year t_0; BK_{i,t_1} is the book value of capital stock when firm i first appears in our dataset in year t_1; and g_i is the average capital stock growth rate of firm i for the period we observe in the data since year t_1.

The calibration of δ is based on the law of motion of capital (12), which implies that

$$\log\left(1+\frac{I_{i,t}}{K_{i,t}}\right) = \triangle \log \hat{K}_{i,t} - \log(1-\delta)$$
$$\simeq \triangle \log \hat{K}_{i,t} + \delta.$$

The model implies that both $\hat{K}_{i,t}$ and $Y_{i,t}$ grow at the same rate in the long run.[28] So, the above equation suggests calibrating δ by matching the difference between $\log(1+I_{i,t}/K_{i,t})$ and $\Delta \log Y_{i,t}$. This gives $\delta = 0.05$. Investment expenditure $I_{i,t}$ is then recovered according to equation (12).

Four key variables for estimation are then constructed by definition: profit-revenue ratio, $(\pi_{i,t}/Y_{i,t})$, log revenue-capital ratio, $\left(\log\left(Y_{i,t}/\hat{K}_{i,t}\right)\right)$, investment rate, $(I_{i,t}/K_{i,t})$, and revenue growth rate, $(\Delta \log Y_{i,t})$. The revenue and profit data are deflated by the GDP deflator for the secondary industry from the *China Statistical Yearbook*. We exclude outliers by trimming the top and bottom 5 percent of observations for each variable in each year. The model assumes that firms are on the balanced-growth path. In the presence of capital adjustment costs, however, it would take years for firms to reach their balanced-growth path. Therefore, we exclude firms that are less than 5 years old when they first enter our dataset. Furthermore, our investment model does not consider entry and exit, which means that the model's implications are valid only for existing and ongoing firms. Finally, many existing non-state-owned firms with sales revenue beyond RMB 5 millions were missing from the survey in the early years but have appeared in the NBS data since 2004 thanks to the economic census conducted in that year. For these reasons, our empirical exercise utilizes a sample of firms surviving from 2004 to 2007 and being at least 5 years old in 2004. This gives us a balanced panel consisting of 107,579 firms and spanning 4 years. The annual mean values of each of the four variables in the balanced panel are reported in Table A.3.

[Insert Table A.3]

Our simulations in the structural estimation assume firms to be around their balanced growth paths. Since the estimation is to match moments of the four variables, we need to check the stationarity of the four variables from a fast-changing economy like China's. It is hard to give a formal test, given the fact that the panel has only four time-series observations. Nevertheless, one can still see from Table A.3 that except for the falling investment rate, none of the other three variables features an obvious trend.

[28] This result carries over to the case with capital adjustment costs. Bloom (2000) shows that when a firm is on its balanced growth path, the gap between capital stock with and without adjustment costs is bounded.

7.4 Compustat Data

We construct capital stock and deflate the data strictly following Bloom (2009). To be specific, capital stocks for firm i in industry m in year t are constructed by the perpetual inventory method: $K_{i,t} = (1-\delta) K_{i,t-1} (P_{m,t}/P_{m,t-1}) + I_{i,t}$, initialized using the net book value of capital, where $\delta = 0.10$, $I_{i,t}$ is net capital expenditures on plant, property, and equipment, and $P_{m,t}$ is the industry-level capital goods deflator from Bartelsman et al. (2000). Sales revenue and cost of goods sold are deflated by the CPI. We use a sample from 2002 to 2005 since $P_{m,t}$ is not available after 2005. Finally, we also trim the top and bottom 5 percent of observations for each variable in each year in Compustat.

7.5 Back-of-the-Envelope Calculations for the Effects of Market Beta and Market Incompleteness

We back out the following distribution of τ_i:

$$\log(1+\tau_i) \overset{i.i.d.}{\sim} N\left(0, 0.684^2\right).$$

Suppose that all the heterogeneity comes from J_i or, more precisely, r_i, the firm-specific discount factor. $\log J_i$ would follow a log-normal distribution with variance of 0.684^2. Since $\log J_i \simeq \log(r_i + \delta)$, it implies that

$$\log(r_i + \delta) \overset{i.i.d.}{\sim} N\left(\log(0.25), 0.684^2\right).$$

We can, thus, back out $var(r_i) = 0.244^2$.

Risk-averse investors would assign higher discount rates to firms with $Z_{i,t}$ that are more correlated to aggregate shocks. To see the capacity of market beta in generating $var(r_i)$, consider a typical CAPM,

$$r_i = r_f + (r_m - r_f) \cdot beta_i,$$

where r_f is the interest rate on riskless assets, r_m is the expected market return and $r_m - r_f$ is the expected market risk premium.

If all the heterogeneity in r_i is driven by heterogeneous market beta, $var(beta_i)$ has to be 2.44^2 to match $var(r_i) = 0.244^2$ with a 10 percent risk premium. Since few firms in the NBS data are listed, we cannot calculate the dispersion of market beta. Morck et al. (2000) find the stock returns in emerging economies to be more synchronous, probably due to a poor capitalization of firm-specific information. This implies that the market betas tend to be less dispersed in emerging economics than those in developed economies. Picking up the number from Mankiw and Shapiro (1986), $var(beta_i)$ is 0.38^2 for 464 U.S. stocks over 92 quarters.

Even if we take this value as the upper limit for $var\,(beta_i)$ in China, $var\,(r_i)$ would be 0.038^2, merely 2.4 percent of the 0.244^2 that is needed to explain the estimated σ_τ^2.

To see the importance of market incompleteness in generating heterogeneity in r_i, we adopt the framework in Angeletos and Panousi (2011). Following their notations in equation (18), we have

$$r_i = r_f + \sqrt{\frac{2\theta\gamma\,(d - r_f)}{\theta + 1}}\sigma_i,$$

where $d > 0$ is the discount factor, $\gamma > 0$ is the coefficient of relative risk aversion, and $\theta > 0$ is the elasticity of intertemporal substitution. What we need is a measure of σ_i and its dispersion in the data.

Assume that $Z_{i,t}$ follows the stochastic process in (19), where σ_i parameterizes the risk level for firm i. Bloom (2000) shows that revenue and capital will still grow at the same rate of μ in the long run. This implies

$$\begin{aligned} \Delta \log Y_{i,t} &= \Delta \log Z_{i,t} \\ &= \mu + z_{i,t} - z_{i,t-1} \\ &\simeq \mu + e_{i,t}, \end{aligned}$$

if $\rho \rightarrow 1$. Therefore, when panel data are available, the variance of a firm's sales growth, $var\,(\Delta \log Y_{i,t})$, serves as a proxy for its risk level, σ_i^2.

In the NBS sample, the value of $var\,(\Delta \log Y_{i,t})$ has an estimate of 0.142^2. If all the heterogeneity in J_i is driven by idiosyncratic risks, $\sqrt{2\theta\gamma\,(d - r_f)\,/\,(\theta + 1)}$ should be as large as 1.718. If $d - r_f = 0.10$ and $\theta = 1$, the coefficient of relative risk aversion γ has to be as large as 30 to match $var\,(r_i) = 0.244^2$. Alternatively, if $\theta = 1$ and $\gamma = 5$, then $var\,(r_i)$ would be equal to 0.1^2, which accounts for about 16.9 percent of the estimated σ_τ^2.

7.6 The Simulated Methods of Moments

The SMM estimator Θ^* solves the following minimal quadratic distance problem (Gouriéroux and Monfort, 1996):

$$\Theta^* = \arg\min_{\Theta} \left(\hat{\Phi}^D - \frac{1}{S}\sum_{s=1}^{S} \hat{\Phi}_s^M\,(\Theta) \right)' \Omega \left(\hat{\Phi}^D - \frac{1}{S}\sum_{s=1}^{S} \hat{\Phi}_s^M\,(\Theta) \right), \tag{39}$$

where Θ is the vector of parameters of interest; $\hat{\Phi}^D$ is a set of empirical moments estimated from an empirical dataset; $\hat{\Phi}^M\,(\Theta)$ is the same set of simulated moments estimated from a simulated dataset based on the model; S is the number of simulation paths; and Ω is a positive definite weighting matrix. See Wu (2009) for the technical details on how to solve the minimal quadratic distance problem of (39), to draw optimal weighting matrix from the data and to calculate the numerical standard errors for the estimates.

Suppose that the empirical dataset is a panel with N firms and T years. We use the asymptotics of fixed T and large N. At the efficient choice for Ω^*, the SMM procedure provides a global specification test of the overidentifying restrictions of the model:

$$\begin{aligned} OI &= \frac{NS}{1+S}\left(\hat{\Phi}^D - \frac{1}{S}\sum_{s=1}^{S}\hat{\Phi}_s^M(\Theta)\right)'\Omega^*\left(\hat{\Phi}^D - \frac{1}{S}\sum_{s=1}^{S}\hat{\Phi}_s^M(\Theta)\right) \\ &\sim \chi^2\left[\dim\left(\hat{\Phi}\right) - \dim(\Theta)\right]. \end{aligned}$$

7.7 Robustness Tests

Table A.4 presents results for a set of robustness checks. Column (1) corresponds to the benchmark model, where $r = 0.20$ and $\rho = 0.90$. Columns (2) and (3) test the sensitivity to the discount factor by imposing $r = 0.15$ and 0.25, respectively. Columns (4) and (5) report the results with $\rho = 0.85$ and 0.95, respectively. Overall, we see only some modest variations in the estimates. In particular, the estimated σ_τ, ranging from 0.67 to 0.75, appears to be robust to the alternative choices of r and ρ.

[Insert Table A.4]

Column (6) increases the number of type in each dimension of heterogeneity from 3 to 5. The alternative simulation specification triples the estimation time but causes virtually no change in any of the estimates.

Columns (7) and (8) allow the long-run growth rate of $Z_{i,t}$, μ, and the level of uncertainty, σ, to be firm-specific. In Column (7), μ_i follows a normal distribution with mean μ and standard deviation σ_μ. In Column (8), σ_i follows a normal distribution with mean σ and standard deviation σ_σ. Introducing an additional dimension of heterogeneity involves an additional state variable. The estimation time increases by 2.5 times accordingly. Not surprisingly, allowing more heterogeneities improves the overall fitness. The estimate of σ_τ, however, is almost unaffected.

Column (9) replaces measurement errors in capital with measurement errors in investment. To be specific, $K_{i,t+1} = (1-\delta)(K_{i,t} + I_{i,t})$, where $I_{i,t} = I_{i,t}^{true}\exp\left(e_{i,t}^I\right)$ and $e_{i,t}^I \overset{i.i.d}{\sim} N(0, \sigma_{meI}^2)$. The alternative specification implies a persistent effect on the measurement of capital through capital accumulation. We find much larger capital adjustment costs. The estimated σ_τ, however, increases little, by less than 5 percent.

7.8 Specification Tests

To evaluate the importance of each of the three components (i.e., the unobserved heterogeneities, capital adjustment costs and measurement errors), Table A.5 reports specification tests for three restricted models. The full-blown model is taken as the benchmark, with estimation results listed in Column (1).

[Insert Table A.5]

Column (2) reports the results with homogeneous α_i and η_i – i.e., $\sigma_{\log \alpha} = \sigma_{\log \eta} = 0$. The estimated σ_τ increases from 0.706 to 0.924, implying a large bias by omitting the unobserved heterogeneities. Moreover, the model fails to match the data in a number of aspects, including all moments of the profit-revenue ratio (except for the mean) and the correlation between the revenue-capital and profit-revenue ratio. As a result, the overall fitness of the restricted model degenerates substantially.

Column (3) reports the results with no capital adjustment costs – i.e., $b^q = b^i = b^f = 0$. The estimate for σ_τ is just 7 percent lower than the benchmark result. For reasons discussed in the text, the unobserved heterogeneities can essentially be identified by the five core moments, on which capital adjustment costs have little impact. Nevertheless, without capital adjustment costs, the model cannot match some salient features, such as positive serial correlation, in the investment rate and revenue growth.

Column (4) reports the results with no measurement errors – i.e., $\sigma_{meK} = \sigma_{meY} = \sigma_{me\pi} = 0$. The estimate for σ_τ is, once again, very close to the benchmark result, with a difference of 3 percent. Like capital adjustment costs, measurement errors have only second-order effects on the between-group standard deviations. Consequently, the estimation of the unobserved heterogeneities is largely unaffected by measurement errors. Regarding the fitness, the restricted model generates too small within-group standard deviations of the profit-revenue and revenue-capital ratios and too much serial correlation in these two ratios.

Appendix for Tables

Table A.1: Illustration for Identification of Capital Adjustment Costs

Parameters	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	$\boldsymbol{b^q = 0.0}$	$\boldsymbol{b^q = 1.0}$	$b^q = 0.0$	$b^q = 0.0$	$b^q = 1.0$
	$\boldsymbol{b^i = 0.0}$	$b^i = 0.0$	$\boldsymbol{b^i = 0.1}$	$b^i = 0.0$	$b^i = 0.1$
	$\boldsymbol{b^f = 0.0}$	$b^f = 0.0$	$b^f = 0.0$	$\boldsymbol{b^f = 0.1}$	$b^f = 0.1$
Set of Moments					
mean(π/Y)	0.170	0.170	0.170	0.170	0.170
mean(log(Y/Khat))	0.837	0.865	0.780	0.852	1.006
mean(I/K)	0.202	0.111	0.137	0.187	0.112
mean(ΔlogY)	0.050	0.049	0.050	0.051	0.048
bsd(π/Y)	0.061	0.061	0.061	0.061	0.061
wsd(π/Y)	0.000	0.000	0.000	0.000	0.000
bsd(log(Y/Khat))	0.682	0.676	0.682	0.689	0.674
wsd(log(Y/Khat))	0.009	**0.131**	**0.096**	**0.124**	0.128
bsd(I/K)	0.233	0.080	0.170	0.266	0.098
wsd(I/K)	0.451	**0.065**	**0.242**	0.505	0.114
bsd(ΔlogY)	0.216	0.135	0.171	0.188	0.142
wsd(ΔlogY)	0.336	0.190	0.232	0.270	0.199
skew(π/Y)	0.176	0.176	0.176	0.176	0.176
skew(log(Y/Khat))	0.000	0.036	0.019	0.059	0.039
skew(I/K)	1.412	0.674	**2.712**	**3.380**	1.376
skew(dlogY)	0.000	0.037	0.431	0.774	0.200
scorr(π/Y)	1.000	1.000	1.000	1.000	1.000
scorr(log(Y/Khat))	1.000	0.967	0.980	0.967	0.968
scorr(I/K)	-0.058	**0.619**	**0.184**	-0.049	0.331
scorr(ΔlogY)	-0.067	0.009	0.043	0.001	0.013
bcorr(π/Y, log(Y/Khat))	-0.379	-0.392	-0.383	-0.376	-0.381

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 $\sigma_{meK} = \sigma_{meY} = \sigma_{me\pi} = 0$.

Table A.2: Illustration for Identification of Measurement Errors

Parameters	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	$\boldsymbol{\sigma_{meK}}$ **= 0.0**	$\boldsymbol{\sigma_{meK}}$ **= 0.5**	σ_{meK} = 0.0	σ_{meK} = 0.0	σ_{meK} = 0.5
	$\boldsymbol{\sigma_{meY}}$ **= 0.0**	σ_{meY} = 0.0	$\boldsymbol{\sigma_{meY}}$ **= 0.5**	σ_{meY} = 0.0	σ_{meY} = 0.5
	$\boldsymbol{\sigma_{me\pi}}$ **= 0.0**	$\sigma_{me\pi}$ = 0.0	$\sigma_{me\pi}$ = 0.0	$\boldsymbol{\sigma_{me\pi}}$ **= 0.5**	$\sigma_{me\pi}$ = 0.5
Set of Moments					
mean(π/Y)	0.170	0.170	0.193	0.170	0.193
mean(log(Y/Khat))	1.006	0.996	1.007	1.006	0.997
mean(I/K)	0.112	0.127	0.112	0.112	0.127
mean(ΔlogY)	0.048	0.048	0.048	0.048	0.048
bsd(π/Y)	0.061	0.061	0.088	0.066	0.094
wsd(π/Y)	0.000	0.000	**0.094**	**0.045**	0.111
bsd(log(Y/Khat))	0.674	0.711	0.720	0.674	0.755
wsd(log(Y/Khat))	0.128	**0.416**	**0.451**	0.128	0.600
bsd(I/K)	0.098	0.125	0.098	0.098	0.125
wsd(I/K)	0.114	**0.163**	0.114	0.114	0.163
bsd(ΔlogY)	0.142	0.142	0.276	0.142	0.276
wsd(ΔlogY)	0.199	0.199	**0.695**	0.199	0.695
skew(π/Y)	0.176	0.176	1.989	0.706	2.284
skew(log(Y/Khat))	0.039	-0.021	0.016	0.039	-0.016
skew(I/K)	1.376	2.871	1.376	1.376	2.871
skew(dlogY)	0.200	0.200	0.006	0.200	0.006
scorr(π/Y)	1.000	1.000	**0.288**	**0.575**	0.227
scorr(log(Y/Khat))	0.968	**0.669**	**0.634**	0.968	0.490
scorr(I/K)	0.331	0.229	0.331	0.331	0.229
scorr(ΔlogY)	0.013	0.013	**-0.445**	0.013	-0.445
bcorr(π/Y, log(Y/Khat))	-0.381	-0.359	-0.474	-0.351	-0.423

Note: The imposed parameter values are δ = 0.05, r = 0.15, $\mu_{log\alpha} = \mu_{log\eta}$ = –2.50, $\sigma_{log\alpha} = \sigma_{log\eta}$ = 0.50, ρ = 0.90, μ = 0, σ = 0.40, b^q =1.0, and $b^i = b^f$ = 0.1.

Table A.3: The 2004-2007 Balanced Panel for NBS Firms

Year	2004	2005	2006	2007
No. of firms	107579	107579	107579	107579
mean(π/Y)	0.155	0.159	0.157	0.160
mean(log(Y/Khat))	1.143	1.145	1.129	1.134
mean(I/K)	..	0.187	0.161	0.144
mean(ΔlogY)	..	0.109	0.083	0.097

Table A.4: Robustness Tests

Parameters	(1) benchmark	(2) r = 0.15	(3) r = 0.25	(4) ρ = 0.85	(5) ρ = 0.95
σ_τ	0.714	0.670	0.746	0.712	0.729
$\mu_{log\alpha}$	-2.606	-2.727	-2.496	-2.595	-2.602
$\sigma_{log\alpha}$	0.557	0.606	0.524	0.559	0.549
$\mu_{log\eta}$	-2.808	-2.672	-2.973	-2.826	-2.812
$\sigma_{log\eta}$	0.725	0.666	0.794	0.729	0.730
b^q	0.278	0.387	0.273	0.258	0.346
b^i	0.000	0.000	0.005	0.000	0.014
b^f	0.034	0.060	0.025	0.024	0.029
μ	0.080	0.080	0.083	0.081	0.085
σ	0.425	0.412	0.447	0.427	0.416
σ_{meK}	0.401	0.379	0.410	0.405	0.411
σ_{meY}	0.001	0.000	0.001	0.000	0.004
$\sigma_{me\pi}$	0.578	0.575	0.579	0.581	0.574
Moments					
mean(π/Y)	0.154	0.153	0.155	0.154	0.154
mean(log(Y/Khat))	1.146	1.127	1.165	1.143	1.159
mean(I/K)	0.173	0.170	0.177	0.174	0.179
mean(ΔlogY)	0.080	0.080	0.083	0.081	0.084
bsd(π/Y)	0.075	0.075	0.074	0.074	0.075
wsd(π/Y)	0.049	0.049	0.049	0.049	0.049
bsd(log(Y/Khat))	0.878	0.874	0.882	0.879	0.884
wsd(log(Y/Khat))	0.332	0.319	0.337	0.333	0.337
bsd(I/K)	0.164	0.153	0.170	0.155	0.178
wsd(I/K)	0.215	0.209	0.215	0.217	0.214
bsd(ΔlogY)	0.163	0.160	0.165	0.158	0.168
wsd(ΔlogY)	0.219	0.222	0.215	0.224	0.211
skew(π/Y)	0.854	0.856	0.846	0.855	0.857
skew(log(Y/Khat))	0.004	0.007	0.006	0.002	-0.002
skew(I/K)	2.251	2.320	2.206	2.168	2.181
skew(dlogY)	0.176	0.208	0.146	0.165	0.169
scorr(π/Y)	0.599	0.608	0.590	0.596	0.600
scorr(log(Y/Khat))	0.838	0.849	0.834	0.837	0.835
scorr(I/K)	0.243	0.200	0.274	0.202	0.297
scorr(ΔlogY)	0.053	0.026	0.073	0.015	0.099
bcorr(π/Y, log(Y/Khat))	-0.271	-0.280	-0.275	-0.270	-0.259
OI/100	183	208	179	215	157

Table A.4: Robustness Tests – Continued

Parameters	(1) benchmark	(6) type-5	(7) $\sigma_\mu > 0$	(8) $\sigma_\sigma > 0$	(9) $\sigma_{meI} > 0$
σ_τ	0.714	0.690	0.721	0.712	0.745
$\mu_{log\alpha}$	-2.606	-2.620	-2.604	-2.592	-2.654
$\sigma_{log\alpha}$	0.557	0.557	0.551	0.556	0.577
$\mu_{log\eta}$	-2.808	-2.851	-2.805	-2.805	-2.776
$\sigma_{log\eta}$	0.725	0.692	0.728	0.719	0.716
b^q	0.278	0.284	0.325	0.308	0.405
b^i	0.000	0.001	0.000	0.000	0.479
b^f	0.034	0.034	0.039	0.031	0.059
μ	0.080	0.082	0.083	0.080	0.061
σ	0.425	0.424	0.411	0.403	0.465
σ_{meK}	0.401	0.402	0.404	0.390	..
σ_{meY}	0.001	0.000	0.002	0.001	0.001
$\sigma_{me\pi}$	0.578	0.597	0.576	0.575	0.561
σ_μ	..	..	0.080	..	..
σ_σ	..	..	..	0.151	..
σ_{meI}	..	..	..	..	0.114
Moments					
mean(π/Y)	0.154	0.148	0.154	0.155	0.153
mean(log(Y/Khat))	1.146	1.155	1.154	1.147	1.104
mean(I/K)	0.173	0.175	0.177	0.171	0.135
mean(ΔlogY)	0.080	0.082	0.083	0.080	0.060
bsd(π/Y)	0.075	0.071	0.074	0.074	0.075
wsd(π/Y)	0.049	0.048	0.048	0.049	0.047
bsd(log(Y/Khat))	0.878	0.880	0.883	0.875	0.870
wsd(log(Y/Khat))	0.332	0.331	0.334	0.324	0.144
bsd(I/K)	0.164	0.165	0.180	0.161	0.135
wsd(I/K)	0.215	0.214	0.213	0.213	0.169
bsd(ΔlogY)	0.163	0.163	0.168	0.162	0.165
wsd(ΔlogY)	0.219	0.217	0.210	0.218	0.230
skew(π/Y)	0.854	1.010	0.852	0.846	0.846
skew(log(Y/Khat))	0.004	0.013	0.004	0.006	0.029
skew(I/K)	2.251	2.193	2.225	2.295	2.412
skew(dlogY)	0.176	0.176	0.195	0.157	0.268
scorr(π/Y)	0.599	0.581	0.604	0.598	0.620
scorr(log(Y/Khat))	0.838	0.839	0.838	0.844	0.976
scorr(I/K)	0.243	0.250	0.292	0.237	0.268
scorr(ΔlogY)	0.053	0.058	0.103	0.051	0.021
bcorr(π/Y, log(Y/Khat))	-0.271	-0.319	-0.263	-0.274	-0.299
OI/100	183	229	148	179	741

Table A.5: Specification Tests

	col (1)	col (2)	col (3)	col (4)
	benchmark	$\sigma_{log\alpha}=\sigma_{log\eta}=0$	$b^q=b^i=b^f=0$	$\sigma_{meK}=\sigma_{meY}=\sigma_{me\pi}=0$
Parameters				
σ_τ	0.714	0.924	0.665	0.734
$\mu_{log\alpha}$	-2.606	-2.351	-2.645	-2.742
$\sigma_{log\alpha}$	0.557	**0.000**	0.587	0.500
$\mu_{log\eta}$	-2.808	-2.494	-2.716	-2.998
$\sigma_{log\eta}$	0.725	**0.000**	0.660	0.885
b^q	0.278	0.443	**0.000**	0.163
b^i	0.000	0.000	**0.000**	0.476
b^f	0.034	0.082	**0.000**	0.041
μ	0.080	0.078	0.100	0.054
σ	0.425	0.354	0.205	0.443
σ_{meK}	0.401	0.380	0.420	**0.000**
σ_{meY}	0.001	0.123	0.110	**0.000**
$\sigma_{me\pi}$	0.578	0.816	0.541	**0.000**
Moments				
mean(π/Y)	0.154	0.171	0.155	0.141
mean(log(Y/Khat))	1.146	1.011	1.151	1.218
mean(I/K)	0.173	0.168	0.206	0.127
mean(ΔlogY)	0.080	0.078	0.100	0.053
bsd(π/Y)	0.075	0.042	0.073	0.071
wsd(π/Y)	0.049	0.073	0.047	0.000
bsd(log(Y/Khat))	0.878	0.848	0.872	0.851
wsd(log(Y/Khat))	0.332	0.328	0.343	0.137
bsd(I/K)	0.164	0.146	0.145	0.136
wsd(I/K)	0.215	0.218	0.274	0.177
bsd(ΔlogY)	0.163	0.153	0.123	0.160
wsd(ΔlogY)	0.219	0.254	0.227	0.221
skew(π/Y)	0.854	0.184	0.887	0.391
skew(log(Y/Khat))	0.004	0.008	0.011	0.013
skew(I/K)	2.251	2.220	1.586	2.450
skew(dlogY)	0.176	0.213	0.002	0.370
scorr(π/Y)	0.599	-0.001	0.604	1.000
scorr(log(Y/Khat))	0.838	0.830	0.822	0.977
scorr(I/K)	0.243	0.126	-0.047	0.242
scorr(ΔlogY)	0.053	-0.149	-0.223	0.027
bcorr(π/Y, log(Y/Khat))	-0.271	-0.019	-0.304	-0.208
OI/100	183	1510	653	3127