The Sources of Capital Misallocation*

Joel M. David† Venky Venkateswaran‡

USC NYU Stern

January 3, 2019

Abstract

We develop a methodology to disentangle sources of capital 'misallocation', i.e. dispersion in value-added/capital. It measures the contributions of technological/informational frictions and a rich class of firm-specific factors. An application to Chinese manufacturing firms reveals that adjustment costs and uncertainty, while significant, explain only a modest fraction of the dispersion, which stems largely from other factors: a component correlated with productivity and a fixed effect. Adjustment costs are more salient for large US firms, though other factors still account for bulk of the dispersion. Technological/markup heterogeneity explains a limited fraction in China, but a potentially large share in the US.

JEL Classifications: D24, E22, O11, O47

Keywords: Productivity, Misallocation, Distortions, Adjustment costs, Uncertainty

*A previous version of this paper was circulated under the title "Capital Misallocation: Frictions or Distortions?". We thank the editor and anonymous referees, Pete Klenow, Diego Restuccia, Francisco Buera, Daniel Xu, Richard Rogerson, Matthias Kehrig, Loukas Karabarbounis, Virgiliu Midrigan, Andy Atkeson, Hugo Hopenhayn and Russ Cooper for their helpful suggestions, Nicolas Petrosky-Nadeau and Oleg Itshkoki for insightful discussions, Junjie Xia for assistance with the Chinese data, and many seminar and conference participants for useful comments.

†joeldavi@usc.edu.

‡vvenkate@stern.nyu.edu.

1 Introduction

A large and growing body of work analyzes the 'misallocation' of productive resources across firms – usually measured by dispersion in the static average products of inputs (e.g., value-added/input ratios) – and the resulting adverse effects on aggregate productivity and output. A number of recent studies examine the role of specific factors hindering period-by-period equalization of input productivity ratios. Examples of such factors include adjustment costs, imperfect information, financial frictions, as well as firm-specific 'distortions' stemming from economic policies or other institutional features. The importance of disentangling the role of these forces is self-evident. For one, a central question, particularly from a policy standpoint, is whether observed variation in input products stems largely from efficient sources, e.g., technological factors like adjustment costs or heterogeneity in production technologies, or inefficient ones, such as policy-induced distortions or markups. Similarly, understanding the exact nature of distortions – e.g., the extent to which they are correlated with firm characteristics – is essential to analyze their implications beyond static misallocation, for example, on firm entry and exit decisions and investments that influence future productivity.[1]

In this paper, we develop and implement a tractable methodology to distinguish various sources of dispersion in average revenue products of capital (*arpk*) using observable data on value-added and inputs. Our analysis proceeds in two steps. First, we augment a standard general equilibrium model of firm dynamics with a number of forces that contribute to *ex-post* dispersion in the static *arpk*, specifically (i) capital adjustment costs, (ii) informational frictions, in the form of imperfect knowledge about firm-level fundamentals (e.g., productivity or demand) and (iii) other firm-specific factors, meant to capture all other forces influencing investment decisions, including unobserved heterogeneity in markups and/or production technologies, financial frictions, or institutional/policy-related distortions. In this first part of our analysis, rather than take a stand on the exact nature of these factors, we adopt a flexible specification that allows for time-variation and correlation with firm characteristics. The environment is an extension of the canonical Hsieh and Klenow (2009) framework to include dynamic considerations in firms' investment decisions. The main contribution of this part is an empirical strategy that precisely measures the contribution of each force to observed *arpk* dispersion using widely available firm-level data.

In the second part of our analysis, we explore various candidates for the firm-specific factors in (iii) above. First, we extend our methodology to investigate the extent to which the observed dispersion in *arpk* could stem from unobserved heterogeneity in markups and production technologies. Next, we analyze policies that restrict the size of firms and study a model

[1] See Restuccia and Rogerson (2017) for an in-depth discussion of these margins.

of financial/liquidity considerations. We show how these two forces can manifest themselves as firm-specific factors similar to those considered in the first part.

Our key innovation is to explore the sources of *arpk* dispersion within a unified framework and thus provide a more robust decomposition. In contrast, focusing on particular sources while abstracting from others – a common approach in the literature – is potentially problematic. When the data reflect the combined influence of a number of factors, examining them one-by-one can lead to biased assessments of their severity and contributions to the observed dispersion.

To understand the measurement difficulty, consider, as an example, convex adjustment costs. When they are the only force present, there is an intuitive, one-to-one mapping to a single moment, e.g., investment variability – the more severe the adjustment friction, the less volatile is investment. Now, suppose that there are other factors that also dampen investment volatility (e.g., a distortion correlated with productivity or size). In this case, using the variance of investment alone to draw inferences about adjustment costs leads to an upward bias. As a second example, consider the effects of firm-level uncertainty, which reduces the contemporaneous correlation between investment and productivity. However, a low correlation could also be the result of other firm-specific factors (e.g., markups) that are uncorrelated with productivity, making this single moment an inadequate measure of the quality of information.

Our strategy for disentangling these forces is based on a simple insight: although each moment is a complicated function of multiple factors, making any single one insufficient for identification, combining the information in a wider set of moments can be extremely helpful in disentangling these factors. Indeed, we show that allowing these forces to act in tandem is essential to reconcile a broad set of moments from the covariance matrix of firm-level investment and value-added. We formalize this intuition using a tractable special case – when firm-level productivity follows a random walk. In this case, we derive analytic expressions for the moments and prove that a set of four carefully chosen moments, namely, (1) the variance of investment, (2) the autocorrelation of investment, (3) the correlation of investment with past productivity, and (4) the covariance of *arpk* with productivity together uniquely identify adjustment costs, uncertainty and the magnitude and correlation structure of other firm-specific factors.

The intuition behind this result is easiest to see in a simple pairwise analysis. As an example, consider the challenge described earlier of disentangling adjustment costs from other idiosyncratic factors that dampen the response of investment to productivity. Both forces depress the variability of investment. However, they have opposing effects on its autocorrelation – convex adjustment costs create incentives to smooth investment over time and so increase its serial correlation. A distortion that reduces the responsiveness to productivity, on the other hand, raises the relative weight of other, more transitory considerations in the investment decision, lowering the serial correlation. Thus, holding all else fixed, these two moments allow

us to separate the two forces. Similar arguments can be developed for the remaining factors as well. In our quantitative work, where we depart from the polar random walk case, we demonstrate numerically that the same intuition carries through.

This logic also underlies the second part of our analysis, where we dig deeper into factors other than adjustment/information frictions. For example, we use moments of labor and materials usage to investigate the role of unobserved heterogeneity in markups and technologies (specifically, capital elasticities). Under the assumption that the choice of materials is distorted only by market power, markup dispersion is pinned down by the dispersion in materials' share of revenues. Technology dispersion can be bounded from above using the observed covariance between the average products of capital and labor. Intuitively, holding returns to scale fixed, a high production elasticity of capital implies a low labor elasticity, so this type of heterogeneity is a source of negative covariance between capital and labor products. The more positively correlated these are in the data, the lower the scope for *arpk* dispersion from this channel.

We apply our methodology to data on manufacturing firms in China over the period 1998-2009. We find that adjustment and informational frictions play economically significant roles in influencing observed investment dynamics. However, they account for only a relatively modest fraction of *arpk* dispersion among Chinese firms – about 1% and 10%, respectively – leading to losses in aggregate total factor productivity (TFP) of 1% and 8% (relative to the undistorted first-best). This implies that a substantial portion of *arpk* dispersion in China is due to other firm-specific factors. In particular, we find a large role for factors correlated with productivity and ones that are essentially permanent. These account for about 47% and 44% of overall *arpk* dispersion, respectively, leading to TFP losses of 38% and 36%.[2] These findings are driven in large part by two observations – first, firm-level investment is neither extremely volatile nor highly serially correlated. The latter bounds the potential for convex adjustment costs, which create incentives to smooth investment over time. In combination with the former, this leads us to ascribe a large role to correlated distortions, which reduce investment volatility without increasing the serial correlation. Importantly, as we discuss below, these insights continue to hold even when we introduce non-convexities in the adjustment cost function. Second, uncertainty over future productivity, while significant, is simply not large enough to account for the majority of *arpk* dispersion observed in the data.

We also apply the methodology to data on publicly traded firms in the US. Although the two sets of firms are not directly comparable, the US numbers serve as a useful benchmark to put our results for China in context.[3] As one would expect, the overall degree of *arpk* dispersion

[2]Our method also allows for distortions that are transitory and uncorrelated with firm characteristics. However, our estimation finds them to be negligible.

[3]We also report results for Chinese publicly traded firms as well as Colombian and Mexican manufacturing firms. The results regarding the role of various factors in driving observed dispersion are quite similar to our

is considerably smaller for publicly traded US firms. More interestingly, a larger share (about 11%) of the observed dispersion is accounted for by adjustment costs, which depress aggregate TFP by about 2%. Uncertainty plays a smaller role than among Chinese firms, as do other correlated factors – these account for about 7% and 14% of overall *arpk* dispersion, respectively, reducing aggregate TFP by 1% and 3%. However, even for these firms, firm-specific fixed factors, although considerably smaller in absolute magnitude than in China, generate a large share of the observed dispersion in *arpk*, accounting for about 65% of the total, with associated TFP losses of 13%. In other words, even in the US, factors other than technological and informational frictions play a significant role in determining capital allocations.

What are these firm-specific factors? First, we find modest scope for unobserved variation in markups or production technologies in China – together, they account for at most 27% of *arpk* dispersion (4% and 23%, respectively). Intuitively, we do not see much variation in materials' share of revenues in China, suggesting only small markup dispersion, and the average products of labor and capital are highly correlated, limiting the potential for heterogeneity in capital intensities. In contrast, for US publicly traded firms, variation in markups/technologies can explain as much as 58% of *arpk* dispersion (14% and 44%, respectively). These results suggest that unobserved heterogeneity is a promising explanation for much of the observed 'misallocation' in the US, but that the predominant drivers among Chinese firms lie elsewhere e.g., additional market frictions or institutional/policy-related distortions. For example, we show that our estimates of size/productivity-dependent factors could be picking up the effects of size-dependent government policies and certain forms of financial market imperfections. However, disentangling these two forces from other sources of correlated factors requires data beyond value-added and inputs (e.g., firm-level financial data).

We show that these patterns – in particular, the contributions of the various forces to observed *arpk* dispersion – are robust to a number of variations of our baseline setup. First, they are largely unchanged when we allow for non-convex adjustment costs. The main insight that underlies our baseline estimates emerges here as well – examining moments in isolation can paint a distorted picture of the forces driving investment dynamics. For example, the low serial correlation of investment in the data by itself might seem to indicate large non-convexities, but high fixed costs of adjustment also tend to make firm-level investment more volatile and induce substantial 'inaction' (i.e., periods with little or no investment), patterns which are inconsistent with the data. Models with only adjustment costs, even those with sophisticated specifications, struggle to simultaneously match these patterns and can produce very different estimates depending on the choice of target moments. For example, targeting investment variability and inaction (and ignoring the serial correlation, as in, e.g., Asker et al.

baseline findings for Chinese manufacturers.

(2014)) results in much larger estimates for convex costs relative to a strategy of targeting the serial correlation (and ignoring variability/inaction, as in, e.g., Cooper and Haltiwanger (2006)), which instead produces substantial fixed costs. This underscores the value of a strategy like ours, which targets a broad set of moments with a more flexible model. Our estimation reconciles these seemingly inconsistent data patterns by ascribing an important role to other firm-specific factors, particularly when it comes to generating *arpk* dispersion.

Next, we show that distortions in the labor choice do not alter our main conclusions. A version of our model in which labor is subject to the same frictions and distortions as capital leads to a very similar decomposition of *arpk* dispersion. However, since both inputs are affected by each of the forces, the associated implications for aggregate TFP are much larger. Lastly, we show that our main results remain valid across a number of robustness exercises aimed at addressing measurement-related issues, sectoral heterogeneity and parameter choices.

The paper is organized as follows. Section 2 describes our model of frictional investment. Section 3 spells out our identification strategy using the analytically tractable random walk case, while Section 4 details our numerical analysis and presents our quantitative results. Section 5 further investigates the potential sources of firm-specific idiosyncratic factors. Section 6 explores a number of variants on our baseline approach. We summarize our findings and discuss directions for future research in Section 7.

Related literature. Our paper relates to several branches of literature. We bear a direct connection to the large body of work measuring and quantifying the effects of resource misallocation.[4] Following the seminal work of Hsieh and Klenow (2009) and Restuccia and Rogerson (2008), recent attention has shifted toward analyzing the roles of specific factors. Important contributions include Asker et al. (2014) on adjustment costs, Buera et al. (2011), Moll (2014), Gopinath et al. (2017) and Midrigan and Xu (2014) on financial frictions, David et al. (2016) on uncertainty and Peters (2016) on markup dispersion. Several recent papers analyze subsets of these factors in combination. For example, Gopinath et al. (2017) study the interaction of capital adjustment costs and size-dependent financial frictions in determining the recent dynamics of capital allocation in Spain. Kehrig and Vincent (2017) combine financial and adjustment frictions to investigate misallocation within firms, while Song and Wu (2015) estimate a model with adjustment costs, permanent distortions and heterogeneity in markups/technologies.

Our primary contribution is to develop a unified framework that encompasses many of these factors and devise an empirical strategy based on observable firm-level data to disentangle them. Our results, both analytical and quantitative, highlight the importance of studying a broad set of forces in tandem. This breadth is partly what distinguishes us from the work of Song and

[4]Restuccia and Rogerson (2017) and Hopenhayn (2014) provide recent overviews of this line of work.

Wu (2015), who abstract from time-variation in firm-level distortions (as well as in firm-specific markups/technologies), ruling out, by assumption, any role for so-called 'correlated' or size-dependent distortions.[5] Many papers in the literature – e.g., Restuccia and Rogerson (2008), Bartelsman et al. (2013), Hsieh and Klenow (2014) and Bento and Restuccia (2017) – emphasize the need to distinguish such factors from those that are orthogonal to productivity. This message is reinforced by our quantitative findings, which reveal a significant role for correlated factors (in addition to uncorrelated, permanent ones), particularly in developing countries such as China. Our modeling of these factors as implicit taxes correlated with productivity follows the approach taken by, e.g., Restuccia and Rogerson (2008), Guner et al. (2008), Bartelsman et al. (2013), Buera et al. (2013), Buera and Fattal-Jaef (2016) and Hsieh and Klenow (2014).

Our methodology and findings also have relevance beyond the misallocation context, notably, for studies of adjustment and informational frictions. A large literature has examined the implications of adjustment costs, examples of which include Cooper and Haltiwanger (2006), Khan and Thomas (2008) and Bloom (2009). Our analysis shows that accounting for other firm-specific factors acting on firms' investment decisions is potentially crucial in order to accurately estimate the severity of these frictions and reconcile a broader set of micro-level moments. A similar point applies to recent work on quantifying firm-level uncertainty, for example, Bloom (2009), Bachmann and Elstner (2015) and Jurado et al. (2015).

2 The Model

We consider a discrete time, infinite-horizon economy, populated by a representative household. The household inelastically supplies a fixed quantity of labor N and has preferences over consumption of a final good. The household discounts time at rate β. The household side of the economy is deliberately kept simple as it plays a limited role in our study. Throughout the analysis, we focus on a stationary equilibrium in which all aggregate variables remain constant.

Production. A continuum of firms of fixed measure one, indexed by i, produce intermediate goods using capital and labor according to

$$Y_{it} = K_{it}^{\hat{\alpha}_1} N_{it}^{\hat{\alpha}_2}, \quad \hat{\alpha}_1 + \hat{\alpha}_2 \leq 1 \,. \tag{1}$$

[5] We also differ from Song and Wu (2015) in our explicit modeling (and measurement) of information frictions and in our approach to quantifying heterogeneity in markups/technologies.

These intermediate goods are bundled to produce the single final good using a standard CES aggregator

$$Y_t = \left(\int \hat{A}_{it} Y_{it}^{\frac{\theta-1}{\theta}} di \right)^{\frac{\theta}{\theta-1}} ,$$

where $\theta \in (1, \infty)$ is the elasticity of substitution between intermediate goods and $\hat{A}_{it}$ represents a firm-specific idiosyncratic component in production/demand. This is the only source of fundamental uncertainty in the economy (i.e., we abstract from aggregate risk).

Market structure and revenue. The final good is produced frictionlessly by a representative competitive firm. This yields a standard demand function for intermediate good i:

$$Y_{it} = P_{it}^{-\theta} \hat{A}_{it}^{\theta} Y_t \quad \Rightarrow \quad P_{it} = \left(\frac{Y_{it}}{Y_t} \right)^{-\frac{1}{\theta}} \hat{A}_{it} ,$$

where P_{it} denotes the relative price of good i in terms of the final good, which serves as numeraire. Revenues (here, equal to value-added) for firm i at time t are

$$P_{it} Y_{it} = Y_t^{\frac{1}{\theta}} \hat{A}_{it} K_{it}^{\alpha_1} N_{it}^{\alpha_2} ,$$

where

$$\alpha_j = \left(1 - \frac{1}{\theta} \right) \hat{\alpha}_j, \quad j = 1, 2 .$$

This framework accommodates two alternative interpretations of the idiosyncratic component, $\hat{A}_{it}$: as a firm-specific shifter of either quality/demand or productive efficiency.

Input choices. In our baseline analysis, we assume that firms hire labor period-by-period under full information and in an distorted fashion at a competitive wage, W_t.[6] At the end of each period, firms choose capital for the following period. Investment is subject to quadratic adjustment costs, given by

$$\Phi \left(K_{it+1}, K_{it} \right) = \frac{\hat{\xi}}{2} \left(\frac{K_{it+1}}{K_{it}} - (1 - \delta) \right)^2 K_{it} , \tag{2}$$

[6]We relax this assumption later in the paper in two separate exercises. First, in Appendix E.1, we introduce labor market distortions in the form of firm-specific 'taxes' and show that this formulation changes the interpretation of our measure of productivity, but does not affect our identification strategy or conclusions about the sources of *arpk* dispersion. Second, in Section 6.2, we subject the labor choice to all the frictions that distort investment – adjustment costs, informational frictions and other distortionary factors. This setup also leads to a very similar problem with suitably re-defined productivity and curvature parameters.

where $\hat{\xi}$ parameterizes the severity of the adjustment cost and δ is the rate of depreciation.[7]

Investment decisions are likely to be affected by a number of additional factors (other than productivity/demand and the level of installed capital). These could originate from distortionary government policies (e.g., taxes, size restrictions or regulations, or other features of the institutional environment), from other market frictions that are not explicitly modeled (e.g., financial frictions) or from un-modeled heterogeneity in markups/production technologies. For now, we do not take a stand on the precise nature of these factors and, following, e.g., Hsieh and Klenow (2009), model them as firm-specific proportional 'taxes' on the flow cost of capital. We denote these by T^K_{it+1} and, in a slight abuse of terminology, refer to them as 'distortions' or wedges throughout the paper, even though they may partly reflect efficient factors (for example, heterogeneity in production functions).[8] In Section 5, we demonstrate how progress can be made in further disentangling some of these sources.

The firm's problem in a stationary equilibrium can be represented in recursive form as (we suppress the time subscript on all aggregate variables)

$$\begin{aligned}\mathcal{V}\left(K_{it},\mathcal{I}_{it}\right) = \max_{N_{it},K_{it+1}} \quad & \mathbb{E}_{it}\left[Y^{\frac{1}{\theta}}\hat{A}_{it}K_{it}^{\alpha_1}N_{it}^{\alpha_2} - WN_{it} - T^K_{it+1}K_{it+1}\left(1-\beta\left(1-\delta\right)\right) - \Phi\left(K_{it+1},K_{it}\right)\right] \\ + \quad & \beta\mathbb{E}_{it}\left[\mathcal{V}\left(K_{it+1},\mathcal{I}_{it+1}\right)\right] ,\end{aligned}$$

where $\mathbb{E}_{it}\left[\cdot\right]$ denotes expectations conditional on $\mathcal{I}_{it}$, the firm's information set at the time it chooses K_{it+1} (described in more detail below). Note that the wedge T^K_{it+1} (which applies to $1-\beta(1-\delta)$, the per-unit user cost of capital) distorts both the capital decision and the capital-labor ratio. In other words, it is both a 'scale' and 'mix' distortion.[9]

After maximizing over N_{it}, this becomes

$$\begin{aligned}\mathcal{V}\left(K_{it},\mathcal{I}_{it}\right) = \max_{K_{it+1}} \quad & \mathbb{E}_{it}\left[GA_{it}K_{it}^{\alpha} - T^K_{it+1}K_{it+1}\left(1-\beta\left(1-\delta\right)\right) - \Phi\left(K_{it+1},K_{it}\right)\right] \\ + \quad & \mathbb{E}_{it}\beta\left[\mathcal{V}\left(K_{it+1},\mathcal{I}_{it+1}\right)\right] ,\end{aligned} \tag{3}$$

where $\alpha \equiv \frac{\alpha_1}{1-\alpha_2}$ is the curvature of operating profits (value-added less labor expenses) and $A_{it} \equiv \hat{A}_{it}^{\frac{1}{1-\alpha_2}}$ is the firm-specific profitability of capital. In a slight abuse of terminology, we refer to A_{it} simply as firm-specific productivity. The term $G \equiv \left(1-\alpha_2\right)\left(\frac{\alpha_2}{W}\right)^{\frac{\alpha_2}{1-\alpha_2}} Y^{\frac{1}{\theta}\frac{1}{1-\alpha_2}}$ is a constant that captures the effects of aggregate variables.

[7] We generalize this specification to include non-convex costs in Section 6.1 and show that our main quantitative results continue to hold.

[8] The timing convention implies that the wedge T^K_{it+1} affects the firm's choice of K_{it+1}.

[9] In Section 6.2, when the firm's labor choice is assumed to be subject to the same frictions/distortions as its investment decision, the wedge is a pure scale distortion, i.e., it does not distort the capital-labor ratio.

Equilibrium. We can now define a *stationary equilibrium* in this economy as (i) a set of value and policy functions for the firm, $\mathcal{V}(K_{it}, \mathcal{I}_{it})$, $N_{it}(K_{it}, \mathcal{I}_{it})$ and $K_{it+1}(K_{it}, \mathcal{I}_{it})$, (ii) a wage W and (iii) a joint distribution over $(K_{it}, \mathcal{I}_{it})$ such that (a) taking as given wages and the law of motion for $\mathcal{I}_{it}$, the value and policy functions solve the firm's optimization problem, (b) the labor market clears and (c) the joint distribution remains constant through time.

Characterization. We use perturbation methods to solve the model.[10] In particular, we log-linearize the firm's optimality conditions and laws of motion around $A_{it} = \bar{A}$ (the unconditional average level of productivity) and $T_{it}^K = 1$ (i.e., no distortions). Appendix A.1 derives the following log-linearized Euler equation:[11]

$$k_{it+1}\left((1+\beta)\xi + 1 - \alpha\right) = \mathbb{E}_{it}\left[a_{it+1} + \tau_{it+1}\right] + \beta\xi\mathbb{E}_{it}\left[k_{it+2}\right] + \xi k_{it} \ , \tag{4}$$

where ξ and τ_{it+1} are re-scaled versions of the adjustment cost parameter, $\hat{\xi}$, and the distortion, $\log T_{it+1}^K$, respectively.

Stochastic processes. We assume that the productivity, A_{it}, follows an AR(1) process in logs with normally distributed i.i.d. innovations, i.e.,

$$a_{it} = \rho a_{it-1} + \mu_{it}, \quad \mu_{it} \sim \mathcal{N}\left(0, \sigma_\mu^2\right) \ , \tag{5}$$

where ρ is the persistence and σ_μ^2 the variance of the innovations.[12]

For the distortion, τ_{it}, we adopt a specification that allows for a rich correlation structure, both over time as well as with firm-level productivity. Specifically, τ_{it} takes the form:

$$\tau_{it} = \gamma a_{it} + \varepsilon_{it} + \chi_i, \quad \varepsilon_{it} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \quad \chi_i \sim \mathcal{N}\left(0, \sigma_\chi^2\right) \ , \tag{6}$$

where the parameter γ controls the extent to which τ_{it} co-moves with productivity. If $\gamma < 0$, the distortion discourages (encourages) investment by firms with higher (lower) productivity – arguably, the empirically relevant case. The opposite is true if $\gamma > 0$. The uncorrelated component of τ_{it} has both permanent and iid (over time) components, denoted χ_i and ε_{it}, respectively. Thus, the severity of these factors is summarized by 3 parameters: $(\gamma, \sigma_\varepsilon^2, \sigma_\chi^2)$.[13]

[10] The results in Section 6.1, where we solve a version of the model with non-convexities without linearization, suggest that the perturbation yields reasonably accurate estimates.

[11] We use lower-case to denote natural logs, a convention we follow throughout, so that, e.g., $x_{it} = \log X_{it}$.

[12] Appendix I.2 extends the analysis to allow for firm fixed-effects in the process for a_{it}. This has no effect on our analytical results in Section 3 (where we work exclusively with growth rates) and Section 5. Our quantitative results regarding the sources of *arpk* dispersion are very similar with these effects.

[13] Appendix I.2 considers a more flexible process for τ_{it}. Our results there confirm the highly persistent nature

Information. Next, we spell out $\mathcal{I}_{it}$, the information set of the firm at the time of choosing K_{it+1}. This includes the entire history of productivity realizations through period t, i.e., $\{a_{it-s}\}_{s=0}^{\infty}$. Given the AR(1) assumption, this can be summarized by the most recent observation, namely, a_{it}. The firm also observes a noisy signal of the following period's innovation:

$$s_{it+1} = \mu_{it+1} + e_{it+1}, \qquad e_{it+1} \sim \mathcal{N}\left(0, \sigma_e^2\right) \ ,$$

where e_{it+1} is an i.i.d., mean-zero and normally distributed noise term. This is in essence an idiosyncratic 'news shock,' since it contains information about future productivity. Finally, firms also perfectly observe the uncorrelated transitory component of distortions, ε_{it+1}, at the time of choosing period t investment (as noted above, the distortion is indexed by the date it influences the firm's capital choice, so that, e.g., ε_{it+1} is in the firm's information set at date t and affects its choice of k_{it+1}). Firms also observe the fixed component of the distortion, χ_i. They do not see the correlated component, but are aware of its structure, i.e., they know γ.

Thus, the firm's information set is given by $\mathcal{I}_{it} = (a_{it}, s_{it+1}, \varepsilon_{it+1}, \chi_i)$. Direct application of Bayes' rule yields the conditional expectation of productivity, a_{it+1}:

$$\begin{aligned} a_{it+1}|\mathcal{I}_{it} \quad &\sim \quad N\left(\mathbb{E}_{it}\left[a_{it+1}\right], \mathbb{V}\right) \qquad \text{where} \\ \mathbb{E}_{it}\left[a_{it+1}\right] \quad &= \quad \rho a_{it} + \frac{\mathbb{V}}{\sigma_e^2} s_{it+1}, \quad \mathbb{V} = \left(\frac{1}{\sigma_\mu^2} + \frac{1}{\sigma_e^2}\right)^{-1} \ . \end{aligned}$$

There is a one-to-one mapping between the posterior variance $\mathbb{V}$ and the noisiness of the signal, σ_e^2 (given the volatility of productivity, σ_μ^2). In the absence of 'news', i.e., $\sigma_e^2 = \infty$, we have $\mathbb{V} = \sigma_\mu^2$, that is, posterior uncertainty is simply the variance of the innovation. This corresponds to a standard one period time-to-build assumption with $\mathbb{E}_{it}\left[a_{it+1}\right] = \rho a_{it}$. At the other extreme, $\sigma_e^2 = 0$ implies $\mathbb{V} = 0$, so the firm is perfectly informed about a_{it+1}. It turns out to be more convenient to work directly with the posterior variance, $\mathbb{V}$, as the measure of uncertainty.

Law of motion. In Appendix A.1, we solve the Euler equation in (4) to obtain:

$$k_{it+1} \quad = \quad \psi_1 k_{it} + \psi_2 \left(1+\gamma\right) \mathbb{E}_{it}\left[a_{it+1}\right] + \psi_3 \varepsilon_{it+1} + \psi_4 \chi_i \ , \tag{7}$$

$$\begin{aligned} \text{where} \qquad \xi\left(\beta\psi_1^2 + 1\right) \quad &= \quad \psi_1\left((1+\beta)\xi + 1 - \alpha\right) \\ \psi_2 \quad &= \quad \frac{\psi_1}{\xi\left(1 - \beta\rho\psi_1\right)}, \qquad \psi_3 = \frac{\psi_1}{\xi}, \qquad \psi_4 = \frac{1-\psi_1}{1-\alpha} \ . \end{aligned} \tag{8}$$

of the uncorrelated component, suggesting that the simpler specification here with fixed and iid elements largely captures the time-series properties of the distortion.

The coefficients ψ_1–ψ_4 depend only on production (and preference) parameters, including the adjustment cost, and are independent of assumptions about information and distortions. The coefficient ψ_1 is increasing and ψ_2-ψ_4 decreasing in the severity of adjustment costs, ξ. If there are no adjustment costs (i.e., $\xi = 0$) , $\psi_1 = 0$ and $\psi_2 = \psi_3 = \psi_4 = \frac{1}{1-\alpha}$. At the other extreme, as ξ tends to infinity, $\psi_1 \to 1$ and ψ_2-ψ_4 vanish. Intuitively, as adjustment costs become large, the firm's choice of capital becomes more autocorrelated and less responsive to productivity and distortions. Our empirical strategy essentially identifies the coefficients in the policy function, ψ_1 and $\psi_2 (1+\gamma)$, from observable moments. Given values of α and β, the estimate of ψ_1 pins down ξ from (8). Given ξ, β and ρ, we can use the estimate of $\psi_2 (1+\gamma)$ to recover γ.

Aggregation. In Appendix A.2, we show that aggregate output can be expressed as

$$\log Y \equiv y = a + \hat{\alpha}_1 k + \hat{\alpha}_2 n \ ,$$

where k and n denote the (logs of the) aggregate capital stock and labor inputs, respectively. Aggregate TFP, denoted by a, is given by

$$a = a^* - \frac{(\theta\hat{\alpha}_1 + \hat{\alpha}_2)\,\hat{\alpha}_1}{2}\sigma^2_{arpk} \qquad \frac{da}{d\sigma^2_{arpk}} = -\frac{(\theta\hat{\alpha}_1 + \hat{\alpha}_2)\,\hat{\alpha}_1}{2} \ , \tag{9}$$

where a^* is aggregate TFP if static capital products ($arpk_{it}$) are equalized across firms and σ^2_{arpk} is the cross-sectional dispersion in (the log of) the static average product of capital ($arpk_{it} = p_{it}y_{it} - k_{it}$). Thus, aggregate TFP monotonically decreases in the extent of dispersion in capital productivities, summarized in this log-normal world by σ^2_{arpk}. The effect of σ^2_{arpk} on aggregate TFP depends on the elasticity of substitution, θ, and the relative shares of capital and labor in production. The higher is θ, that is, the closer we are to perfect substitutability, the more severe the losses from dispersion in capital products. Similarly, fixing the degree of overall returns to scale in production, for a larger capital share, $\hat{\alpha}_1$, a given degree of dispersion has larger effects on aggregate outcomes.[14]

In our framework, a number of forces – adjustment costs, information frictions, and distortions – will lead to *arpk* dispersion. Once we quantify their contributions to σ^2_{arpk}, equation (9) allows us to directly map those contributions to their aggregate implications.

Measuring the contribution of each factor is a challenging task, since all the data moments confound all the factors (i.e., each moment reflects the influence of more than one factor). As a result, there is no one-to-one mapping between individual moments and parameters – to accurately identify the contribution of any factor, we need to explicitly control for the others.

[14] Aggregate output effects are larger than TFP losses by a factor $\frac{1}{1-\hat{\alpha}_1}$. This is because dispersion in capital products also reduces the incentives for capital accumulation and therefore, the steady-state capital stock.

In the following section, we overcome this challenge by exploiting the fact that these forces have different implications for different moments.

3 Identification

In this section, we lay out our identification strategy – specifically, we provide a methodology to tease out the role of adjustment costs, informational frictions and other factors using observable moments of firm-level value-added and investment. We use a tractable special case – when productivity follow a random walk, i.e., $\rho = 1$ – to derive analytic expressions for key moments, allowing us to prove our identification result formally and make clear the underlying intuition. When we return to the more general model (with $\rho < 1$) in the following section, we will demonstrate numerically that this intuition applies there as well.

We assume that the preference and technology parameters – the discount factor, β, the curvature of the profit function, α, and the depreciation rate, δ – are known to the econometrician (e.g., calibrated using aggregate or industry-level data). The remaining parameters of interest are the costs of capital adjustment, ξ, the quality of firm-level information (summarized by $\mathbb{V}$), and the severity of distortions, parameterized by γ, σ_ε^2 and σ_χ^2.

Our methodology uses a set of carefully chosen elements from the covariance matrix of firm-level capital and productivity (the latter can be measured using data on value-added and capital, along with an assumed curvature, α). Note that $\rho = 1$ implies non-stationarity in levels, so we work with moments of (log) changes. This means that we cannot identify σ_χ^2, the variance of the fixed component.[15] Here, we focus on the four remaining parameters, namely ξ, γ, $\mathbb{V}$ and σ_ε^2. Our main result is to show that these are exactly identified by the following four moments: (1) the autocorrelation of investment, denoted $\rho_{k,k_{-1}}$, (2) the variance of investment, σ_k^2, (3) the correlation of period t investment with the innovation in productivity in period $t-1$, denoted $\rho_{k,a_{-1}}$ and (4) the coefficient from a regression of $\Delta arpk_{it}$ on Δa_{it}, denoted $\lambda_{arpk,a}$.

Several of these moments have been used in the literature to quantify the various factors in isolation. For example, $\rho_{k,k_{-1}}$ and σ_k^2 are standard targets in the literature on adjustment costs – see, e.g., Cooper and Haltiwanger (2006) and Asker et al. (2014). The responsiveness to lagged fundamentals, $\rho_{k,a_{-1}}$, is used by Klenow and Willis (2007) to quantify information frictions in a price-setting context. The covariance of *arpk* with productivity – which we proxy with $\lambda_{arpk,a}$ – is highlighted in the misallocation literature as suggestive of correlated distortions, e.g., Bartelsman et al. (2013) and Buera and Fattal-Jaef (2016). The tractable random walk special case will shed light on the value of analyzing these moments/factors in tandem (and the

[15] For our numerical analysis in Section 4, we use a stationary model (i.e., with $\rho < 1$) and use σ_{arpk}^2, a moment computed using levels of capital and productivity, to pin down σ_χ^2.

potential biases from doing so in isolation).

Our main result is stated formally in the following proposition:

Proposition 1. *The parameters ξ, γ, $\mathbb{V}$ and σ_ε^2 are uniquely identified by the moments $\rho_{k,k_{-1}}$, σ_k^2, $\rho_{k,a_{-1}}$ and $\lambda_{arpk,a}$.*

3.1 Intuition

The proof of Proposition 1 (in Appendix A.3) involves tedious, if straightforward, algebra. Here, we provide a more heuristic argument for the intuition behind the result. We do this by analyzing parameters in pairs and showing that they can be uniquely identified by a pair of moments, holding the other parameters fixed. To be clear, this is a local argument – our goal here is simply to provide intuition about how the different moments can be combined to disentangle the different forces. The identification result in Proposition 1 is a global one and shows that there is a unique mapping from the four moments to the four parameters.

Adjustment costs and correlated distortions. We begin with adjustment costs, parameterized by ξ, and correlated distortions, γ. The relevant moment pair is the variance and autocorrelation of investment, σ_k^2 and $\rho_{k,k_{-1}}$. Both of these moments are commonly used to estimate quadratic adjustment costs – for example, Asker et al. (2014) target the former and Cooper and Haltiwanger (2006) (among other moments), the latter. In our setting, these moments are given by:

$$\sigma_k^2 = \left(\frac{\psi_2^2}{1-\psi_1^2}\right)(1+\gamma)^2\sigma_\mu^2 + \frac{2\psi_3^2}{1+\psi_1}\sigma_\varepsilon^2 \tag{10}$$

$$\rho_{k,k_{-1}} = \psi_1 - \psi_3^2\frac{\sigma_\varepsilon^2}{\sigma_k^2}\,, \tag{11}$$

where $\psi_1 - \psi_3$ are as defined in equation (8). Our argument exploits the fact that the two forces have similar effects on the variability of investment, but opposing effects on the autocorrelation. To see this, recall that ψ_1 is increasing and ψ_2 and ψ_3 decreasing in adjustment costs, but all three are independent of γ. Thus, holding all other parameters fixed, σ_k^2 is decreasing in both the severity of adjustment costs (higher ξ) and correlated factors (more negative γ).[16] The autocorrelation, $\rho_{k,k_{-1}}$, on the other hand, increases with ξ but decreases as γ becomes more negative (through its effect on σ_k^2). Intuitively, while both forces dampen the volatility of investment, they do so for different reasons – adjustment costs make it optimal to smooth

[16]The latter is true only for $\gamma > -1$, which is the empirically relevant region.

investment over time (increasing its autocorrelation) while correlated factors reduce sensitivity to the serially correlated productivity process (reducing the autocorrelation of investment).

The top left panel of Figure 1 shows how these properties help identify the two parameters. The panel plots a pair of 'isomoment' curves: each curve traces out combinations of the two parameters that give rise to a given value of the relevant moment, holding the other parameters fixed. Take the σ_k^2 curve: it slopes upward because higher ξ and lower γ have similar effects on σ_k^2 – if γ is relatively small (in absolute value), adjustment costs must be high in order to maintain a given level of σ_k^2. Conversely, a low ξ is consistent with a given value of σ_k^2 only if γ is very negative. An analogous argument applies to the $\rho_{k,k_{-1}}$ isomoment curve: since higher ξ and more negative γ have opposite effects on $\rho_{k,k_{-1}}$, the curve slopes downward. As a result, the two curves cross only once, yielding the unique combination of the parameters that is consistent with both moments. By plotting curves corresponding to the empirical values of these moments, we can uniquely pin down the pair (ξ, γ) (holding all other parameters fixed).

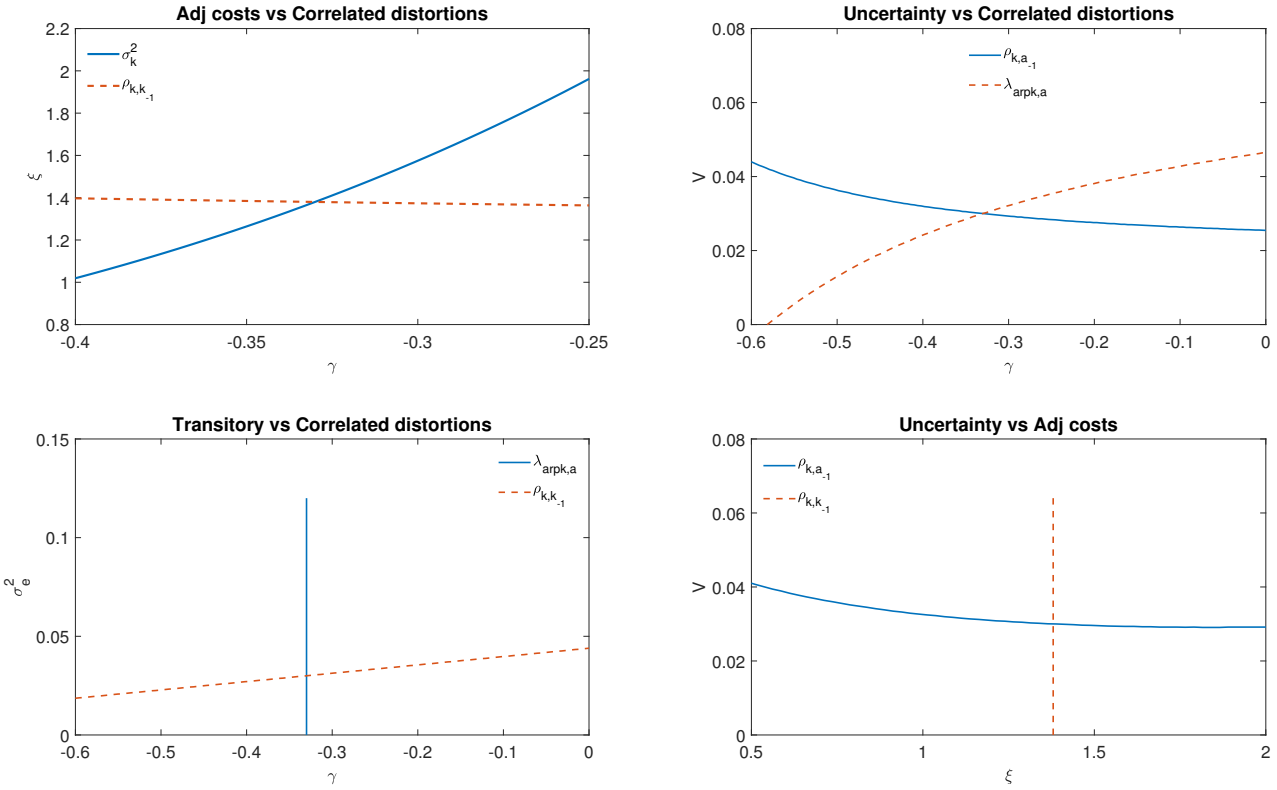


Figure 1: Pairwise Identification - Isomoment Curves

The graph also illustrates the potential bias introduced when examining these forces in isolation. For example, estimating adjustment costs while ignoring correlated distortions (i.e., imposing $\gamma = 0$) puts the estimate on the very right-hand side of the horizontal axis. The estimate for ξ can be read off the vertical height of the isomoment curve corresponding to the

targeted moment. Because the σ_k^2 curve is upward sloping, targeting this moment alone leads to an overestimate of adjustment costs (at the very right of the horizontal axis, the curve is above the point of intersection, which corresponds to the true value of the parameters).[17] Targeting $\rho_{k,k_{-1}}$ alone leads to a bias in the opposite direction – since the $\rho_{k,k_{-1}}$ curve is downward sloping, imposing $\gamma = 0$ yields an underestimate of adjustment costs.

The remaining panels in Figure 1 repeat this analysis for other combinations of parameters. Each relies on the same logic as shown in the top left panel.

Uncertainty and correlated distortions. To disentangle information frictions from correlated factors (the top right panel), we use the correlation of investment with past innovations in productivity, $\rho_{k,a_{-1}}$, and the regression coefficient $\lambda_{arpk,a}$. These moments can be written as:

$$\rho_{k,a_{-1}} = \left[\frac{\mathbb{V}}{\sigma_\mu^2}(1-\psi_1)+\psi_1\right]\frac{\sigma_\mu \psi_2 (1+\gamma)}{\sigma_k} \tag{12}$$

$$\lambda_{arpk,a} = 1-(1-\alpha)(1+\gamma)\psi_2\left(1-\frac{\mathbb{V}}{\sigma_\mu^2}\right) . \tag{13}$$

A higher $\mathbb{V}$ implies a higher correlation of investment with lagged productivity innovations. Intuitively, the more uncertain is the firm, the greater the tendency for its actions to reflect productivity with a 1-period lag. In contrast, a higher (more negative) γ increases the relative importance of transitory factors in the firm's investment decision, reducing its correlation with productivity. Therefore, to maintain a given level of $\rho_{k,a_{-1}}$, a decrease in $\mathbb{V}$ must be accompanied by a less negative γ, i.e., the isomoment curve slopes downward. On the other hand, higher uncertainty and a more negative gamma both cause *arpk* to covary more positively with contemporaneous productivity, a, leading to an upward sloping $\lambda_{arpk,a}$ curve. Together, these two curves pin down $\mathbb{V}$ and γ, holding other parameters fixed.

As before, the graph also reveals the direction of bias when estimating these factors in isolation. Assuming full information ($\mathbb{V} = 0$) and using $\lambda_{arpk,a}$ to discipline the strength of correlated distortions – e.g., as in Bartelsman et al. (2013) and Buera and Fattal-Jaef (2016) – overstates their importance. Using the lagged responsiveness to productivity to discipline information frictions while abstracting from correlated factors understates uncertainty.

Transitory and correlated distortions. To disentangle correlated and uncorrelated factors, consider $\lambda_{arpk,a}$ and $\rho_{k,k_{-1}}$. The former is increasing in the severity of correlated distortions, but independent of transitory ones, implying a vertical isomoment curve. The latter is decreasing in both types of distortions – a more negative γ and higher σ_ε^2 both increase the importance

[17] This approach would also predict a counterfactually high level of the autocorrelation of investment.

of the transitory determinants of investment, yielding an upward sloping isomoment curve.

Uncertainty and adjustment costs. Finally, the bottom right panel shows the intuition for disentangling uncertainty from adjustment costs. An increase in the severity of either of these factors contributes to sluggishness in the response of actions to productivity, i.e., raises the correlation of investment with past productivity shocks $\rho_{k,a_{-1}}$. However, the autocorrelation of investment $\rho_{k,k_{-1}}$ is independent of uncertainty and determined only by adjustment costs (and other factors). Thus, holding those other factors fixed, the two moments, $\rho_{k,a_{-1}}$ and $\rho_{k,k_{-1}}$, jointly pin down the magnitude of adjustment frictions and the extent of uncertainty.

4 Quantitative Analysis

The analytical results in the previous section showed a tight relationship between the moments $\left(\rho_{k,a_{-1}}, \rho_{k,k_{-1}}, \sigma_k^2, \lambda_{arpk,a}\right)$ and the parameters $(\mathbb{V}, \xi, \sigma_\varepsilon^2, \gamma)$ for the special case of $\rho = 1$. In this section, we use this insight to develop an empirical strategy for the more general case where productivity follows a stationary AR(1) process and apply it to data on Chinese manufacturing firms. This allows us to quantify the severity of the various forces and their impact on *arpk* dispersion and economic aggregates. For purposes of comparison, we also provide results for publicly traded firms in the US.[18] In Section 5, we extend our methodology to explore some specific candidates for firm-specific factors other than adjustment/informational frictions.

4.1 Parameterization

We begin by assigning values to the more standard preference and production parameters of our model. We assume a period length of one year and accordingly set the discount factor $\beta = 0.95$. We use an annual depreciation rate of $\delta = 0.10$. We keep the elasticity of substitution θ common across countries and set its value to 6, roughly in the middle of the range of values in the literature.[19] We assume constant returns to scale in production, but allow the parameters $\hat{\alpha}_1$ and $\hat{\alpha}_2$ to vary across countries. In the US, we set these to standard values of 0.33 and 0.67, respectively, which implies $\alpha = 0.62$.[20] For China, we set a higher capital share, namely,

[18]The two sets of firms are not directly comparable due to their differing coverage. For example, the Chinese data include many more small firms. Similarly, there may be selection biases when using data on publicly traded firms. To partly address this concern, in Appendix J, we repeat the analysis on Chinese publicly traded firms. We find patterns that are quite similar to those for Chinese manufacturing firms, suggesting that cross-country differences in frictions and distortionary factors are quite significant. This conclusion is further supported by results for two additional countries, Colombia and Mexico, also presented in Appendix J.

[19]In Appendix I.3, we report results for $\theta = 3$, the value used in Hsieh and Klenow (2009).

[20]This is very close to the estimate of 0.59 in Cooper and Haltiwanger (2006). We also estimated α following the indirect inference approach in, e.g., Cooper et al. (2015). Specifically, we find the value of α so that the

$\hat{\alpha}_1 = \hat{\alpha}_2 = 0.5$, in line with evidence from a number of recent papers, for example, Bai et al. (2006). These values imply an α equal to 0.71 in China.[21]

Next, we turn to the parameters of the productivity process, a_{it}: the persistence, ρ, and the variance of the innovations, σ^2_μ. Under our assumptions, firm-level productivity is directly given by (up to an additive constant) $a_{it} = va_{it} - \alpha k_{it}$ where va_{it} denotes the log of value-added.[22] Controlling for industry-year fixed effects to isolate the firm-specific component, we use a standard autoregression to estimate the parameters ρ and σ^2_μ.

To pin down the remaining parameters – the adjustment cost, ξ, the quality of information, $\mathbb{V}$, and the size of other factors, γ and σ^2_ε – we follow a strategy informed by the results in the previous section. Specifically, we target the correlation of investment growth with lagged innovations in productivity ($\rho_{\iota,a_{-1}}$), the autocorrelation of investment growth ($\rho_{\iota,\iota_{-1}}$), the variance of investment growth (σ^2_ι) and the correlation of the average product of capital with productivity ($\rho_{arpk,a}$).[23] Finally, to infer σ^2_χ, the variance of the fixed component in (6), we match the overall dispersion in the average product of capital, σ^2_{arpk}, which is clearly increasing in σ^2_χ. Thus, by construction, our parameterized model will match the observed *arpk* dispersion in the data, allowing us to decompose the contribution of each factor. Appendix C describes our numerical estimation procedure in detail. We summarize our empirical approach in Table 1.

4.2 Data

The data on Chinese manufacturing firms are from the Annual Surveys of Industrial Production conducted by the National Bureau of Statistics. The surveys include all industrial firms (both state-owned and non-state owned) with sales above 5 million RMB (about $600,000). We use data spanning the period 1998-2009.[24] The original data come as a repeated cross-section.

coefficient from an OLS regression of value-added on capital using model-simulated data matches its counterpart from an identical regression in the data. This procedure also yields $\alpha = 0.62$.

[21] Using the same capital share for both countries yields a very similar decomposition of observed σ^2_{arpk}. More generally, the curvature of the profit function, α, plays a key role in determining the TFP/output implications of a given degree of σ^2_{arpk}, but does not materially change the estimated contributions of various factors, the main focus of this paper. See also Section 6.2 (where labor distortions leads to a higher α), Appendix I.3 (where a lower elasticity of substitution leads to a lower α), as well as Section 6.4 (sectoral heterogeneity in α).

[22] An alternative strategy is to measure the true productivity directly, i.e., $\hat{a}_{it} = va_{it} - \alpha_1 k_{it} - \alpha_2 n_{it}$, and construct the implied $a_{it} = \frac{1}{1-\alpha_2}\hat{a}_{it}$. The two approaches are equivalent under an undistorted labor choice, but Appendix E.1 shows that more generally, firm-specific capital profitability, a_{it}, is a combination of productivity and a labor distortion. As a result, inferring a_{it} from $\hat{a}_{it}$ without adjusting for a potentially distorted labor choice can lead to biased estimates, while the strategy of directly measuring a_{it}, as we do here, remains valid.

[23] We use investment growth to partly cleanse the data of firm-level fixed-effects, which have been shown to play a significant role in firm-level investment data (in the analytical cases studied earlier, we used the level of investment). See Morck et al. (1990) for more on this issue. Appendix I.4 shows that our results are largely unchanged if we use the autocorrelation and variance of investment in levels, rather than growth rates.

[24] Industrial firms correspond to Chinese Industrial Classification codes 0610-1220, 1311-4392 and 4411-4620, which includes mining, manufacturing and utilities. Early vintages of the NBS data did not report all variables

Table 1: Parameterization - Summary

Parameter	Description	Target/Value
Preferences/production		
θ	Elasticity of substitution	6
β	Discount rate	0.95
δ	Depreciation	0.10
$\hat{\alpha}_1$	Capital share	0.33 US/0.50 China
$\hat{\alpha}_2$	Labor share	0.67 US/0.50 China
Productivity/frictions		
ρ	Persistence of productivity	$\rho_{a,a_{-1}}$
σ^2_μ	Shocks to productivity	σ^2_a
$\mathbb{V}$	Signal precision	$\rho_{\iota,a_{-1}}$
ξ	Adjustment costs	$\rho_{\iota,\iota_{-1}}$
γ	Correlated factors	$\rho_{arpk,a}$
σ^2_ε	Transitory factors	σ^2_ι
σ^2_χ	Permanent factors	σ^2_{arpk}

A panel is constructed following almost directly the method outlined in Brandt et al. (2014), which also contains an excellent overview of the data for the interested reader. The Chinese data have been used multiple times and are by now familiar in the misallocation literature – for example, Hsieh and Klenow (2009) – although our use of the panel dimension is rather new. The data on US publicly traded firms comes from Compustat North America. We use data covering the same period as for the Chinese firms.

We measure the firm's capital stock, k_{it}, in each period as the value of fixed assets in China and of property, plant and equipment (PP&E) in the US.[25] Value-added is estimated as a constant fraction of revenues using a share of intermediates of 0.5. We measure the average product of capital as $arpk_{it} = va_{it} - k_{it}$. Net investment and productivity growth are obtained by first differencing k_{it} and a_{it}, respectively. To isolate the firm-specific variation in our data series, we extract a time-by-industry fixed-effect from each and use the residual. In both countries, industries are classified at the 4-digit level. This is equivalent to deviating each firm from the unweighted average within its industry in each period and also eliminates aggregate

for the full set of firms in the years after 2007. Although this does not seem to be an issue in our sample (all the variables we use are well populated in all years), we have also redone our analysis using data only through 2007. The estimates are very similar (as noted below, the moments are fairly stable over time).

[25]Our baseline measure of the capital stock uses reported book values. In Section 6.4 (details in Appendix I.5), we construct the capital stock using the perpetual inventory method for the US firms and re-estimate the model. This yields slightly different point estimates, but very similar patterns for the role of various factors.

components. After eliminating duplicates and problematic observations (for example, firms reporting in foreign currencies), outliers, observations with missing data etc., our final sample consists of 797,047 firm-year observations in China and 34,260 in the US. Appendix B provides further details on how we build our sample and construct the moments, as well as summary statistics from 2009.[26]

Table 2 reports the target moments for both countries.[27] The first two columns show the productivity moments, which have similar persistence but higher volatility in China. The remaining columns show that, in China, investment growth is more correlated with past shocks, more volatile and less autocorrelated. The Chinese data also show a higher correlation between productivity and *arpk* and substantially larger dispersion in *arpk*. These patterns will lead us to significantly different estimates of the severity of various factors across the two sets of firms.

Table 2: Target Moments

	ρ	σ_μ^2	$\rho_{\iota,a_{-1}}$	$\rho_{\iota,\iota_{-1}}$	$\rho_{arpk,a}$	σ_ι^2	σ_{arpk}^2
China	0.91	0.15	0.29	−0.36	0.76	0.14	0.92
US	0.93	0.08	0.13	−0.30	0.55	0.06	0.45

4.3 Identification

Before turning to the estimation results, we revisit the issue of identification. Although we no longer have analytical expressions for the mapping between moments and parameters, we use a numerical experiment to show that the intuition developed in Section 3 for the random walk case applies here as well. In that section, we used a pairwise analysis to demonstrate how various moments combine to help disentangle the sources of observed *arpk* dispersion. Here, we repeat that analysis by plotting numeric isomoment curves in Figure 2, using the moments and parameter values for US firms (from Tables 2 and 3, respectively). They reveal the same patterns as Figure 1, indicating that the logic of that special case goes through here as well.[28]

[26] We have also examined the moments year-by-year. They are reasonably stable over time.

[27] We report bootstrapped standard errors for the moments in Table 9 in Appendix C. Given the large sample sizes (almost 800,000 in China and 35,000 in the US), the estimates of the moments are extremely precise.

[28] The differences in the precise shape of some of the curves in the two figures come partly from the departure from the random walk case and also from the fact that they use slightly different moments (Figure 2 works with changes in investment and $\rho_{arpk,a}$ while Figure 1 used changes in k and $\lambda_{arpk,a}$).

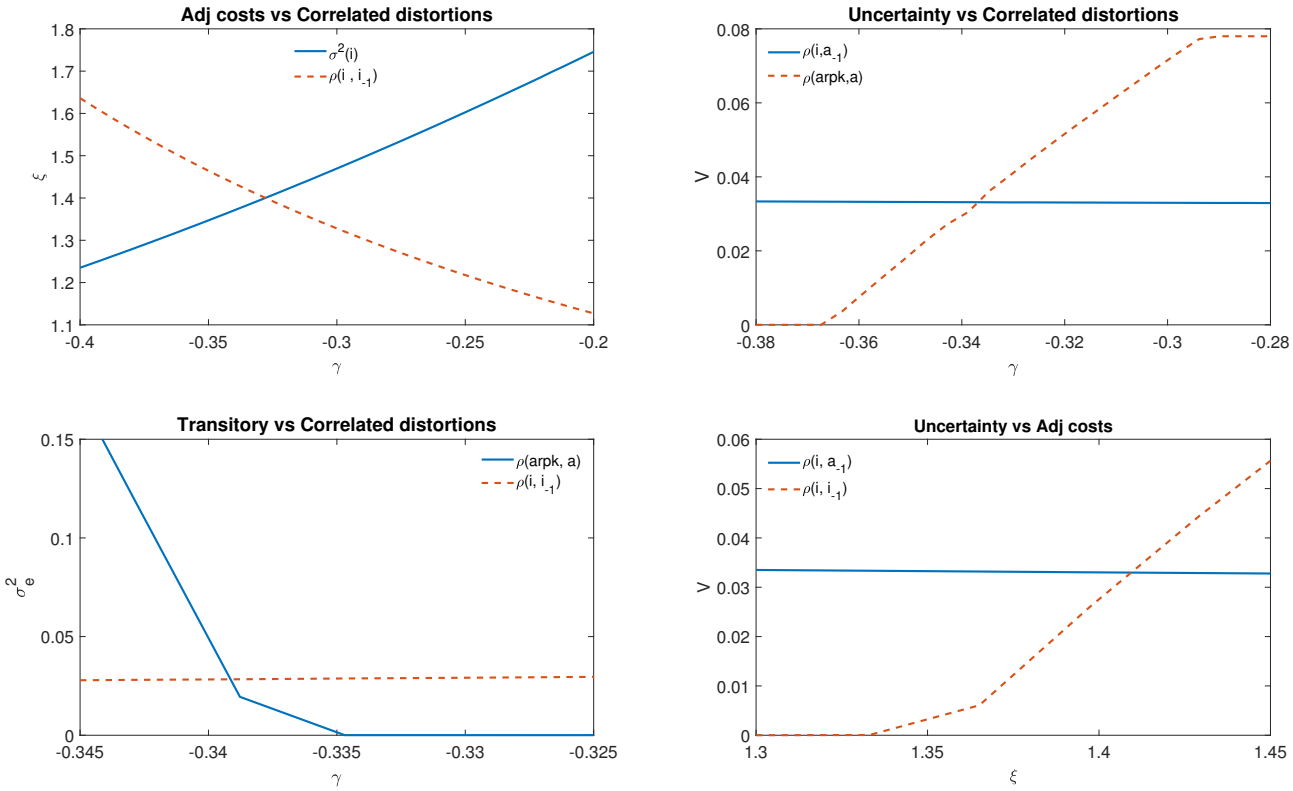


Figure 2: Isomoment Curves - Quantitative Model

4.4 The Sources of 'Misallocation'

Table 3 contains our baseline results. In the top panel we display the parameter estimates.[29] In the second panel of Table 3, we report the contribution of each factor to *arpk* dispersion, which we denote $\Delta\sigma^2_{arpk}$.[30] These are calculated under the assumption that only the factor of interest is operational, i.e., in the absence of the others, so that the contribution of each one is measured relative to the undistorted first-best.[31] The third panel expresses this contribution as a percentage of the total *arpk* dispersion measured in the data, denoted $\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$. Because of interactions between the factors, there is no *a priori* reason to expect these relative contributions to sum to one. In practice, however, we find that the total is reasonably close to one, allowing

[29]Table 9 in Appendix C reports standard errors and compares the model-simulated moments (at the estimated parameters) to their empirical counterparts. The parameters are quite precisely estimated (again, both of our firm-level datasets have a relatively large number of observations) and the model matches the five moments almost exactly in both countries.

[30]For adjustment costs, we do not have an analytic mapping between the severity of these costs and σ^2_{arpk}, but this is a straightforward calculation to make numerically; for each of the other factors, we can compute their contributions to *arpk* dispersion analytically.

[31]An alternative would be to calculate the contribution of each factor holding the others constant at their estimated values. It turns out that the interactions between the factors are small at the estimated parameter values, so the two approaches yield similar results. Table 10 in Appendix D shows that the effects of each factor on *arpk* dispersion in the US are close under either approach. Interaction effects are even smaller in China.

us to interpret this exercise as a decomposition of total observed dispersion. In the bottom panel of the table, we compute the implied losses in aggregate TFP, again relative to the undistorted first-best level, i.e., $\Delta a = a^* - a$. Once we have the contribution of each factor to *arpk* dispersion, computing these values is simply an application of expression (9).

Table 3: Contributions to 'Misallocation'

			Other Factors		
	Adjustment Costs	Uncertainty	Correlated	Transitory	Permanent
Parameters	ξ	$\mathbb{V}$	γ	σ^2_ε	σ^2_χ
China	0.13	0.10	−0.70	0.00	0.41
US	1.38	0.03	−0.33	0.03	0.29
$\Delta\sigma^2_{arpk}$					
China	0.01	0.10	0.44	0.00	0.41
US	0.05	0.03	0.06	0.03	0.29
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$					
China	1.3%	10.3%	47.4%	0.0%	44.4%
US	10.8%	7.3%	14.4%	6.3%	64.7%
Δa					
China	0.01	0.08	0.38	0.00	0.36
US	0.02	0.01	0.03	0.01	0.13

Adjustment costs. Our results show evidence of economically significant adjustment frictions. For example, the estimate of 1.38 for ξ in the US implies a value of 0.2 for $\hat{\xi}$ in the adjustment cost function.[32] This puts us in the middle of previous estimates of convex costs in the literature, though differences in data and empirical strategies complicate direct comparisons. For example, using US manufacturing data, Asker et al. (2014) estimate a convex adjustment cost of 8.8 in a monthly model, which translates to an annual $\hat{\xi} = 0.73$, roughly a factor of four above our estimate.[33] Our estimate is closer to, and slightly higher than, Cooper and Haltiwanger (2006), who find $\hat{\xi} = 0.05$ for US manufacturing plants.

What leads us to find different estimates? The answer lies primarily in the fact that our model explicitly includes additional factors that may act on the investment decision – e.g., distortions – and consequently, our empirical strategy is designed to match a broader set of

[32]The mapping between ξ and $\hat{\xi}$ is in equation (22) in Appendix A.1.

[33] To interpret this difference, a firm that doubles its capital stock in a year would incur an adjustment cost equal to 11% of the value of the investment according to our estimate, but equal to 60% at the Asker et al. (2014) estimate.

moments.[34] The papers mentioned above abstract from these factors and focus on matching different moments. For example, Asker et al. (2014) target the variability of investment (among other moments), but do not try to match the autocorrelation, while Cooper and Haltiwanger (2006) do the reverse. As we saw in Section 3, in the presence of correlated factors, the first strategy overstates the true extent of adjustment costs, while the second understates it. It turns out that this bias can be quite large: an adjustment cost-only model (i.e., ignoring other factors) estimated to match the volatility of investment growth yields an estimate of $\hat{\xi}$ about 60% higher than our baseline estimate, but predicts a counterfactually high autocorrelation of investment growth: -0.17 vs -0.30 in the data. A strategy targeting only the serial correlation leads to the opposite conclusion – a lower estimate of $\hat{\xi}$, but at the cost of excessively high variability compared to the data. These patterns are exactly in line with the arguments developed in Section 3. More broadly, these exercises can partly explain the wide range of adjustment cost estimates in the literature – when adjustment costs are estimated without explicitly controlling for other factors, the results can be quite sensitive to the particular moments chosen.[35] Indeed, our results suggest that explicitly accounting for these additional factors is essential in order to reconcile a broad set of moments in firm-level investment dynamics.

The estimated value of ξ is lower in China. Investment growth in China is both more volatile and less serially correlated than for US firms, which (together with the other moments), leads the estimation to find a lower degree of adjustment frictions. Importantly, as in the US, one would reach a very different conclusion from examining a model with only adjustment costs: for example, estimating such a model by targeting σ_ι^2 in China yields an estimate for ξ of about 1.5, roughly 10 times larger than the one in Table 3.

Perhaps most importantly for purposes of our analysis, in both countries, the estimated adjustment costs do not contribute significantly to *arpk* dispersion. If adjustment costs were the only friction in China, σ^2_{arpk} would be 0.01 (the observed level is 0.92). The higher estimate of ξ in the US implies a slightly higher, though still modest, contribution (by themselves, adjustment costs lead to $\sigma^2_{arpk} = 0.05$ or 11% of the observed dispersion). The corresponding aggregate TFP losses are 1% and 2% in the two countries, respectively.

This does not mean that adjustment costs are irrelevant for understanding firm-level investment dynamics. Setting adjustment costs to zero in the US while holding the other parameters

[34]There are a few other differences between our approach and these papers: (i) they have convex and non-convex (fixed) adjustment costs. In Section 6.1, we show that our estimates of $\hat{\xi}$ change little when we introduce a fixed cost; (ii) they use moments of investment in levels while we work with growth rates. In Appendix I.4, we show that targeting the variance and autocorrelation of investment in levels changes the estimate of $\hat{\xi}$ only slightly; (iii) Asker et al. (2014) follow a different strategy to estimate the process for profitability, a_{it}: they directly measure productivity $\hat{a}_{it}$ and use the implied a_{it}. See footnote 22. As we show in Appendix E.1, this strategy can overstate the volatility of a_{it}, i.e., σ^2_μ, and bias adjustment cost estimates upward.

[35]Bloom (2009) points out the wide variation in these estimates, ranging from 0 to 20 (Table IV).

at their estimated values causes the variance of investment growth to spike to 1.68 (compared to 0.06 in the data) and the autocorrelation to plummet to -0.62 (data: -0.30). However, σ^2_{arpk} falls only modestly, from 0.45 to 0.41. Re-estimating the model without adjustment costs (and dropping the autocorrelation as a target) also leads to a counterfactually low autocorrelation (-0.50).[36] In other words, while adjustment frictions are an important determinant of investment dynamics, they do not generate significant dispersion in average products of capital.[37]

Uncertainty. Table 3 shows that firms in both countries make investment decisions under considerable uncertainty, with the information friction more severe for Chinese firms. As a share of the prior uncertainty, σ^2_μ, residual uncertainty, $\frac{\mathbb{V}}{\sigma^2_\mu}$, is 0.42 in the US and 0.63 in China.[38] In an environment where imperfect information is the only friction, we have $\sigma^2_{arpk} = \mathbb{V}$, so the contribution of uncertainty alone to observed *arpk* dispersion can be directly read off the second column in Table 3 – namely 0.10 in China and 0.03 in the US. These represent about 10% and 7% of total *arpk* dispersion in the two countries, respectively. The implications for aggregate TFP are substantial in China – losses are about 8% – and are lower in the US, about 1%. Note, however, that imposing a one period time-to-build assumption where firms install capital in advance without any additional information about innovations in productivity, i.e. setting $\mathbb{V} = \sigma^2_\mu$, would overstate uncertainty (and bias the estimates of adjustment costs and other parameters). Indeed, doing so yields estimates of $\mathbb{V}$ that are about 55% higher in China and a factor of 2.5 times higher in the US.

'Distortions'. The last three columns of Table 3 show that other, potentially distortionary, factors play a significant role in generating the observed *arpk* dispersion in both countries. Turning first to the correlated component, the negative values of γ suggest that they act to disincentivize investment by more productive firms and especially so in China. The contribution of these distortions to *arpk* dispersion is given by $\gamma^2\sigma^2_a$, which amounts to 0.44 in China, or 47% of total dispersion. The associated aggregate consequences are also quite sizable – TFP

[36] The estimates for other parameters also change: notably, a more negative γ is needed to match σ^2_ι.

[37] Asker et al. (2014) make a similar observation – across various specifications of adjustment costs (including one with zero adjustment costs and a one period time-to-build), their model's performance in capturing dispersion in *arpk* is not dramatically altered, even though the implications for other moments (e.g., the variability of investment) are quite different. See Table 9 and the accompanying discussion in that paper.

[38] Our values for $\frac{\mathbb{V}}{\sigma^2_\mu}$ are similar to those in David et al. (2016), who find 0.41 and 0.63 for publicly traded firms in the US and China, respectively. The estimates of $\mathbb{V}$ are different but are not directly comparable – David et al. (2016) focus on longer time horizons (they analyze 3-year time intervals). This might lead one to conclude that ignoring other factors – as David et al. (2016) do – leads to negligible bias in the estimate of uncertainty. But, this is not a general result and rests on the fact that adjustment costs and uncorrelated distortions are estimated to be modest. Then, as Figure 2 shows, the sensitivity of actions to signals turns out to be a very good indicator of uncertainty. If, on the other hand, adjustment costs and/or uncorrelated factors were much larger, the bias from estimating uncertainty alone can be quite significant.

losses from these sources are 38%. In contrast, the estimate of γ in the US is significantly less negative than in China, suggesting that these types of correlated factors are less of an issue for firms in the US, both in an absolute sense – the *arpk* dispersion from these factors in the US is 0.06, less than one-seventh that in China – and in relative terms – they account for only 14% of total observed *arpk* dispersion in the US. The corresponding TFP effects are also considerably smaller for the US - losses from correlated sources are only about 3%.

Next, we consider the role of distortions that are uncorrelated with firm productivity. Table 3 shows that purely transitory factors (measured by σ^2_ε) are negligible in both countries, but permanent firm-specific factors (measured by σ^2_χ) play a prominent role. Their contribution to *arpk* dispersion, which is also given by σ^2_χ, amounts to 0.41 in China and 0.29 in the US. Thus, their absolute magnitude in the US is considerably below that in China, but in relative terms, these factors seem to account for a substantial portion of measured *arpk* dispersion in both countries. The aggregate consequences of these types of distortions are also significant, with TFP losses of 36% in China and about 13% in the US.

In sum, the estimation results point to the presence of substantial distortions to investment, especially in China, where they disproportionately disincentivize investment by more productive firms. Section 6 and Appendix I show that these results are robust to a number of modifications to our baseline setup, e.g., allowing for non-convex adjustment costs, a frictional labor choice, richer stochastic processes on productivity and distortions, curvature assumptions and additive measurement error. Further, we have applied the methodology to data on Colombian and Mexican firms (in addition to the set of publicly traded firms in China) – the results resemble those for Chinese manufacturing firms, in that they point to a substantial role for correlated factors, as well as fixed ones (details are in Appendix J).

What patterns in the data lead us to this conclusion? In both countries, we see considerable dispersion in *arpk*, which tends to be correlated with firm productivity. The fact that investment growth is not very correlated through time and responds only modestly to past shocks limits the role of adjustment and informational frictions and assigns a substantial role to other factors. The high correlation of *arpk* with productivity, particularly in China, suggests that these factors vary systematically with productivity. In both countries, large fixed distortions, uncorrelated with productivity, are necessary to rationalize the total observed dispersion in *arpk*. In the next section, we explore some candidates for these firm-specific factors.

5 Firm-Specific Factors: Some Candidates

Our results suggest a large role for firm-specific 'distortions' in explaining the observed *arpk* dispersion. In this section, we extend our baseline framework and empirical methodology to

investigate three potential sources – heterogeneity in markups and production technologies, size-dependent policies and financial considerations.

5.1 Heterogeneity in Markups and Technologies

In our baseline setup, all firms within an industry (1) operated identical production technologies and (2) were monopolistically competitive facing CES demand curves and therefore, had identical markups. As a result, any firm-level heterogeneity in technologies and/or markups would show up in our estimates of other factors (note also that this type of variation would drive a wedge between *arpk* and the true marginal product of capital, implying that dispersion in the former is not necessarily a sign of misallocation). Here, we explore this possibility using a modified version of our baseline model. This requires more assumptions and additional data, but allows us to provide an upper bound on the contribution of these elements.

We begin by generalizing the production function from Section 2 to include intermediate inputs and to allow for (potentially time-varying) heterogeneity in input intensities. Specifically, the output of firm i is now given by

$$Y_{it} = K_{it}^{\hat{\alpha}_{it}} N_{it}^{\hat{\zeta}_{it}-\hat{\alpha}_{it}} M_{it}^{1-\hat{\zeta}_{it}} ,$$

where M_{it} denotes intermediate or materials input. The price of these inputs, potentially firm-specific, is denoted P_{it}^M. Throughout this section, we abstract from adjustment/information frictions in firms' input decisions. This is largely for simplicity, but is also supported by the relatively modest role played by these dynamic considerations in our baseline estimates.[39]

Capital and labor choices are assumed to be subject to a factor-specific 'distortion' (in addition to the markup), denoted T_{it}^K and T_{it}^N, respectively, but the choice of intermediates is undistorted except for the markup. Since the method remains valid even with unobserved firm-specific variation in the price of intermediate goods, it does allow for distortions in the market for intermediate inputs, so long as they are reflected in prices. Formally, these assumptions imply that the firm's optimal choices solve the following cost minimization problem:

$$\min_{K_{it},N_{it},M_{it}} \; R_t T_{it}^K K_{it} + W_t T_{it}^N N_{it} + P_{it}^M M_{it} \quad s.t. \quad Y_{it} \leq K_{it}^{\hat{\alpha}_{it}} N_{it}^{\hat{\zeta}_{it}-\hat{\alpha}_{it}} M_{it}^{1-\hat{\zeta}_{it}} .$$

The contribution of markup dispersion. To quantify the contribution of markup dispersion, we use the powerful methodology of De Loecker and Warzynski (2012), which allows us to

[39]It is possible to extend the identification methodology from Section 3 to explicitly incorporate these forms of heterogeneity. Although this would require more assumptions (e.g., on the correlation structure of markups/technologies with productivity and over time) and make the intuition more complicated, the basic insights should still go through.

measure firm-level markups without taking a stand on the nature of competition/demand. To do so, we assume a common materials elasticity across firms within an industry, i.e., $\hat{\zeta}_{it} = \hat{\zeta}_t \; \forall \; i$.[40] Cost minimization then implies the following optimality condition (we suppress the time subscript on ζ, though the method remains valid with arbitrary time-variation):

$$P_{it}^M = MC_{it}\left(1-\hat{\zeta}\right)\frac{Y_{it}}{M_{it}} \quad \Rightarrow \quad \frac{P_{it}^M M_{it}}{P_{it}Y_{it}} = (1-\hat{\zeta})\frac{MC_{it}}{P_{it}}\,, \tag{14}$$

where MC_{it} is the marginal cost of the firm. This condition states that, at the optimum, the firm sets the materials share in gross output equal to the inverse of the markup, $\frac{MC_{it}}{P_{it}}$, multiplied by the materials elasticity $1-\hat{\zeta}$.

Expression (14) suggests a simple way to estimate the cross-sectional dispersion in markups. The left-hand side is materials' share of revenue – the within-industry dispersion in this object (in logs) maps one-for-one into (log) markup dispersion. Data on materials expenditures are directly reported in the Chinese data. In the US, we follow, e.g., De Loecker and Eeckhout (2017) and İmrohoroğlu and Tüzel (2014) and calculate intermediate expenditures as total expenses less labor expenses. The former are defined as sales less operating income and the latter are imputed using number of employees and the average industry wage, from the NBER-CES Manufacturing Industry Database.[41]

The results of applying this procedure are reported in Table 4. The variance of the share of materials in revenue (the first row in the top panel) is about 0.06 in the US Compustat data and 0.05 in China. This accounts for about 14% of σ^2_{arpk} among the US firms, but only about 4% among Chinese manufacturing firms. Thus, markup heterogeneity composes a non-negligible fraction of observed *arpk* dispersion among US publicly traded firms but seems to be an almost negligible force in China.

Can markup variation help explain the large role for correlated distortions in Section 4? In theory, yes – markups that increase with size would be consistent with the patterns uncovered in that section. Quantitatively, however, this does not seem to hold much promise. In China, this is clear simply from the rather modest dispersion in markups – the contribution of markups to *arpk* dispersion is much smaller (0.05) than the estimated total contribution of correlated factors (0.44). In the US, where dispersion in markups is more substantial, the data suggest they are largely independent of firm size. For example, projecting the measured markup on

[40] In Appendix F, we allow the materials elasticity to vary across firms within an industry (our baseline calculation already allows for variation across industries). Under certain orthogonality assumptions, we show that the covariance of the materials share with *arpk* and *arpn* (average product of labor) pins down the dispersion in markups. This approach yields very similar (if slightly lower) estimates of markup dispersion in both countries.

[41] Details of these calculations are provided in Appendix B. In an earlier version of this paper, we also used the reported wage bill, available for a smaller subset of firms. The results were broadly similar, though the share of both markup and technology dispersion was somewhat higher (28% and 62 % of σ^2_{arpk} for those firms).

revenues yields a statistically significant, yet economically small, coefficient of 0.01 (the raw correlation between markups and revenues is 0.07).

The contribution of technology dispersion. Cost minimization also implies that the average revenue products of capital and labor are given by:[42]

$$arpk_{it} \equiv \log\left(\frac{P_{it}Y_{it}}{K_{it}}\right) = \log\frac{P_{it}}{MC_{it}} - \log\hat{\alpha}_{it} + \tau_{it}^{K} + \text{Constant} \tag{15}$$

$$arpn_{it} \equiv \log\left(\frac{P_{it}Y_{it}}{N_{it}}\right) = \log\frac{P_{it}}{MC_{it}} - \log(\hat{\zeta} - \hat{\alpha}_{it}) + \tau_{it}^{N} + \text{Constant} \tag{16}$$

$$\approx \log\frac{P_{it}}{MC_{it}} + \left(\frac{\bar{\alpha}}{\hat{\zeta} - \bar{\alpha}}\right)\log\hat{\alpha}_{it} + \tau_{it}^{N} + \text{Constant}\ , \tag{17}$$

where τ_{it}^{K} and τ_{it}^{N} denote (the logs of) the capital and labor wedges T_{it}^{K} and T_{it}^{N}, respectively, and $\bar{\alpha}$ is the average capital elasticity. Average revenue products are combinations of firm-specific production elasticities, markups and wedges. Note that the capital elasticity, $\hat{\alpha}_{it}$, has opposing effects on the average products of capital and labor: firms with a high $\hat{\alpha}_{it}$ will, *ceteris paribus*, have a low *arpk* and a high *arpn*. In other words, this form of heterogeneity acts like a 'mix' distortion (as opposed to 'scale' factors, which distort all input decisions in the same direction). We will make use of this property to derive an upper bound for variation in $\hat{\alpha}_{it}$ using the observed covariance of *arpk* and *arpn*. Let

$$\widetilde{arpk}_{it} \equiv \log\left(\frac{P_{it}Y_{it}}{K_{it}}\right) - \log\left(\frac{P_{it}}{MC_{it}}\right)$$

$$\widetilde{arpn}_{it} \equiv \log\left(\frac{P_{it}Y_{it}}{N_{it}}\right) - \log\left(\frac{P_{it}}{MC_{it}}\right)$$

denote the markup-adjusted average revenue products of capital and labor. Appendix F proves the following result:

Proposition 2. *Suppose* $\log\hat{\alpha}_{it}$ *is uncorrelated with the distortions* τ_{it}^{K} *and* τ_{it}^{N}*. Then, the cross-sectional dispersion in* $\log\hat{\alpha}_{it}$ *satisfies*

$$\sigma^2(\log\hat{\alpha}_{it}) \leq \frac{\sigma^2_{\widetilde{arpk}}\sigma^2_{\widetilde{arpn}} - cov\left(\widetilde{arpk}, \widetilde{arpn}\right)^2}{2\frac{\bar{\alpha}}{\hat{\zeta}-\bar{\alpha}}cov\left(\widetilde{arpk}, \widetilde{arpn}\right) + \left(\frac{\bar{\alpha}}{\hat{\zeta}-\bar{\alpha}}\right)^2\sigma^2_{\widetilde{arpk}} + \sigma^2_{\widetilde{arpn}}}\ . \tag{18}$$

The bound in (18) is obtained by setting the correlation between τ_{it}^{K} and τ_{it}^{N} to 1. Given the observed second moments of $\left(\widetilde{arpk}_{it}, \widetilde{arpn}_{it}\right)$, this maximizes the potential for variation in

[42] See Appendix F for details. The third equation is derived by log-linearizing (16) around $\hat{\alpha}_{it} = \bar{\alpha}$.

$\hat{\alpha}_{it}$, which, as noted earlier, is a source of negative correlation between $\widetilde{arpk}_{it}$ and $\widetilde{arpn}_{it}$. The expression for the bound reveals the main insight: the more positive the covariance between $\left(\widetilde{arpk}_{it}, \widetilde{arpn}_{it}\right)$, the lower is the scope for heterogeneity in $\hat{\alpha}_{it}$.

To compute this bound for the two countries, we set $\hat{\zeta}$, the share of materials in gross output, to 0.5. The results, along with the moments used, are reported in Table 4. They show that heterogeneity in technologies can potentially account for a substantial portion of σ^2_{arpk} in the US – as much as 44% – and a more modest, though still significant, fraction in China, about 23%.[43] In other words, a substantial portion of measured *arpk* dispersion in both countries may not be a sign of misallocated resources at all. This is most striking in the US, where the average products of capital and labor covary less positively than in China.[44]

Table 4: Heterogeneous Markups and Technologies

	China		US	
Moments				
$\sigma^2\left(\log \frac{P_{it}Y_{it}}{P_t^M M_{it}}\right)$	0.05		0.06	
$\text{cov}\left(\widetilde{arpk}_{it}, \widetilde{arpn}_{it}\right)$	0.41		0.23	
$\sigma^2\left(\widetilde{arpk}_{it}\right)$	1.37		0.52	
$\sigma^2\left(\widetilde{arpn}_{it}\right)$	0.76		0.35	
Estimated $\Delta\sigma^2_{arpk}$				
Dispersion in Markups	0.05	(3.8%)	0.06	(13.6%)
Dispersion in $\log \hat{\alpha}_{it}$	0.30	(23.1%)	0.18	(44.4%)
Total	0.35	(26.9%)	0.24	(58.1%)

Notes: The values in parentheses in the bottom panel are the contributions to *arpk* dispersion expressed as a fraction of total σ^2_{arpk}.

An alternative approach to assessing the potential for technology dispersion is discussed in Hsieh and Klenow (2009), in which all the variation in firm-level capital-labor ratios is attributed to heterogeneity in $\hat{\alpha}_{it}$. This amounts to assuming that $\tau^K_{it} = \tau^N_{it}$, which implies:

$$k_{it} - n_{it} \;=\; arpn_{it} - arpk_{it} \;\approx\; \frac{\hat{\zeta}}{\hat{\zeta} - \bar{\alpha}} \log \hat{\alpha}_{it} \quad \Rightarrow \quad \sigma^2\left(k_{it} - n_{it}\right) = \left(\frac{\hat{\zeta}}{\hat{\zeta} - \bar{\alpha}}\right)^2 \sigma^2(\log \hat{\alpha}_{it}) \; .$$

This procedure yields estimates for $\sigma^2\left(\log \hat{\alpha}_{it}\right)$ that are quite close to those in Table 4: 0.27

[43] There is some evidence that the share of intermediates may be higher in China than the US, see, e.g., Table 1 in Brandt et al. (2014). We re-computed the bound with $\hat{\zeta} = 0.25$ and obtained very similar results. We also verified the accuracy of the approximation by working directly with (16) instead of the log-linearized version in (17). This yielded slightly lower bounds: 38% and 17% of σ^2_{arpk} in the US and China, respectively.

[44] In Appendix F, we derive an analogous bound in the case with within-industry variation in the materials elasticity. The results are very similar.

(compared to 0.30) for China and 0.16 (compared to 0.18) in the US.

In sum, unobserved heterogeneity in markups and technologies seem to be promising candidates for firm-specific factors, which drive most of the *arpk* dispersion in the data. This is particularly true for the US, where they can explain as much as 58% of the observed dispersion in the US. In China, their role is more modest, but still meaningful, at 27%.

5.2 Size-Dependent Policies

Our results show a significant role for factors correlated with firm-level productivity – especially in China – in explaining the observed variation in *arpk* across firms. Here, we show how policies that affect or restrict the size of firms can lead to a correlated factor of this form. A number of papers have pointed out the prevalence of distortionary size-dependent policies across a range of countries, for example, Guner et al. (2008). These policies often take the form of restrictions (or additional costs) associated with acquiring capital and/or other inputs. To be clear, our goal is not to explore the role of a particular policy in China or the US. Rather, we show how policies that are common in a number of countries can generate patterns that are, in a sense, isomorphic to factors correlated with productivity.

Towards this end, we generalize our baseline specification of firm-specific factors in equation (6) to allow for a component that varies with the chosen level of capital. Formally,

$$\tau_{it} = \gamma_k k_{it} + \gamma a_{it} + \varepsilon_{it} + \chi_i \ ,$$

where the parameter γ_k indexes the severity of this additional component. The empirically relevant case is $\gamma_k < 0$, which implicitly penalizes larger firms. This specification captures the essence of the policies discussed above in a tractable way (e.g., it allows us to continue to use perturbation methods). With this formulation, the log-linearized Euler equation takes the form

$$k_{it+1}\left((1+\beta)\xi + 1 - \alpha - \gamma_k\right) = (1+\gamma)\,\mathbb{E}_{it}\left[a_{it+1}\right] + \varepsilon_{it+1} + \chi_i + \beta\xi\mathbb{E}_{it}\left[k_{it+2}\right] + \xi k_{it} \ . \qquad (19)$$

Expression (19) is identical to expression (4), but with $\alpha + \gamma_k$ taking the place of α. It is straightforward to derive the firm's investment policy function and verify that the same adjustment goes through, i.e., expressions (7) and (8) hold, with α everywhere replaced by $\alpha + \gamma_k$. Intuitively, the size-dependent component, γ_k, changes the effective degree of curvature in the firm's investment problem – although the curvature of the profit function remains α, the firm acts as if it is $\alpha + \gamma_k$. If $\gamma_k < 0$, the distortion dampens the responsiveness of investment to shocks. If $\gamma_k > 0$, the responsiveness of investment is amplified.

Importantly, these effects are broadly similar to those coming from γ: indeed, if γ_k were

the only factor distorting investment choices, the implied law of motion for k_{it} is identical (up to a first-order) to one with only productivity-dependent factors, where $\gamma = \frac{\gamma_k}{1-\alpha-\gamma_k}$. The implication of this isomorphism is that we cannot distinguish the two factors using observed series of capital and value-added alone. This challenge also applies to the case when other factors are present, though the mapping between the two is more complicated (and affects the other parameters as well). We detail this mapping in Appendix G.

What about the contribution to *arpk* dispersion? Table 5 reports the results for Chinese firms for two values of γ_k, namely -0.18 and -0.36 (these values imply effective curvatures $\alpha+\gamma_k$ equal to one-quarter and one-half of the true α, respectively). The table shows two key results: first, a more negative γ_k reduces the estimated γ (i.e., makes it less negative), suggesting that our baseline estimates of correlated factors could be picking up such size-dependent policies. The total contribution of both types of distortions is quite stable, ranging between 40% and 47%. Second, the estimates of adjustment costs remain modest over this wide range of curvature.

Table 5: Size vs Productivity-Dependent Factors

	Correlated Factors			
	Size-Dependent γ_k	Prod.-Dependent γ	Total	Adj. Costs ξ
$\alpha+\gamma_k = 0.71$ (baseline)				
Parameters	0.00	-0.70		0.13
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$	0.0%	47.4%	47.4%	1.3%
$\alpha+\gamma_k = 0.54$				
Parameters	-0.18	-0.51		0.21
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$	14.2%	25.4%	39.6%	2.3%
$\alpha+\gamma_k = 0.36$				
Parameters	-0.36	-0.33		0.29
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$	29.6%	10.2%	39.8%	3.2%

5.3 Financial Frictions

In this section, we show that liquidity considerations can lead to size-dependent distortions of the form analyzed in the previous subsection. We assume that firms face a cost $\Upsilon\left(K_{it+1}, B_{it+1}\right)$, where B_{it+1} denotes holdings of liquid assets, which earn an exogenous rate of return $R < \frac{1}{\beta}$. The cost is increasing (decreasing) in K_{it+1} (B_{it+1}). This specification captures the idea that firms need costly liquidity in order to operate (e.g., to meet working capital needs). Using a

continuous penalty function rather than an occasionally binding constraint allows us to continue using perturbation methods. Note also that this differs from the standard borrowing constraint used widely in the literature on financial frictions. Our firms are not constrained in terms of their ability to raise funds. This implies that self-financing, which often significantly weakens the long-run bite of borrowing constraints, plays no role here.[45]

We use the following flexible functional form for the liquidity cost:

$$\Upsilon\left(K_{it+1}, B_{it+1}\right) = \hat{\nu}\frac{K_{it+1}^{\omega_1}}{B_{it+1}^{\omega_2}} ,$$

where $\hat{\nu}$, ω_1 and ω_2 are all positive parameters. The marginal liquidity cost of capital, after optimizing over the choice of B_{it+1} is given by (derivations in Appendix H)

$$\Upsilon_{1,t+1} \equiv \frac{d\Upsilon\left(K_{it+1}, B_{it+1}\right)}{dK_{it+1}} = \nu\left(1-\beta R\right)^{\frac{\omega_2}{\omega_2+1}} K_{it+1}^{\omega} , \tag{20}$$

where ν and ω are composite parameters. The former is always positive, while the latter is of indeterminate sign. If ω is positive (negative), the marginal cost of liquidity is increasing (decreasing) in K_{it+1}.

The log-linearized Euler equation takes the same form as (19), with

$$\gamma_k = -\omega\left(\frac{\bar{\Upsilon}_1}{\bar{\Upsilon}_1+\kappa}\right) , \tag{21}$$

where $\bar{\Upsilon}_1$ is the marginal cost of liquidity in the deterministic steady state and $\kappa = 1-\beta(1-\delta)+\hat{\xi}\delta(1-\beta(1-\frac{\delta}{2}))$. Intuitively, the fraction $\frac{\bar{\Upsilon}_1}{\bar{\Upsilon}_1+\kappa}$ is the steady state share of liquidity in the total marginal cost of capital. Thus, liquidity considerations manifest themselves as a size-dependent factor of the form described in Section 5.2. The sign depends on the sign of ω: if $\omega > 0$, then $\gamma_k < 0$, so costly liquidity dampens incentives to adjust capital in response to productivity (since the liquidity cost is convex). The opposite happens if $\omega < 0$.

Thus, liquidity considerations are a promising candidate for correlated and/or size-dependent factors. Cross-country differences in liquidity requirements (summarized by the parameters ν and ω) and/or costs (i.e., $1-\beta R$) will translate into variation in the severity of our measures of correlated firm-specific factors. However, our results here also highlight the difficulty in separating them from other factors using production-side data alone. One would need additional data, e.g., on firm-level liquidity holdings, to disentangle the role of liquidity from other forces.

[45]See, for example, Midrigan and Xu (2014) and Moll (2014). Gopinath et al. (2017) show that a richer variant of the standard collateral constraint can have important implications during a period of transition, even if it generates only modest amounts of *arpk* dispersion in the long-run.

In sum, our findings in sections 5.1-5.3 provide some guidance on the factors beyond adjustment and information frictions that influence investment decisions. For US publicly traded firms, observed dispersion in value-added/capital ratios could be driven to a significant extent by unobserved heterogeneity in production technologies and therefore, as we emphasized above, may not be a sign of misallocated capital. On the other hand, the scope for this type of heterogeneity appears limited among Chinese manufacturing firms, suggesting a greater role for inefficient factors like size-dependent policies or financial imperfections.

6 Robustness and Extensions

In this section, we explore a number of variants on our baseline approach. We generalize our specification of adjustment costs to include a non-convex component. We also use this exercise to assess the accuracy of the log-linearized solution, since this case requires nonlinear solution techniques. We consider the implications of a frictional labor choice. We also explore a number of measurement concerns, including the potential for measurement error. Appendix I contains additional extensions and robustness exercises – alternative stochastic processes on productivity and distortions, different assumptions on the elasticity of substitution and variants on the set of target moments. Our main conclusions about the relative contribution of various factors to observed *arpk* dispersion is robust across these exercises.

6.1 Non-Convex Adjustment Costs

Our baseline specification with only convex adjustment costs allowed us to use perturbation techniques to solve and estimate the model, which yielded both analytical tractability for our identification arguments and computational efficiency. However, it raises two questions: one, how well does the log-linearized version approximate the true solution? And two, are the results robust to allowing for non-convex adjustment costs? In this section, we address both of these concerns. We modify the adjustment cost function to include a non-convex component:

$$\Phi\left(K_{it+1}, K_{it}\right) = \frac{\hat{\xi}}{2}\left(\frac{K_{it+1}}{K_{it}} - (1-\delta)\right)^2 K_{it} + \hat{\xi}_f \mathbb{I}\left\{I_{it} \neq 0\right\} \pi\left(A_{it}, K_{it}\right) \ ,$$

where $I_{it} = K_{it+1} - (1-\delta) K_{it}$ denotes period t investment and $\mathbb{I}\{\cdot\}$ the indicator function. The adjustment cost is now composed of two components: the first is a quadratic term, the same as before. The second is a fixed cost, which is parameterized by $\hat{\xi}_f$, the fraction of profits

that must be incurred if the firm undertakes any non-zero investment.[46]

Due to the fixed cost, we can no longer use perturbation methods. We therefore solve the model by value function iteration and re-estimate the parameters using simulated method of moments. Note that there is an additional parameter in this version relative to the baseline, $\hat{\xi}_f$. To pin this down, we add a new target moment: inaction, defined as the fraction of firms with (gross) investment rates of less than 5% in absolute value. This value is equal to 20% of firms in the Chinese data and 18% of firms in the US.[47] Formally, we estimate the model by searching over the six parameters, $\hat{\xi}, \hat{\xi}_f, \mathbb{V}, \gamma, \sigma_\varepsilon^2$ and σ_χ^2, to find the combination that minimizes the equally-weighted sum of squared deviations of the model-implied values for the six target moments – the five moments from Table 2 and inaction – from their empirical counterparts.[48]

The results are reported in Table 6. The estimated value for the fixed cost, $\hat{\xi}_f$, is modest in both countries, about 0.2% of annual profits. The other parameters and their contributions to σ^2_{arpk} are quite close to the baseline estimates. These results demonstrate that (1) allowing for non-convex adjustment costs does not alter our main conclusions regarding the sources of *arpk* dispersion and (2) the perturbation approach produces reasonably accurate estimates.[49]

Table 6: Non-Convex Adjustment Costs

Parameters	$\hat{\xi}$	(ξ)	$\hat{\xi}_f$	$\mathbb{V}$	γ	σ_ε^2	σ_χ^2
China	0.075	(0.51)	0.002	0.09	−0.64	0.00	0.44
US	0.250	(1.70)	0.002	0.03	−0.30	0.02	0.29
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$							
China	6.5%		0.8%	10.1%	35.6%	0.0%	47.7%
US	13.0%		1.1%	7.1%	11.5%	4.4%	64.4%

Notes: The second column (in parentheses) reports the value of the normalized adjustment cost parameter, ξ, for purposes of comparison to Table 3. The mapping between ξ and $\hat{\xi}$ is given in expression (22).

Our estimates for the fixed adjustment cost are lower than many previous estimates in the literature.[50] The primary reason for this difference is the fact that we explicitly control for other

[46]The scaling with profits is a common formulation in the literature – see, e.g., Asker et al. (2014) – and ensures that the fixed cost does not become negligible for large firms.

[47]Appendix I.1 shows the results are robust to using alternative moments to pin down the non-convex component, for example, investment 'spikes.'

[48]We provide further details of the estimation technique and results in Appendix I.1. Table 12 in that appendix reports the fit of the model with respect to both targeted moments and non-targeted moments.

[49]Note that if $\hat{\xi} = \hat{\xi}_f = 0$, there is no approximation involved under the perturbation approach, i.e., the model is exactly log-linear. This property, along with the modest estimates for adjustment costs, is the main reason why the approximation works reasonably well in this region of the parameter space.

[50]For example, Bloom (2009) estimates a fixed adjustment cost of 1% of annual sales for US Compustat firms. Asker et al. (2014) and Cooper and Haltiwanger (2006) work with data on US manufacturing firms and

factors and target a broader set of moments – indeed, our results here further underscore the importance of doing so. In Appendix I.1, we explore this finding in greater depth by estimating a number of variants of our model with only adjustment costs, i.e., abstracting from other factors. This typically yields larger estimates of these costs, but at the expense of counterfactual implications for other, non-targeted moments. For example, a strategy which fits the low serial correlation of investment (e.g., Cooper and Haltiwanger (2006)) yields much larger fixed costs (and smaller convex ones), but implies counterfactually high levels of investment variability and, even more strikingly, inaction. Importantly, however, even under this approach, adjustment costs generate only modest dispersion in *arpk*. Conversely, a strategy which matches the low variability of investment growth (along the lines of Asker et al. (2014)) implies larger convex costs but significantly over-predicts the autocorrelation of investment (and similarly leaves much of the *arpk* dispersion unexplained). Finally, a strategy that jointly fits both the serial correlation and variability yields larger costs of both types, but misses widely on other moments: for example, the predicted degree of inaction is extremely high relative to the data.

These patterns lead our estimation to ascribe an important role to other distortionary factors, even after allowing for non-convexities – these factors reduce investment volatility without increasing its serial correlation and so reconcile these two moments. Further, in conjunction with the estimated level of adjustment costs, the model still performs well on additional moments such as inaction and investment spikes (see Appendix I.1 for details).

6.2 Frictional Labor

Our baseline analysis makes the rather stark assumption of no adjustment or information frictions in labor choice, making it a static decision with full information. Although not uncommon in the literature, this may not be a good description of labor markets. Here, we depart from this assumption and assume that labor is subject to the same forces as capital – adjustment and informational frictions and other factors. In Appendix E.2, we show that, under these conditions, the firm's investment problem takes the same form as in expression (3), but with a modified curvature parameter of $\alpha = \alpha_1 + \alpha_2$ (and appropriately re-defined G and A_{it}). With this re-definition, our identification strategy goes through unchanged. Table 7 reports results for Chinese firms under this specification. The top panel shows the target moments recomputed under this assumption. A comparison to Table 2 reveals that assuming frictional labor raises the correlation of investment with lagged shocks as well as the correlation of the *arpk* with productivity. The second panel reports the associated parameter estimates. They imply higher adjustment costs, greater uncertainty and more severe correlated distortions. As a result, a

estimate this parameter at 12.5% of annual output and 4% of the capital stock, respectively.

lower level of the permanent factor, σ^2_χ, is needed to match σ^2_{arpk}.

The bottom panel of Table 7 reports the contribution of each factor to total *arpk* dispersion and computes the implications for aggregate TFP. There is a noticeable increase in the impact of adjustment costs from the baseline case – now, they account for almost 13% of *arpk* dispersion in China (compared to 1% above). There is also a slight increase in the impact of uncertainty (from 10% to 11%). Further, the effects on aggregate productivity are much larger than in the baseline scenario – here, these forces distort both inputs into production. Adjustment costs and imperfect information now lead to TFP losses of about 36% and 32%, respectively. Thus, this version of our model illustrates the potential for large aggregate consequences of adjustment/information frictions. However, despite the increased impact of these forces (in both relative and absolute terms), the results also confirm a key finding from before, namely, the important role of other correlated and permanent factors. Indeed, these factors compose about 80% of the measured *arpk* dispersion, leading to TFP gaps relative to the first-best of about 144% and 90%, respectively.

Table 7: Frictional Labor - China

Moments	ρ	σ^2_μ	$\rho_{\iota,a_{-1}}$	$\rho_{\iota,\iota_{-1}}$	$\rho_{arpk,a}$	σ^2_ι	σ^2_{arpk}
	0.92	0.16	0.33	−0.36	0.81	0.14	0.94
Parameters			ξ	$\mathbb{V}$	γ	σ^2_ε	σ^2_χ
			0.78	0.11	−0.68	0.04	0.30
Aggregate Effects							
$\Delta\sigma^2_{arpk}$			0.12	0.11	0.48	0.04	0.30
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$			12.8%	11.3%	51.2%	4.0%	32.2%
Δa			0.36	0.32	1.44	0.11	0.90

6.3 Measurement Error

Measurement error is an important and challenging concern for the misallocation literature more broadly. In an important recent contribution, Bils et al. (2017) propose a method to estimate the role of additive measurement error. Here, we apply their methodology to our data. It essentially involves estimating the following regression:

$$\Delta va_{it} = \Phi arpk_{it} + \Psi \Delta k_{it} - \Psi\left(1-\lambda\right) arpk_{it} \cdot \Delta k_{it} + D_{jt} + \epsilon_{it} \ ,$$

where Δva_{it} and Δk_{it} denote changes in (log) value-added and capital respectively, D_{jt} is a full set of industry-year fixed effects and $arpk_{it}$ is (the log of the) average revenue product of

capital. The key object is the coefficient on the interaction term. Bils et al. (2017) show that, under certain assumptions, λ is the ratio of the true dispersion in the *arpk* to its measured counterpart (and inversely, $1-\lambda$ is the contribution of measurement error to the observed σ^2_{arpk}). Intuitively, to the extent measured *arpk* deviations are due to additive measurement error, value-added of firms with high observed *arpk* will display a lower elasticity w.r.t. capital.

Estimating this regression in our data yields estimates for λ of 0.92 in China and 0.88 in the US. These values suggest that, in both countries, only about 10% of the observed σ^2_{arpk} can be accounted for by additive measurement error. Of course, it must be pointed out that this method is silent about other forms of measurement error (e.g., multiplicative).[51]

6.4 Additional Measurement Concerns

In this subsection, we address two other measurement-related issues. The first stems from our use of book values for capital. Although this is a common approach in the misallocation literature, e.g., Hsieh and Klenow (2009) and Gopinath et al. (2017), other papers use the perpetual inventory method along with data on investment good price deflators to construct an alternative measure for capital. To address this concern, we compute firm-level capital stocks for US firms, where data on the relevant price indices are readily available, using the approach outlined in Eberly et al. (2012). The results from re-estimating the model using these measures, presented in Appendix I.5, are broadly in line with our baseline findings. They point to a somewhat larger role for adjustment costs (the autocorrelation of investment growth is higher under this method and the variance lower, leading to a higher estimate of ξ), which account for about 27% of total σ^2_{arpk} (compared to 11% under our baseline approach). The contribution of uncertainty is essentially unchanged at about 6%. Importantly, other firm-specific factors continue to play a key role in generating the observed *arpk* dispersion.

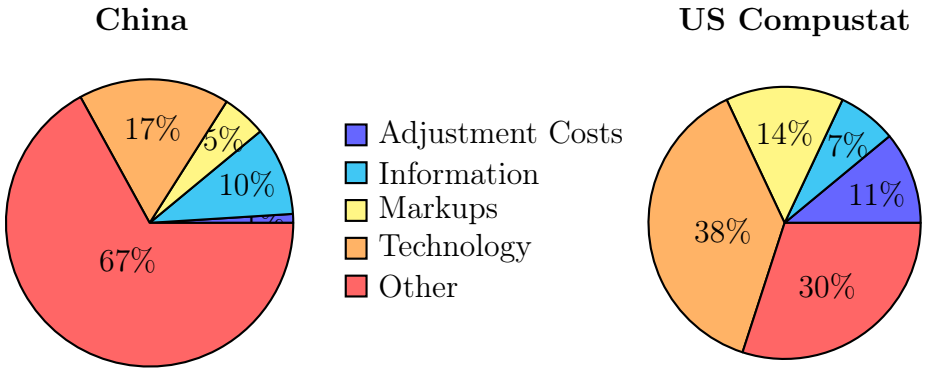
The second concern relates to sectoral heterogeneity in the structural parameters. We have estimated our model separately for US firms for the 9 major sectors of the industrial classification (e.g., manufacturing, construction, services, etc.). Specifically, we allowed for sector-specific parameters in production (we infer sector-specific α's using sectoral labor shares obtained from the BEA), adjustment frictions, uncertainty, as well as other factors. The details of this procedure are outlined in Appendix I.6 and the results are presented in Table 20 in that appendix. Although there is some variation across sectors, the overall patterns in the role of various factors (bottom panel of that table) are similar to those from our baseline analysis. The contribution of adjustment costs to observed *arpk* dispersion is generally modest – the highest

[51] There are a few approaches in the literature to deal with multiplicative measurement error, e.g. Collard-Wexler and De Loecker (2016) and Song and Wu (2015) make some progress on this dimension after imposing additional structure.

contribution is about 20% of σ^2_{arpk} in Manufacturing and the lowest is 2% in Finance, Insurance and Real Estate. Uncertainty accounts for 5-10% across sectors, leaving the bulk of observed *arpk* dispersion within each sector to be accounted for by other factors.

7 Conclusion

In this paper, we have laid out a model of investment featuring multiple factors that interfere with the equalization of static capital products, along with an empirical strategy to disentangle them using widely available firm-level production data. Figure 3 summarizes our results on the sources of *arpk* dispersion in China (left panel) and the US (right panel). They show that much of the *arpk* dispersion stems not from adjustment and informational frictions, but from other firm-specific factors, either systematically correlated with firm productivity/size or almost permanent. Moreover, unobserved heterogeneity in demand and production technologies can potentially account for a significant portion of observed *arpk* dispersion in the US, but not in China, where size-dependent policies and/or financial imperfections may be more fruitful avenues to pursue. Crucially, analyzing these forces in isolation would have led to very different conclusions, highlighting the value of using a unified framework and empirical approach.



Notes: The numbers for the contribution of technological dispersion denote the upper bound as calculated in footnote 43.

Figure 3: The Sources of 'Misallocation'

There are several promising directions for future work. A key message of our analysis is that although various frictions/distortions/policies may all generate dispersion in input products, they often have different effects on various moments of the data, helping to tease out their individual contributions. This insight, along with our quantitative findings, should be useful both in identifying candidate factors driving that dispersion and in guiding empirical strategies to measure their impact. For example, policies and/or frictions that introduce persistent wedges into firms' investment decisions would be very promising; ones that are transitory, particularly

those unrelated to firm size/productivity, less so. On the empirical side, our formulation and findings on these factors point to a strategy for investigating specific forces even while controlling for others – thereby reaching more accurate estimates of their contributions – in a tractable, albeit reduced-form, way. For example, one recent paper that builds on our results – both methodological and substantive – is David, Schmid, and Zeke (2018). First, they propose a theory of firm-level risk premia that delivers a firm-specific fixed wedge in the capital choice. Second, they verify that their identification strategy is robust to the presence of other distortions of the same form as we lay out here.

Our analysis focused primarily on capital allocation, but a natural extension is to use similar methods to study the allocation of labor. Such an analysis holds much promise, both for understanding the role of various forces (e.g., adjustment, informational or other) in explaining observed dispersion in labor products, and further, in narrowing the list of candidate factors driving input allocations more broadly. As we showed in Section 5.1, combining data on multiple inputs can help shed light on the nature of distortionary factors – for example, the correlation between capital and labor products can be very useful in disciplining the potential for 'scale' factors that distort all input products in the same direction vs 'mix' factors that distort the capital-ratio ratio.

Our findings have implications beyond static *arpk* dispersion. Midrigan and Xu (2014) show that the same factors behind static variation in input products can have larger effects on aggregate outcomes by influencing entry and exit decisions. Similarly, a number of recent papers examine the impact of distortions on the life-cycle of the firm and the distribution of productivity itself, e.g., Hsieh and Klenow (2014), Bento and Restuccia (2017) and Da-Rocha et al. (2017). An important insight from these papers is that the exact nature of the underlying distortions (e.g., their correlation with firm productivity or demand) is key to understanding their dynamic implications. An ambitious next step would be to use an empirical strategy like the one in this paper to analyze richer environments featuring some of these elements.

References

Asker, J., A. Collard-Wexler, and J. De Loecker (2014): "Dynamic Inputs and Resource (Mis)Allocation," *Journal of Political Economy*, 122, 1013–1063.

Atkeson, A. and P. J. Kehoe (2005): "Modeling and measuring organization capital," *Journal of Political Economy*, 113, 1026–1053.

Bachmann, R. and S. Elstner (2015): "Firm Optimism and Pessimism," *European Economic Review*, 79, 297–325.

Bai, C.-E., C.-T. Hsieh, and Y. Qian (2006): "The Return to Capital in China," *Brookings Papers on Economic Activity*, 2006, 61–101.

Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2013): "Cross Country Differences in Productivity: The Role of Allocative Efficiency," *American Economic Review*, 103, 305–334.

Bento, P. and D. Restuccia (2017): "Misallocation, establishment size, and productivity," *American Economic Journal: Macroeconomics*, 9, 267–303.

Bils, M., P. J. Klenow, and C. Ruane (2017): "Misallocation or Mismeasurement?" *Working Paper*.

Bloom, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77, 623–685.

Brandt, L., J. Van Biesebroeck, and Y. Zhang (2014): "Challenges of Working with the Chinese NBS Firm-Level Data," *China Economic Review*, 30, 339–352.

Broda, C. and D. E. Weinstein (2006): "Globalization and the Gains from Variety," *The Quarterly journal of economics*, 121, 541–585.

Buera, F. J. and R. N. Fattal-Jaef (2016): "The Dynamics of Development: Innovation and Reallocation," *working paper*.

Buera, F. J., J. P. Kaboski, and Y. Shin (2011): "Finance and Development: A Tale of Two Sectors," *American Economic Review*, 101, 1964–2002.

Buera, F. J., B. Moll, and Y. Shin (2013): "Well-Intended Policies," *Review of Economic Dynamics*, 16, 216–230.

Collard-Wexler, A. and J. De Loecker (2016): "Production Function Estimation with Measurement Error in Inputs," Tech. rep., National Bureau of Economic Research.

Cooper, R., G. Gong, and P. Yan (2015): "Dynamic Labor Demand in China: Public and Private Objectives," *The RAND Journal of Economics*, 46, 577–610.

Cooper, R. W. and J. C. Haltiwanger (2006): "On the Nature of Capital Adjustment Costs," *The Review of Economic Studies*, 73, 611–633.

Da-Rocha, J.-M., M. M. Tavares, and D. Restuccia (2017): "Policy Distortions and Aggregate Productivity with Endogenous Establishment-Level Productivity," Tech. rep., National Bureau of Economic Research.

David, J. M., H. A. Hopenhayn, and V. Venkateswaran (2016): "Information, Misallocation, and Aggregate Productivity," *The Quarterly Journal of Economics*, 131, 943–1005.

David, J. M., L. Schmid, and D. Zeke (2018): "Risk-Adjusted Capital Allocation and Misallocation," Working paper, Duke University.

De Loecker, J. and J. Eeckhout (2017): "The rise of market power and the macroeconomic implications," Tech. rep., National Bureau of Economic Research.

De Loecker, J. and F. Warzynski (2012): "Markups and Firm-Level Export Status," *The American Economic Review*, 102, 2437–2471.

Eberly, J., S. Rebelo, and N. Vincent (2012): "What Explains the Lagged-Investment Effect?" *Journal of Monetary Economics*, 59, 370–380.

Eslava, M., J. Haltiwanger, A. Kugler, and M. Kugler (2004): "The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia," *Journal of Development Economics*, 75, 333–371.

Gopinath, G., Ş. Kalemli-Özcan, L. Karabarbounis, and C. Villegas-Sanchez (2017): "Capital Allocation and Productivity in South Europe," *The Quarterly Journal of Economics*, 132, 1915–1967.

Guner, N., G. Ventura, and Y. Xu (2008): "Macroeconomic Implications of Size-Dependent Policies," *Review of Economic Dynamics*, 11, 721–744.

Hopenhayn, H. A. (2014): "Firms, Misallocation, and Aggregate Productivity: A Review," *Annu. Rev. Econ.*, 6, 735–770.

Hsieh, C. and P. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124, 1403–1448.

Hsieh, C.-T. and P. J. Klenow (2014): "The Life Cycle of Plants in India and Mexico," *The Quarterly Journal of Economics*, 129, 1035–1084.

İmrohoroğlu, A. and Ş. Tüzel (2014): "Firm-Level Productivity, Risk, and Return," *Management Science*, 60, 2073–2090.

Jurado, K., S. C. Ludvigson, and S. Ng (2015): "Measuring Uncertainty," *The American Economic Review*, 105, 1177–1216.

Kehrig, M. and N. Vincent (2017): "Do Firms Mitigate or Magnify Capital Misallocation? Evidence from Planet-Level Data," *Working Paper*.

Keller, W. and S. R. Yeaple (2009): "Multinational enterprises, international trade, and productivity growth: firm-level evidence from the United States," *The Review of Economics and Statistics*, 91, 821–831.

Khan, A. and J. K. Thomas (2008): "Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics," *Econometrica*, 76, 395–436.

Klenow, P. J. and J. L. Willis (2007): "Sticky Information and Sticky Prices," *Journal of Monetary Economics*, 54, 79–99.

Midrigan, V. and D. Y. Xu (2014): "Finance and Misallocation: Evidence from Plant-Level Data," *The American Economic Review*, 104, 422–458.

Moll, B. (2014): "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?" *The American Economic Review*, 104, 3186–3221.

Morck, R., A. Shleifer, R. W. Vishny, M. Shapiro, and J. M. Poterba (1990): "The Stock Market and Investment: Is the Market a Sideshow?" *Brookings Papers on Economic Activity*, 1990, 157–215.

Peters, M. (2016): "Heterogeneous Mark-Ups, Growth and Endogenous Misallocation," *Working Paper*.

Restuccia, D. and R. Rogerson (2008): "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments," *Review of Economic Dynamics*, 11, 707–720.

——— (2017): "The Causes and Costs of Misallocation," Tech. rep., National Bureau of Economic Research.

Song, Z. and G. L. Wu (2015): "Identifying Capital Misallocation," *Working Paper*.

Tybout, J. R. and M. D. Westbrook (1995): "Trade Liberalization and the Dimensions of Efficiency Change in Mexican Manufacturing Industries," *Journal of International Economics*, 39, 53–78.

Appendix: For Online Publication

A Baseline Model

This appendix provides detailed derivations and proofs for our baseline model.

A.1 Solution

The first order condition and envelope conditions associated with (3) are, respectively,

$$\begin{aligned} T^K_{it+1}\left(1-\beta\left(1-\delta\right)\right)+\Phi_1\left(K_{it+1},K_{it}\right) &= \beta\mathbb{E}_{it}\left[\mathcal{V}_1\left(K_{it+1},\mathcal{I}_{it+1}\right)\right] \\ \mathcal{V}_1\left(K_{it},\mathcal{I}_{it}\right) &= \Pi_1\left(K_{it},A_{it}\right)-\Phi_2\left(K_{it+1},K_{it}\right) \end{aligned}$$

and combining yields the Euler equation

$$\mathbb{E}_{it}\left[\beta\Pi_1\left(K_{it+1},A_{it+1}\right)-\beta\Phi_2\left(K_{it+2},K_{it+1}\right)-T^K_{it+1}\left(1-\beta\left(1-\delta\right)\right)-\Phi_1\left(K_{it+1},K_{it}\right)\right]=0$$

where

$$\begin{aligned} \Pi_1\left(K_{it+1},A_{it+1}\right) &= \alpha G A_{it+1}K^{\alpha-1}_{it+1} \\ \Phi_1\left(K_{it+1},K_{it}\right) &= \hat{\xi}\left(\frac{K_{it+1}}{K_{it}}-\left(1-\delta\right)\right) \\ \Phi_2\left(K_{it+1},K_{it}\right) &= -\hat{\xi}\left(\frac{K_{it+1}}{K_{it}}-\left(1-\delta\right)\right)\frac{K_{it+1}}{K_{it}}+\frac{\hat{\xi}}{2}\left(\frac{K_{it+1}}{K_{it}}-\left(1-\delta\right)\right)^2 \\ &= \frac{\hat{\xi}}{2}\left(1-\delta\right)^2-\frac{\hat{\xi}}{2}\left(\frac{K_{it+1}}{K_{it}}\right)^2 \end{aligned}$$

In the undistorted $\left(\bar{T}^K=1\right)$ non-stochastic steady state, these are equal to

$$\begin{aligned} \bar{\Phi}_1 &= \hat{\xi}\delta \\ \bar{\Phi}_2 &= \frac{\hat{\xi}}{2}\left(1-\delta\right)^2-\frac{\hat{\xi}}{2} \\ \bar{\Pi}_1 &= \alpha\bar{G}\bar{A}\bar{K}^{\alpha-1} \end{aligned}$$

Log-linearizing the Euler equation around this point yields

$$\mathbb{E}_{it}\left[\beta\bar{\Pi}_1\pi_{1,it+1}-\beta\bar{\Phi}_2\phi_{2,it+1}-\tau^K_{it+1}\left(1-\beta\left(1-\delta\right)\right)-\bar{\Phi}_1\phi_{1,it}\right]=0$$

where $\tau_{it+1}^K = \log T_{it+1}^K$ and

$$\begin{aligned}
\bar{\Pi}_1 \pi_{1,it+1} &\approx \alpha \bar{G} \bar{A} \bar{K}^{\alpha-1} \left(a_{it+1} + (\alpha - 1) k_{it+1} \right) \\
\bar{\Phi}_1 \phi_{1,it} &\approx \hat{\xi} \left(k_{it+1} - k_{it} \right) \\
\bar{\Phi}_2 \phi_{2,it+1} &\approx -\hat{\xi} \left(k_{it+2} - k_{it+1} \right)
\end{aligned}$$

Rearranging gives

$$k_{it+1} \left((1+\beta)\xi + 1 - \alpha \right) = \mathbb{E}_{it} \left[a_{it+1} + \tau_{it+1} \right] + \beta \xi \mathbb{E}_{it} \left[k_{it+2} \right] + \xi k_{it}$$

where

$$\xi = \frac{\hat{\xi}}{\beta \bar{\Pi}_1}, \qquad \tau_{it+1} = -\frac{1 - \beta(1-\delta)}{\beta \bar{\Pi}_1} \tau_{it+1}^K$$

which is expression (4) in the text. Using the steady state Euler equation,

$$\beta(\bar{\Pi}_1 + 1 - \delta) - \beta \bar{\Phi}_2 = 1 + \bar{\Phi}_1 \quad \Rightarrow \quad \alpha \beta \bar{G} \bar{A} \bar{K}^{\alpha-1} = 1 - \beta(1-\delta) + \hat{\xi} \delta \left(1 - \beta \left(1 - \frac{\delta}{2} \right) \right)$$

we have

$$\begin{aligned}
\xi &= \frac{\hat{\xi}}{1 - \beta(1-\delta) + \hat{\xi} \delta \left(1 - \beta \left(1 - \frac{\delta}{2} \right) \right)} \\
\tau_{it+1} &= -\frac{1 - \beta(1-\delta)}{1 - \beta(1-\delta) + \hat{\xi} \delta \left(1 - \beta \left(1 - \frac{\delta}{2} \right) \right)} \tau_{it+1}^K
\end{aligned} \tag{22}$$

To derive the investment policy function, we conjecture that it takes the form in (7). Then,

$$\begin{aligned}
k_{it+2} &= \psi_1 k_{it+1} + \psi_2 (1+\gamma) \mathbb{E}_{it+1} a_{it+2} + \psi_3 \varepsilon_{it+2} + \psi_4 \chi_i \\
\mathbb{E}_{it} \left[k_{it+2} \right] &= \psi_1 k_{it+1} + \psi_2 (1+\gamma) \rho \mathbb{E}_{it} \left[a_{it+1} \right] + \psi_4 \chi_i \\
&= \psi_1 \left(\psi_1 k_{it} + \psi_2 (1+\gamma) \mathbb{E}_{it} \left[a_{it+1} \right] + \psi_3 \varepsilon_{it+1} + \psi_4 \chi_i \right) + \psi_2 (1+\gamma) \rho \mathbb{E}_{it} \left[a_{it+1} \right] + \psi_4 \chi_i \\
&= \psi_1^2 k_{it} + (\psi_1 + \rho) \psi_2 (1+\gamma) \mathbb{E}_{it} \left[a_{it+1} \right] + \psi_1 \psi_3 \varepsilon_{it+1} + \psi_4 (1 + \psi_1) \chi_i
\end{aligned}$$

where we have used $\mathbb{E}_{it} \left[\varepsilon_{it+2} \right] = 0$ and $\mathbb{E}_{it} \left[\mathbb{E}_{it+1} \left[a_{it+2} \right] \right] = \rho \mathbb{E}_{it} \left[a_{it+1} \right]$. Substituting and rearranging,

$$\begin{aligned}
& \left(1 + \beta \xi \psi_4 (1 + \psi_1) \right) \chi_i + \left(1 + \beta \xi \psi_1 \psi_3 \right) \varepsilon_{it+1} \\
+ \; & \left(1 + \beta \xi (\psi_1 + \rho) \psi_2 \right) (1+\gamma) \mathbb{E}_{it} \left[a_{it+1} \right] + \xi \left(1 + \beta \psi_1^2 \right) k_{it} \\
= \; & \left((1+\beta)\xi + 1 - \alpha \right) \left(\psi_1 k_{it} + \psi_2 (1+\gamma) \mathbb{E}_{it} \left[a_{it+1} \right] + \psi_3 \varepsilon_{it+1} + \psi_4 \chi_i \right)
\end{aligned}$$

Finally, matching coefficients gives

$$\begin{aligned}
\xi\left(\beta\psi_1^2+1\right) &= \psi_1\left((1+\beta)\xi+1-\alpha\right) \\
1+\beta\xi\left(\psi_1+\rho\right)\psi_2 &= \psi_2\left((1+\beta)\xi+1-\alpha\right) \Rightarrow \psi_2 = \frac{1}{1-\alpha+\beta\xi\left(1-\psi_1-\rho\right)+\xi} \\
1+\beta\xi\psi_1\psi_3 &= \psi_3\left((1+\beta)\xi+1-\alpha\right) \Rightarrow \psi_3 = \frac{1}{1-\alpha+\left(1-\psi_1\right)\beta\xi+\xi} \\
1+\beta\xi\psi_4\left(1+\psi_1\right) &= \psi_4\left((1+\beta)\xi+1-\alpha\right) \Rightarrow \psi_4 = \frac{1}{1-\alpha+\xi\left(1-\beta\psi_1\right)}
\end{aligned}$$

A few lines of algebra yields the expressions in (8).

A.2 Aggregation

To derive aggregate TFP and output, substitute the firm's optimality condition for labor

$$N_{it} = \left(\frac{\alpha_2 Y^{\frac{1}{\theta}}}{W}\hat{A}_{it}K_{it}^{\alpha_1}\right)^{\frac{1}{1-\alpha_2}}$$

into the production function (1) to get

$$Y_{it} = \left(\frac{\alpha_2 Y^{\frac{1}{\theta}}}{W}\right)^{\frac{\hat{\alpha}_2}{1-\alpha_2}} \hat{A}_{it}^{\frac{\hat{\alpha}_2}{1-\alpha_2}} K_{it}^{\frac{\hat{\alpha}_1}{1-\alpha_2}}$$

and using the demand function, revenues are

$$P_{it}Y_{it} = Y^{\frac{1}{\theta}\frac{1}{1-\alpha_2}}\left(\frac{\alpha_2}{W}\right)^{\frac{\alpha_2}{1-\alpha_2}} A_{it}K_{it}^{\alpha}$$

Labor market clearing implies

$$\int N_{it}di = \int\left(\frac{\alpha_2 Y^{\frac{1}{\theta}}}{W}\right)^{\frac{1}{1-\alpha_2}} A_{it}K_{it}^{\alpha}di = N$$

so that

$$\left(\frac{\alpha_2}{W}\right)^{\frac{\alpha_2}{1-\alpha_2}} = \left(\frac{N}{\int A_{it}K_{it}^{\alpha}di}\frac{1}{Y^{\frac{1}{\theta}\frac{1}{1-\alpha_2}}}\right)^{\alpha_2} \Rightarrow P_{it}Y_{it} = Y^{\frac{1}{\theta}}\frac{A_{it}K_{it}^{\alpha}}{\left(\int A_{it}K_{it}^{\alpha}di\right)^{\alpha_2}}N^{\alpha_2}$$

By definition,

$$ARPK_{it} = \frac{A_{it}K_{it}^{\alpha-1}}{\left(\int A_{it}K_{it}^{\alpha}di\right)^{\alpha_2}}Y^{\frac{1}{\theta}}N^{\alpha_2}$$

so that

$$K_{it} = \left(\frac{Y^{\frac{1}{\theta}} A_{it}}{ARPK_{it}} \right)^{\frac{1}{1-\alpha}} \left(\frac{N}{\int A_{it} K_{it}^{\alpha} di} \right)^{\frac{\alpha_2}{1-\alpha}}$$

and capital market clearing implies

$$K = \int K_{it} di = \left(Y^{\frac{1}{\theta}} \right)^{\frac{1}{1-\alpha}} \left(\frac{N}{\int A_{it} K_{it}^{\alpha} di} \right)^{\frac{\alpha_2}{1-\alpha}} \int A_{it}^{\frac{1}{1-\alpha}} ARPK_{it}^{-\frac{1}{1-\alpha}} di$$

The latter two equations give

$$K_{it}^{\alpha} = \left(\frac{A_{it}^{\frac{1}{1-\alpha}} ARPK_{it}^{-\frac{1}{1-\alpha}}}{\int A_{it}^{\frac{1}{1-\alpha}} ARPK_{it}^{-\frac{1}{1-\alpha}} di} K \right)^{\alpha}$$

Substituting into the expression for $P_{it}Y_{it}$ and rearranging, we can derive

$$P_{it}Y_{it} = \frac{\frac{A_{it}^{\frac{1}{1-\alpha}} ARPK_{it}^{-\frac{\alpha}{1-\alpha}}}{\left(\int A_{it}^{\frac{1}{1-\alpha}} ARPK_{it}^{-\frac{1}{1-\alpha}} di \right)^{\alpha}}}{\left(\frac{\int A_{it}^{\frac{1}{1-\alpha}} ARPK_{it}^{-\frac{\alpha}{1-\alpha}} di}{\left(\int A_{it}^{\frac{1}{1-\alpha}} ARPK_{it}^{-\frac{1}{1-\alpha}} di \right)^{\alpha}} \right)^{\alpha_2}} Y^{\frac{1}{\theta}} K^{\alpha_1} N^{\alpha_2}$$

Using the fact that $Y = \int P_{it}Y_{it} di$, we can derive

$$Y = \int P_{it}Y_{it} di = Y^{\frac{1}{\theta}} A K^{\alpha_1} N^{\alpha_2}$$

where

$$A = \left(\frac{\int A_{it}^{\frac{1}{1-\alpha}} ARPK_{it}^{-\frac{\alpha}{1-\alpha}} di}{\left(\int A_{it}^{\frac{1}{1-\alpha}} ARPK_{it}^{-\frac{1}{1-\alpha}} di \right)^{\alpha}} \right)^{1-\alpha_2}$$

or in logs,

$$a = (1-\alpha_2) \left[\log \left(\int A_{it}^{\frac{1}{1-\alpha}} ARPK_{it}^{-\frac{\alpha}{1-\alpha}} \right) - \alpha \log \left(\int A_{it}^{\frac{1}{1-\alpha}} ARPK_{it}^{-\frac{1}{1-\alpha}} \right) \right]$$

The first term inside brackets is equal to

$$\frac{1}{1-\alpha} \overline{a} - \frac{\alpha}{1-\alpha} \overline{arpk} + \frac{1}{2} \left(\frac{1}{1-\alpha} \right)^2 \sigma_a^2 + \frac{1}{2} \left(\frac{\alpha}{1-\alpha} \right)^2 \sigma_{arpk}^2 - \frac{\alpha}{(1-\alpha)^2} \sigma_{arpk,a}$$

and the second,

$$\frac{\alpha}{1-\alpha}\overline{a} - \frac{\alpha}{1-\alpha}\overline{arpk} + \frac{1}{2}\alpha\left(\frac{1}{1-\alpha}\right)^2\sigma_a^2 + \frac{1}{2}\alpha\left(\frac{1}{1-\alpha}\right)^2\sigma_{arpk}^2 - \frac{\alpha}{(1-\alpha)^2}\sigma_{arpk,a}$$

Combining,

$$a = (1-\alpha_2)\left[\overline{a} + \frac{1}{2}\frac{1}{1-\alpha}\sigma_a^2 - \frac{1}{2}\frac{\alpha}{1-\alpha}\sigma_{arpk}^2\right]$$

and

$$\begin{aligned} y &= \frac{1}{\theta}y + (1-\alpha_2)\,\bar{a} + \frac{1}{2}\frac{1-\alpha_2}{1-\alpha}\sigma_a^2 - \frac{1}{2}\alpha\frac{1-\alpha_2}{1-\alpha}\sigma_{arpk}^2 + \alpha_1 k + \alpha_2 n \\ &= \frac{\theta}{\theta-1}(1-\alpha_2)\,\bar{a} + \frac{\theta}{\theta-1}\frac{1}{2}\frac{1-\alpha_2}{1-\alpha}\sigma_a^2 - \frac{\theta}{\theta-1}\frac{1}{2}\alpha\frac{1-\alpha_2}{1-\alpha}\sigma_{arpk}^2 + \hat{\alpha}_1 k + \hat{\alpha}_2 n \\ &= a + \hat{\alpha}_1 k + \hat{\alpha}_2 n \end{aligned}$$

where, using $a_{it} = \frac{1}{1-\alpha_2}\hat{a}_{it}$, $\sigma_a^2 = \left(\frac{1}{1-\alpha_2}\right)^2\sigma_{\hat{a}}^2$ and $\alpha = \frac{\alpha_1}{1-\alpha_2}$,

$$\begin{aligned} a &= \frac{\theta}{\theta-1}\bar{\hat{a}} + \frac{1}{2}\frac{\theta}{\theta-1}\frac{1}{1-\alpha_1-\alpha_2}\sigma_{\hat{a}}^2 - \frac{1}{2}\left(\theta\hat{\alpha}_1 + \hat{\alpha}_2\right)\hat{\alpha}_1\sigma_{arpk}^2 \\ &= a^* - \frac{1}{2}\left(\theta\hat{\alpha}_1 + \hat{\alpha}_2\right)\hat{\alpha}_1\sigma_{arpk}^2 \end{aligned}$$

which is equation (9) in the text.

To compute the effect on output, notice that the aggregate production function is

$$y = \hat{\alpha}_1 k + \hat{\alpha}_2 n + a$$

so that

$$\begin{aligned} \frac{dy}{d\sigma_{arpk}^2} &= \hat{\alpha}_1\frac{dk}{da}\frac{da}{d\sigma_{arpk}^2} + \frac{da}{d\sigma_{arpk}^2} \\ &= \frac{da}{d\sigma_{arpk}^2}\left(1 + \hat{\alpha}_1\frac{dk}{da}\right) \end{aligned}$$

In the stationary equilibrium, the aggregate marginal product of capital must be a constant, denote it by $\bar{R}$, i.e., $\log\hat{\alpha}_1 + y - k = \bar{r}$ so that

$$k = \frac{1}{1-\hat{\alpha}_1}\left(\log\hat{\alpha}_1 + \hat{\alpha}_2 n + a - \bar{r}\right)$$

and

$$\frac{dk}{da} = \frac{1}{1-\hat{\alpha}_1}$$

Combining,

$$\frac{dy}{d\sigma^2_{arpk}} = \frac{da}{d\sigma^2_{arpk}}\left(1+\frac{\hat{\alpha}_1}{1-\hat{\alpha}_1}\right) = \frac{da}{d\sigma^2_{arpk}}\frac{1}{1-\hat{\alpha}_1}$$

A.3 Identification

In this appendix, we derive analytical expressions for the four moments in the random walk case, i.e., when $\rho = 1$, and prove Proposition 1.

Moments. From expression (7), we have the firm's investment policy function

$$k_{it+1} = \psi_1 k_{it} + \psi_2 (1+\gamma)\,\mathbb{E}_{it}[a_{it+1}] + \psi_3 \varepsilon_{it+1} + \psi_4 \chi_i$$

and substituting for the expectation,

$$k_{it+1} = \psi_1 k_{it} + \psi_2 (1+\gamma)\,(a_{it} + \phi\,(\mu_{it+1} + e_{it+1})) + \psi_3 \varepsilon_{it+1} + \psi_4 \chi_i$$

where $\phi = \frac{\mathbb{V}}{\sigma^2_e}$ so that $1-\phi = \frac{\mathbb{V}}{\sigma^2_\mu}$. Then,

$$\Delta k_{it+1} = \psi_1 \Delta k_{it} + \psi_2 (1+\gamma)\,((1-\phi)\,\mu_{it} + \phi\mu_{it+1} + \phi\,(e_{it+1} - e_{it})) + \psi_3\,(\varepsilon_{it+1} - \varepsilon_{it})$$

We will use the fact that

$$\begin{aligned}
\text{cov}\,(\Delta k_{it+1}, \mu_{it+1}) &= \psi_2 (1+\gamma)\,\phi\sigma^2_\mu \\
\text{cov}\,(\Delta k_{it+1}, e_{it+1}) &= \psi_2 (1+\gamma)\,\phi\sigma^2_e \\
\text{cov}\,(\Delta k_{it+1}, \varepsilon_{it+1}) &= \psi_3 \sigma^2_\varepsilon
\end{aligned}$$

Now,

$$\begin{aligned}
\text{var}\,(\Delta k_{it+1}) &= \psi_1^2 \text{var}\,(\Delta k_{it}) + \psi_2^2 (1+\gamma)^2 (1-\phi)^2 \sigma^2_\mu \\
&+ \psi_2^2 (1+\gamma)^2 \phi^2 \sigma^2_\mu + 2\psi_2^2 (1+\gamma)^2 \phi^2 \sigma^2_e + 2\psi_3^2 \sigma^2_\varepsilon \\
&+ 2\psi_1\psi_2 (1+\gamma)(1-\phi)\,\text{cov}\,(\Delta k_{it}, \mu_{it}) - 2\psi_1\psi_2 (1+\gamma)\,\phi \text{cov}\,(\Delta k_{it}, e_{it}) \\
&- 2\psi_1\psi_3 \text{cov}\,(\Delta k_{it}, \varepsilon_{it})
\end{aligned}$$

where substituting, rearranging and using the fact that the moments are stationary gives

$$\sigma_k^2 \equiv \text{var}\left(\Delta k_{it}\right) = \frac{\left(1+\gamma\right)^2 \psi_2^2 \sigma_\mu^2 + 2\left(1-\psi_1\right)\psi_3^2 \sigma_\varepsilon^2}{1-\psi_1^2}$$

which can be rearranged to yield expression (10). Next,

$$\begin{aligned}
\text{cov}\left(\Delta k_{it+1}, \Delta k_{it}\right) &= \psi_1 \text{var}\left(\Delta k_{it}\right) + \psi_2\left(1+\gamma\right)\left(1-\phi\right)\text{cov}\left(\Delta k_{it}, \mu_{it}\right) \\
&- \psi_2\left(1+\gamma\right)\phi \text{cov}\left(\Delta k_{it}, e_{it}\right) - \psi_3 \text{cov}\left(\Delta k_{it}, \varepsilon_{it}\right) \\
&= \psi_1 \text{var}\left(\Delta k_{it}\right) - \psi_3 \text{cov}\left(\Delta k_{it}, \varepsilon_{it}\right) \\
&= \psi_1 \sigma_k^2 - \psi_3^2 \sigma_\varepsilon^2
\end{aligned}$$

so that

$$\rho_{k,k_{-1}} \equiv \text{corr}\left(\Delta k_{it}, \Delta k_{it-1}\right) = \psi_1 - \psi_3^2 \frac{\sigma_\varepsilon^2}{\sigma_k^2}$$

which is expression (11). Similarly,

$$\begin{aligned}
\text{cov}\left(\Delta k_{it+1}, \Delta a_{it}\right) &= \text{cov}\left(\Delta k_{it+1}, \mu_{it}\right) \\
&= \psi_1 \text{cov}\left(\Delta k_{it}, \mu_{it}\right) + \psi_2\left(1+\gamma\right)\left(1-\phi\right)\sigma_\mu^2 \\
&= \psi_1 \psi_2\left(1+\gamma\right)\phi\sigma_\mu^2 + \psi_2\left(1+\gamma\right)\left(1-\phi\right)\sigma_\mu^2 \\
&= \left(1-\phi\left(1-\psi_1\right)\right)\psi_2\left(1+\gamma\right)\sigma_\mu^2
\end{aligned}$$

and from here it is straightforward to derive

$$\rho_{k,a_{-1}} \equiv \text{corr}\left(\Delta k_{it}, \Delta a_{it-1}\right) = \left[\frac{\mathbb{V}}{\sigma_\mu^2}\left(1-\psi_1\right) + \psi_1\right]\frac{\sigma_\mu \psi_2\left(1+\gamma\right)}{\sigma_k}$$

as in expression (12).

Finally,

$$arpk_{it} = p_{it} + y_{it} - k_{it} = \text{Const} + a_{it} + \alpha k_{it} - k_{it} = \text{Const} + a_{it} - \left(1-\alpha\right)k_{it}$$

so that

$$\Delta arpk_{it} = \Delta a_{it} - \left(1-\alpha\right)\Delta k_{it} = \mu_{it} - \left(1-\alpha\right)\Delta k_{it}$$

which implies

$$\text{cov}\left(\Delta arpk_{it}, \mu_{it}\right) = \left(1 - \left(1-\alpha\right)\left(1+\gamma\right)\psi_2\phi\right)\sigma_\mu^2$$

and

$$\begin{aligned}\lambda_{arpk,a} \equiv \frac{\operatorname{cov}\left(\Delta arpk_{it}, \mu_{it}\right)}{\sigma_\mu^2} &= 1-(1-\alpha)(1+\gamma)\psi_2\phi \\ &= 1-(1-\alpha)(1+\gamma)\psi_2\left(1-\frac{\mathbb{V}}{\sigma_\mu^2}\right)\end{aligned}$$

which is expression (13).

To see that the correlation $\rho_{arpk,a}$ is decreasing in σ_ε^2, we derive

$$\begin{aligned}\operatorname{var}\left(\Delta arpk_{it}\right) &= \sigma_\mu^2+(1-\alpha)^2\sigma_k^2-2(1-\alpha)\operatorname{cov}\left(\Delta k_{it},\mu_{it}\right) \\ &= \sigma_\mu^2+(1-\alpha)^2\left(\frac{\psi_2^2(1+\gamma)^2\sigma_\mu^2+2(1-\psi_1)\psi_3^2\sigma_\varepsilon^2}{1-\psi_1^2}\right)-2(1-\alpha)\psi_2(1+\gamma)\phi\sigma_\mu^2 \\ &= \frac{1}{1-\psi_1^2}\left(\left(\left(1-\psi_1^2\right)(1-2(1-\alpha)(1+\gamma)\psi_2\phi)+(1-\alpha)^2(1+\gamma)^2\psi_2^2\right)\sigma_\mu^2\right) \\ &+ \frac{1}{1-\psi_1^2}\left(2(1-\alpha)^2(1-\psi_1)\psi_3^2\sigma_\varepsilon^2\right)\end{aligned}$$

so

$$\rho_{arpk,a}=\frac{\left(1-(1-\alpha)(1+\gamma)\psi_2\phi\right)\sigma_\mu\sqrt{1-\psi_1^2}}{\sqrt{\left(\left(1-\psi_1^2\right)(1-2(1-\alpha)(1+\gamma)\psi_2\phi)+(1-\alpha)^2(1+\gamma)^2\psi_2^2\right)\sigma_\mu^2+2(1-\alpha)^2(1-\psi_1)\psi_3^2\sigma_\varepsilon^2}}$$

Proof of Proposition 1. Write the variance of investment as

$$\sigma_k^2=\psi_1^2\sigma_k^2+(1+\gamma)^2\psi_2^2\sigma_\mu^2+2(1-\psi_1)\psi_3^2\sigma_\varepsilon^2$$

We can rewrite the last term as a function of an observable moment, the autocovariance of investment, which is given by

$$\sigma_{k,k_{-1}}=\psi_1\sigma_k^2-\psi_3^2\sigma_\varepsilon^2\,. \tag{23}$$

Substituting,

$$\sigma_k^2=\psi_1^2\sigma_k^2+(1+\gamma)^2\psi_2^2\sigma_\mu^2+2(1-\psi_1)\left(\psi_1\sigma_k^2-\sigma_{k,k_{-1}}\right) \tag{24}$$

To eliminate the second term, use the equation for $\lambda_{arpk,a}$ to solve for

$$(1+\gamma)\psi_2\phi=\frac{1-\lambda_{arpk,a}}{1-\alpha}=\tilde{\lambda} \tag{25}$$

where $\tilde{\lambda}$ is a decreasing function of $\lambda_{arpk,a}$ that depends only on the known parameter α. Substituting into the expression for the covariance of investment with the lagged shock, $\sigma_{k,a_{-1}}$,

and rearranging yields

$$(1+\gamma)\,\psi_2 = \frac{\sigma_{k,a_{-1}}}{\sigma_\mu^2} + \tilde{\lambda}\,(1-\psi_1) \tag{26}$$

which is an equation in ψ_1 and observable moments. Substituting into (24) gives

$$\sigma_k^2 = \psi_1^2\sigma_k^2 + \left(\frac{\sigma_{k,a_{-1}}}{\sigma_\mu^2} + \tilde{\lambda}\,(1-\psi_1)\right)^2 \sigma_\mu^2 + 2\,(1-\psi_1)\left(\psi_1\sigma_k^2 - \sigma_{k,k_{-1}}\right)$$

and rearranging, we can derive

$$0 = \left(\hat{\lambda}^2 - 1\right)(1-\psi_1)^2 + 2\left(\hat{\lambda}\rho_{k,a_{-1}} - \rho_{k,k_{-1}}\right)(1-\psi_1) + \rho_{k,a_{-1}}^2 \tag{27}$$

where

$$\hat{\lambda} = \frac{\sigma_\mu}{\sigma_k}\tilde{\lambda} = \frac{\sigma_\mu}{\sigma_k}\left(\frac{1-\lambda_{arpk,a}}{1-\alpha}\right)$$

Equation (27) represents a quadratic equation in a single unknown, $1-\psi_1$, or equivalently, in ψ_1. The solution features one positive root and one negative. The positive root corresponds to the true ψ_1 that represents the solution to the firm's investment policy. The value of ψ_1 pins down the adjustment cost parameter ξ as well as ψ_2 and ψ_3. We can then back out γ from (26), ϕ (and so $\mathbb{V}$) from (25) and finally, σ_ε^2 from (23).

□

B Data

Our Chinese data are from the Annual Surveys of Industrial Production conducted by the National Bureau of Statistics. The data span the period 1998-2009 and are built into a panel following quite closely the method outlined in Brandt et al. (2014). We measure the capital stock as the value of fixed assets and calculate investment as the change in the capital stock relative to the preceding period. We construct firm productivity, a_{it}, as the log of value-added less α multiplied by the log of the capital stock and (the log of) the average product of capital, $arpk_{it}$ as the log of value-added less the log of the capital stock. We compute value-added from revenues using a share of intermediates of 0.5 (our data does not include a direct measure of value-added in all years). Investment growth and changes in productivity are the first differences of the investment and productivity series (in logs) respectively.

To extract the firm-specific variation in our variables, we regress each on a year by time fixed-effect and work with the residual. Industries are defined at the 4-digit level. This eliminates the industry-wide component of each series common to all firms in an industry and time period (as well the aggregate component common across all firms) and leaves only the idiosyncratic

variation. To estimate the parameters governing firm productivity, i.e., the persistence ρ and variance of the innovations σ_μ^2, we perform the autoregression implied by (5), again including industry by year controls. We eliminate duplicate observations (firms with multiple observations within a single year) and trim the 3% tails of each series. We additionally exclude observations with excessively high variability in investment (investment rates over 100%). Our final sample in China consists of 797,047 firm-year observations.

Our US data are from Compustat North America and also spans the period 1998-2009. We measure the capital stock using gross property, plant and equipment. We treat the data in exactly the same manner as just described for the set of Chinese firms. We additionally eliminate firms that are not incorporated in the US and/or do not report in US dollars. Our final sample in the US consists of 34,260 firm-year observations.

Table 8 reports a number of summary statistics from one year of our data, 2009: the number of firms (with available data on sales), the share of GDP they account for, and average sales and capital.

Table 8: Sample Statistics 2009

	No. of Firms	Share of GDP	Avg. Sales ($M)	Avg. Capital ($M)
China	303623	0.65	21.51	8.08
US	6177	0.45	2099.33	1811.35

Materials and labor expenses. For the analyses in Section 5.1, labor input is measured as the wage bill. The wage bill is directly reported in the Chinese data. For the US, we follow, e.g., Keller and Yeaple (2009) and impute a measure of the wage bill as the number of employees multiplied by the average industry wage, calculated using data from the NBER-CES Manufacturing Industry Database (available at `http://www.nber.org/nberces/`; the average industry wage is calculated as total industry-wide payroll divided by total employees). Expenditures on intermediate inputs are reported in the Chinese data. In the US, we construct a measure of intermediates following the method in, e.g., De Loecker and Eeckhout (2017) and İmrohoroğlu and Tüzel (2014). Specifically, intermediate expenditures are calculated as total expenses less labor expenses, where total expenses are calculated as sales less operating income (before depreciation and amortization, Compustat series OIBDP) and labor expenses are measured as described earlier. We can then calculate all the series used in Section 5.1, i.e., the raw and 'markup-adjusted' average revenue products of capital, labor and materials (the inverse of materials' share of revenues). We isolate the firm-specific variation in these series following a similar procedure as described above, i.e., by extracting a full set of industry by time fixed-effects and working with the residual. We trim the 1% tails of each series.

C Computation and Estimation of Baseline Model

In this appendix, we provide details of our numerical estimation procedure and results. We estimate the model via method of moments using the following procedure. For a given set of parameters, we compute the cross-sectional moments of interest using the steady state distribution. To do so, we cast the law of motion (7) in matrix form:

$$
\begin{aligned}
BX_{it} &= CX_{it-1} + DU_{it} \qquad \text{where} \\
X_{it} &= \begin{bmatrix} k_{it} \\ \iota_{it} \\ a_{it} \\ \mathbb{E}_{it-1}a_{it} \end{bmatrix} \qquad U_{it} = \begin{bmatrix} \mu_{it} \\ e_{it} \\ \varepsilon_{it} \\ \chi_i \end{bmatrix} \\
B &= \begin{bmatrix} 1 & 0 & 0 & -\psi_2(1+\gamma) \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad C = \begin{bmatrix} \psi_1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & \rho & 0 \\ 0 & 0 & \rho & 0 \end{bmatrix} \quad D = \begin{bmatrix} 0 & 0 & \psi_3 & \psi_4 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1-\frac{\mathbb{V}}{\sigma_\mu^2} & 1-\frac{\mathbb{V}}{\sigma_\mu^2} & 0 & 0 \end{bmatrix}.
\end{aligned}
$$

Pre-multiplying by B^{-1} yields

$$X_{it} = B^{-1}CX_{it-1} + B^{-1}DU_{it} = \tilde{C}X_{it-1} + \tilde{D}U_{it} \ .$$

The steady state covariance matrix of X_{it}, denoted Σ_X, is then obtained by solving the Lyapunov equation:

$$\Sigma_X = \tilde{C}\Sigma_X\tilde{C}' + \tilde{D}\Sigma_U\tilde{D}' \ ,$$

where Σ_U denotes the covariance matrix of U_{it}. It is straightforward to compute other second moments. For example, to obtain the covariance matrix of $\Delta X_{it} = X_{it} - X_{it-1}$, note that

$$\Delta X_{it} = (\tilde{C} - I)X_{it-1} + \tilde{D}U_{it} \quad \Rightarrow \quad \Sigma_{\Delta X} = (\tilde{C} - I)\Sigma_X(\tilde{C} - I)' + \tilde{D}\Sigma_U\tilde{D}' \ .$$

We then use a non-linear solver to search over the parameter vector $(\xi, \mathbb{V}, \gamma, \sigma_\varepsilon^2, \sigma_\chi^2)$ to minimize the equally-weighted distance between the model and data values for the targeted moments.

Table 9 displays the details of our baseline estimation. In the top panel, we report the target moments computed from the data, along with standard errors (in parentheses) and the simulated model counterparts. The estimated parameter vector is shown in the bottom panel along with standard errors and confidence intervals. The model is able to match the full set of target moments quite closely in both countries (in China, the fit is essentially exact) and the standard errors and confidence intervals indicate that the estimates are quite precise.

Standard errors and confidence intervals are calculated using the following bootstrap procedure: we draw 1,000 random samples (with replacement) from the data (these are block-boostraps, i.e., we resample entire histories of firms). For each re-sampled dataset, we re-calculate the target moments and re-estimate the model parameters. Standard errors are computed as the standard deviations of the resulting distribution of estimated moments and parameters. 95% confidence intervals for the parameters are computed as the 2.5th and 97.5th percentiles of the distributions of the parameter estimates.

D Interactions Between Factors

In the main text (specifically, Table 3), we measured the contribution of each factor in isolation, i.e., setting all other forces to zero. The top panel of Table 10 reproduces those estimates (labeled 'In isolation') and compares them to the case where all the other factors are held fixed at their estimated levels (labeled 'Joint'). The table shows some evidence of interactions, but since adjustment and informational frictions are modest, the numbers are quite similar under both approaches.

E Labor Market Distortions

This appendix presents two tractable versions of our model with labor market distortions. In the first, these are modeled as firm-specific 'taxes' with an arbitrary correlation structure. The second describes the environment from Section 6.2 in the main text, where all the factors acting on investment – adjustment, informational and other – are assumed to apply to the labor decision as well. Under both specifications, the profit function takes the same form as in the baseline analysis with suitably re-defined productivity and curvature. This implies that our identification arguments and empirical strategy go through exactly. More importantly, quantifying the sources of *arpk* dispersion still requires only data on value-added and capital as before.

E.1 Firm-Specific Labor Taxes

Here, we introduce labor market distortions in the form of firm-specific taxes on the cost of labor. These distortions change the interpretation of the profitability shifter, A_{it}, which has implications for the correct measurement strategy of this term. But, apart from this re-interpretation, they do not change our estimates/conclusions about the drivers of *arpk* dispersion.

Table 9: Baseline Esimation - Model Fit

Moments	ρ	σ^2_μ	$\rho_{\iota,a_{-1}}$	$\rho_{\iota,\iota_{-1}}$	$\rho_{arpk,a}$	σ^2_ι	σ^2_{arpk}
China							
Data	0.914	0.146	0.287	−0.359	0.757	0.143	0.922
	(0.001)	(0.000)	(0.001)	(0.001)	(0.002)	(0.000)	(0.002)
Model-Simulated			0.287	−0.375	0.737	0.124	0.914
US							
Data	0.933	0.078	0.126	−0.297	0.547	0.056	0.450
	(0.002)	(0.001)	(0.007)	(0.006)	(0.011)	(0.001)	(0.008)
Model-Simulated			0.126	−0.297	0.547	0.056	0.450
Parameter Estimates			ξ	$\mathbb{V}$	γ	σ^2_ε	σ^2_χ
China			0.132	0.095	−0.704	0.000	0.410
			(0.003)	(0.000)	(0.003)	(0.000)	(0.002)
			[0.127,0.137]	[0.095,0.096]	[−0.709,−0.698]	[0.000,0.000]	[0.406,0.413]
US			1.382	0.033	−0.328	0.029	0.292
			(0.163)	(0.001)	(0.024)	(0.009)	(0.006)
			[1.099,1.724]	[0.031,0.034]	[−0.372,−0.279]	[0.015,0.049]	[0.279,0.304]

Notes: Top panel: rows labeled 'Data' display the empirical values of the moments. Values in parentheses are bootstrapped standard errors. Rows labeled 'Model-Simulated' display the value of the moments from the model simulation at the estimated parameter values. Bottom panel displays the parameter estimates. Values in parentheses are bootstrapped standard errors. Values in brackets are 95% confidence intervals.

Table 10: Interactions Between Factors - US

			Other Factors		
	Adj Costs	Uncertainty	Correlated	Transitory	Permanent
In isolation					
$\Delta\sigma^2_{arpk}$	0.05	0.03	0.06	0.03	0.29
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$	10.8%	7.3%	14.4%	6.3%	64.7%
Joint					
$\Delta\sigma^2_{arpk}$	0.04	0.03	0.08	0.00	0.29
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$	8.0%	5.7%	17.4%	0.3%	64.7%

With firm-specific labor taxes, denoted T^N_{it}, the firm's problem becomes

$$\begin{aligned}\mathcal{V}\left(K_{it},\mathcal{I}_{it}\right) = \max_{N_{it},K_{it+1}} \quad & \mathbb{E}_{it}\left[Y^{\frac{1}{\theta}}\hat{A}_{it}K^{\alpha_1}_{it}N^{\alpha_2}_{it} - WT^N_{it}N_{it} - T^K_{it+1}K_{it+1}\left(1-\beta\left(1-\delta\right)\right) - \Phi\left(K_{it+1},K_{it}\right)\right] \\ + \quad & \beta\mathbb{E}_{it}\left[\mathcal{V}\left(K_{it+1},\mathcal{I}_{it+1}\right)\right] .\end{aligned}$$

The labor choice satisfies the first order condition:

$$N_{it} = \left(\alpha_2\frac{Y^{\frac{1}{\theta}}\hat{A}_{it}K^{\alpha_1}_{it}}{WT^N_{it}}\right)^{\frac{1}{1-\alpha_2}} .$$

Substituting, we can derive operating profits (value-added net of total wages) as

$$\begin{aligned}P_{it}Y_{it} - WT^N_{it}N_{it} &= Y_t^{\frac{1}{\theta}}\hat{A}_{it}K^{\alpha_1}_{it}\left(\alpha_2 Y_t^{\frac{1}{\theta}}\frac{\hat{A}_{it}K^{\alpha_1}_{it}}{WT^N_{it}}\right)^{\frac{\alpha_2}{1-\alpha_2}} - WT^N_{it}\left(\alpha_2 Y_t^{\frac{1}{\theta}}\frac{\hat{A}_{it}K^{\alpha_1}_{it}}{WT^N_{it}}\right)^{\frac{1}{1-\alpha_2}} \\ &= \left(1-\alpha_2\right)\left(\frac{\alpha_2}{W}\right)^{\frac{\alpha_2}{1-\alpha_2}} Y^{\frac{1}{\theta}\frac{1}{1-\alpha_2}}\frac{\hat{A}_{it}^{\frac{1}{1-\alpha_2}}}{\left(T^N_{it}\right)^{\frac{\alpha_2}{1-\alpha_2}}}K_{it}^{\frac{\alpha_1}{1-\alpha_2}} \\ &= GA_{it}K^{\alpha}_{it} ,\end{aligned}$$

where

$$A_{it} \equiv \left(\frac{\hat{A}_{it}}{\left(T^N_{it}\right)^{\alpha_2}}\right)^{\frac{1}{1-\alpha_2}} . \tag{28}$$

Thus, the profit function (and therefore, the firm's investment problem) takes the same form as in the baseline version, except that A_{it} now incorporates the effect of the labor distortion as well. With this re-interpretation, our identification strategy remains valid, so long as A_{it} is correctly measured.

Measuring profitability. Recall that our empirical strategy in the baseline analysis measured profitability shocks using $a_{it} = va_{it} - \alpha k_{it}$. Expression (28) shows that this is the correct measure of profitability even with distortions to labor. This result implies that, apart from issues of interpretation, our quantitative analysis is not affected at all. In other words, our empirical strategy requires neither data on labor inputs nor taking a stand on the extent of labor distortions.

This is not the case for an alternative strategy that directly estimates the true productivity, $\hat{a}_{it} \equiv \log \hat{A}_{it} = va_{it} - \alpha_1 k_{it} - \alpha_2 n_{it}$, and uses it to construct the implied profitability term as $a_{it} = \frac{1}{1-\alpha_2}\hat{a}_{it}$, i.e., without controlling for labor distortions. It is easy to see that when there are firm-specific labor distortions, this approach leads to an incorrect measure of a_{it}, and therefore, to biased estimates of (ρ, σ_μ^2) and the other parameters. In particular, consider the empirically relevant case where the labor distortion is positively correlated with productivity: using a measure of A_{it} inferred from the estimated $\hat{A}_{it}$ without adjusting for T_{it}^N will overstate the variability in profitability. Quantitatively, this bias can be very large: in our data, this strategy produces estimates of (ρ, σ_μ) of $(0.90, 0.28)$ and $(0.88, 0.35)$ for the US and China, respectively, compared to our baseline estimates of $(0.93, 0.08)$ and $(0.91, 0.15)$. In other words, it overstates the volatility of shocks by a factor of almost 3.

This is essentially the strategy followed by Asker et al. (2014) and contributes to the difference between our estimates and theirs. To get a sense of the magnitude, a model with only convex adjustment costs estimated to match the variability of investment growth would yield an adjustment cost parameter, ξ, that is about 3 times higher in both countries under this more volatile shock process than under the baseline process (i.e., using a profitability measure that does account for labor distortions). This, in turn, would imply *arpk* dispersion from adjustment costs alone that exceeds the total observed dispersion in the US (and is about 60% of the total in China).

E.2 Frictional Labor

Here, we provide detailed derivations for the case of frictional labor in Section 6.2.

E.2.1 Model Solution

When labor is chosen under the same frictions as capital, the value function takes the form

$$\begin{aligned}\mathcal{V}\left(K_{it},N_{it},\mathcal{I}_{it}\right)=\max_{K_{it+1},N_{it+1}} & \quad \mathbb{E}_{it}\left[Y^{\frac{1}{\theta}}\hat{A}_{it}K_{it}^{\alpha_1}N_{it}^{\alpha_2}\right] \\ & - \mathbb{E}_{it}\left[T_{it+1}K_{it+1}\left(1-\beta\left(1-\delta\right)\right)+\Phi\left(K_{it+1},K_{it}\right)\right] \\ & - \mathbb{E}_{it}\left[T_{it+1}WN_{it+1}\left(1-\beta\left(1-\delta\right)\right)+W\Phi\left(N_{it+1},N_{it}\right)\right] \\ & + \mathbb{E}_{it}\left[\beta\mathcal{V}\left(K_{it+1},N_{it+1},\mathcal{I}_{it+1}\right)\right]\end{aligned} \tag{29}$$

where the adjustment cost function $\Phi\left(\cdot\right)$ is as defined in (2). Because the firm makes a one-time payment to hire incremental labor, the cost of labor W is now to be interpreted as the present discounted value of wages. Capital and labor are both subject to the same adjustment frictions, the same distortions, denoted T_{it+1}, and are chosen under the same information set, though the cost of labor adjustment is denominated in labor units.

The first order and envelope conditions yield two Euler equations:

$$\begin{aligned}\mathbb{E}_{it}\left[T_{it+1}\left(1-\beta\left(1-\delta\right)\right)+\Phi_1\left(K_{it+1},K_{it}\right)\right] &= \mathbb{E}_{it}\left[\beta\alpha_1 Y^{\frac{1}{\theta}}\hat{A}_{it+1}K_{it+1}^{\alpha_1-1}N_{it+1}^{\alpha_2}-\beta\Phi_2\left(K_{it+2},K_{it+1}\right)\right] \\ W\mathbb{E}_{it}\left[T_{it+1}\left(1-\beta\left(1-\delta\right)\right)+\Phi_1\left(N_{it+1},N_{it}\right)\right] &= \mathbb{E}_{it}\left[\beta\alpha_2 Y^{\frac{1}{\theta}}\hat{A}_{it+1}K_{it+1}^{\alpha_1}N_{it+1}^{\alpha_2-1}-\beta W\Phi_2\left(N_{it+2},N_{it+1}\right)\right]\end{aligned}$$

To show that this setup reduces to a Bellman equation of the same form as (3), we guess – and verify – that there exists a constant η such that the firm's labor policy takes the form $N_{it+1}=\eta K_{it+1}$.

Under this conjecture, we can rewrite the firm's problem in (29) as

$$\begin{aligned}\tilde{\mathcal{V}}\left(K_{it},\mathcal{I}_{it}\right)=\max_{K_{it+1}} & \quad \mathbb{E}_{it}\left[\frac{\eta^{\alpha_2}}{1+W\eta}Y^{\frac{1}{\theta}}\hat{A}_{it}K_{it}^{\alpha_1+\alpha_2}-T_{it+1}K_{it+1}\left(1-\beta\left(1-\delta\right)\right)\right] \\ & + \mathbb{E}_{it}\left[-\Phi\left(K_{it+1},K_{it}\right)+\beta\tilde{\mathcal{V}}\left(K_{it+1},\mathcal{I}_{it+1}\right)\right]\end{aligned}$$

Let $\{K_{it}^*\}$ be the solution to this problem. By definition, it must satisfy:

$$\begin{aligned}\mathbb{E}_{it}\left[T_{it+1}\left(1-\beta\left(1-\delta\right)\right)+\Phi_1\left(K_{it+1}^*,K_{it}^*\right)\right] &= \mathbb{E}_{it}\left[\beta\frac{\left(\alpha_1+\alpha_2\right)Y^{\frac{1}{\theta}}\hat{A}_{it+1}K_{it+1}^{*\alpha_1+\alpha_2-1}\eta^{\alpha_2}}{1+W\eta}\right] \\ &- \mathbb{E}_{it}\left[\beta\Phi_2\left(K_{it+2}^*,K_{it+1}^*\right)\right]\end{aligned} \tag{30}$$

Now substitute the conjecture $N_{it}^*=\eta K_{it}^*$ into the optimality condition for labor from the

original problem and rearrange to get:

$$\begin{aligned}\mathbb{E}_{it}\left[T_{it+1}\left(1-\beta\left(1-\delta\right)\right)+\Phi_1\left(K^*_{it+1},K^*_{it}\right)\right] &= \mathbb{E}_{it}\left[\beta\frac{\alpha_2 Y^{\frac{1}{\theta}}\hat{A}_{it+1}K^{*\alpha_1+\alpha_2-1}_{it+1}\eta^{\alpha_2}}{W\eta}\right] \\ &- \mathbb{E}_{it}\left[\beta\Phi_2\left(K^*_{it+2},K^*_{it+1}\right)\right]\end{aligned} \tag{31}$$

If η satisfies

$$\frac{\alpha_1+\alpha_2}{1+W\eta}=\frac{\alpha_2}{W\eta} \qquad \Rightarrow \qquad W\eta=\frac{\alpha_2}{\alpha_1} \tag{32}$$

then (31) is identical to (30). In other words, under (32), the sequence $\{K^*_{it},N^*_{it}\}$ satisfies the optimality condition for labor from the original problem. It is straightforward to verify that this also implies that $\{K^*_{it},N^*_{it}\}$ satisfy the optimality condition for capital from the original problem:

$$\begin{aligned}\mathbb{E}_{it}\left[T_{it+1}\left(1-\beta\left(1-\delta\right)\right)+\Phi_1\left(K^*_{it+1},K^*_{it}\right)\right] &= \mathbb{E}_{it}\left[\beta\alpha_1 Y^{\frac{1}{\theta}}\hat{A}_{it+1}K^{*\alpha_1+\alpha_2-1}_{it+1}\eta^{\alpha_2}-\beta\Phi_2\left(K^*_{it+2},K^*_{it+1}\right)\right] \\ &= \mathbb{E}_{it}\left[\beta\frac{\alpha_2 Y^{\frac{1}{\theta}}\hat{A}_{it+1}K^{*\alpha_1+\alpha_2-1}_{it+1}\eta^{\alpha_2}}{W\eta}-\beta\Phi_2\left(K^*_{it+2},K^*_{it+1}\right)\right]\end{aligned}$$

Thus, this version can be solved following the same steps as the baseline setup. The firm's problem takes the same form as expression (3), with $\alpha=\alpha_1+\alpha_2$, $G=\frac{\eta^{\alpha_2}Y^{\frac{1}{\theta}}}{1+W\eta}$ and $A_{it}=\hat{A}_{it}$.

E.2.2 Aggregation

To derive aggregate output and TFP for this case, we use $N_{it}=\eta K_{it}$ where $\eta=\frac{\alpha_2}{\alpha_1 W}$. Substituting into the revenue function gives

$$P_{it}Y_{it}=Y^{\frac{1}{\theta}}\hat{A}_{it}\eta^{\alpha_2}K^{\alpha_1+\alpha_2}_{it}=Y^{\frac{1}{\theta}}\hat{A}_{it}\eta^{\alpha_2}K^{\alpha}_{it}$$

By definition,

$$ARPK_{it}=Y^{\frac{1}{\theta}}\hat{A}_{it}\eta^{\alpha_2}K^{\alpha-1}_{it}$$

so that

$$K_{it}=\left(\frac{Y^{\frac{1}{\theta}}\hat{A}_{it}\eta^{\alpha_2}}{ARPK_{it}}\right)^{\frac{1}{1-\alpha}}$$

so that

$$\begin{aligned} P_{it}Y_{it} &= Y^{\frac{1}{\theta}}\eta^{\alpha_2}\hat{A}_{it}\left(\frac{Y^{\frac{1}{\theta}}\eta^{\alpha_2}\hat{A}_{it}}{ARPK_{it}}\right)^{\frac{\alpha}{1-\alpha}} \\ &= Y^{\frac{1}{\theta}\frac{1}{1-\alpha}}\eta^{\frac{\alpha_2}{1-\alpha}}\hat{A}_{it}^{\frac{1}{1-\alpha}}ARPK_{it}^{-\frac{\alpha}{1-\alpha}} \end{aligned}$$

and

$$Y = \int P_{it}Y_{it}di = Y^{\frac{1}{\theta}\frac{1}{1-\alpha}}\eta^{\frac{\alpha_2}{1-\alpha}}\int \hat{A}_{it}^{\frac{1}{1-\alpha}}ARPK_{it}^{-\frac{\alpha}{1-\alpha}}di$$

or, rearranging,

$$Y = Y^{\frac{1}{\theta}\frac{\hat{\alpha}_1+\hat{\alpha}_2}{1-\alpha}}\eta^{\frac{\hat{\alpha}_2}{1-\alpha}}\left(\int \hat{A}_{it}^{\frac{1}{1-\alpha}}ARPK_{it}^{-\frac{\alpha}{1-\alpha}}di\right)^{\frac{\theta}{\theta-1}}$$

Capital market clearing implies

$$K = \int K_{it}di = Y_t^{\frac{1}{\theta}\frac{1}{1-\alpha}}\eta^{\frac{\alpha_2}{1-\alpha}}\int \hat{A}_{it}^{\frac{1}{1-\alpha}}ARPK_{it}^{-\frac{1}{1-\alpha}}di$$

so that

$$K^{\hat{\alpha}_1}N^{\hat{\alpha}_2} = Y_t^{\frac{1}{\theta}\frac{\hat{\alpha}_1+\hat{\alpha}_2}{1-\alpha}}\eta^{\hat{\alpha}_2+\frac{\alpha_2}{1-\alpha}(\hat{\alpha}_1+\hat{\alpha}_2)}\left(\int \hat{A}_{it}^{\frac{1}{1-\alpha}}ARPK_{it}^{-\frac{1}{1-\alpha}}di\right)^{\hat{\alpha}_1+\hat{\alpha}_2}$$

Aggregate TFP is

$$A = \frac{Y}{K^{\hat{\alpha}_1}N^{\hat{\alpha}_2}} = \frac{\left(\int \hat{A}_{it}^{\frac{1}{1-\alpha}}ARPK_{it}^{-\frac{\alpha}{1-\alpha}}di\right)^{\frac{\theta}{\theta-1}}}{\left(\int \hat{A}_{it}^{\frac{1}{1-\alpha}}ARPK_{it}^{-\frac{1}{1-\alpha}}di\right)^{\hat{\alpha}_1+\hat{\alpha}_2}}$$

Following similar steps as in the baseline case, we can derive

$$a = a^* - \frac{1}{2}\frac{\theta}{\theta-1}\frac{\alpha}{1-\alpha}\sigma^2_{arpk}$$

Under constant returns to scale in production, this simplifies to

$$a = a^* - \frac{1}{2}\theta\sigma^2_{arpk}$$

The output effects of σ^2_{arpk} are the same as in the baseline case.

F Heterogeneity in Markups/Technologies

Baseline approach. The firm's cost minimization problem is

$$\min_{K_{it},N_{it},M_{it}} R_t T_{it}^K K_{it} + W_t T_{it}^N N_{it} + P_{it}^M M_{it} \quad s.t. \quad Y_{it} \leq K_{it}^{\hat{\alpha}_{it}} N_{it}^{\hat{\zeta}-\hat{\alpha}_{it}} M_{it}^{1-\hat{\zeta}}$$

The first order condition on M_{it} gives

$$P_{it}^M = \left(1-\hat{\zeta}\right)\frac{Y_{it}}{M_{it}} MC_{it} \quad \Rightarrow \quad \frac{P_{it}^M M_{it}}{P_{it} Y_{it}} = \left(1-\hat{\zeta}\right)\frac{MC_{it}}{P_{it}}$$

where MC_{it} is the Lagrange multiplier on the constraint (i.e., the marginal cost). Rearranging gives expression (14). In logs,

$$\log\frac{P_{it}}{MC_{it}} = \log\left(1-\hat{\zeta}\right) + \log\frac{P_{it}Y_{it}}{P_{it}^M M_{it}} \quad \Rightarrow \quad \sigma^2\left(\log\frac{P_{it}}{MC_{it}}\right) = \sigma^2\left(\log\frac{P_{it}Y_{it}}{P_{it}^M M_{it}}\right)$$

Similarly, the optimality conditions for K_{it} and N_{it} yield:

$$\begin{aligned}
\log\frac{P_{it}Y_{it}}{K_{it}} &= \log\frac{P_{it}}{MC_{it}} - \log\hat{\alpha}_{it} + \tau_{it}^K + \text{Constant} \\
\log\frac{P_{it}Y_{it}}{N_{it}} &= \log\frac{P_{it}}{MC_{it}} - \log\left(\hat{\zeta}-\hat{\alpha}_{it}\right) + \tau_{it}^N + \text{Constant}
\end{aligned}$$

Log-linearizing around the average $\hat{\alpha}_{it}$, denote it $\bar{\alpha}$, and ignoring constants yields $\log\left(\hat{\zeta}-\hat{\alpha}_{it}\right) \approx -\frac{\bar{\alpha}}{\hat{\zeta}-\bar{\alpha}}\log\hat{\alpha}_{it}$. Substituting gives expression (17).

Proof of Proposition 2. Assuming $\log\hat{\alpha}_{it}$ is uncorrelated with τ_{it}^K and τ_{it}^N,

$$cov\left(\widetilde{arpk}_{it}, \widetilde{arpn}_{it}\right) = -\frac{\bar{\alpha}}{\hat{\zeta}-\bar{\alpha}}\sigma^2_{\log\hat{\alpha}} + \text{cov}\left(\tau_{it}^k, \tau_{it}^n\right) \tag{33}$$

$$\sigma^2_{\widetilde{arpk}} = \sigma^2_{\log\hat{\alpha}} + \sigma^2_{\tau^k} \tag{34}$$

$$\sigma^2_{\widetilde{arpn}} = \left(\frac{\bar{\alpha}}{\hat{\zeta}-\bar{\alpha}}\right)^2\sigma^2_{\log\hat{\alpha}} + \sigma^2_{\tau^n} \tag{35}$$

From here, we can solve for the correlation of the distortions:

$$\rho\left(\tau_{it}^K, \tau_{it}^N\right) = \frac{\text{cov}\left(\widetilde{arpk}_{it}, \widetilde{arpn}_{it}\right) + \frac{\bar{\alpha}}{\hat{\zeta}-\bar{\alpha}}\sigma^2_{\log\hat{\alpha}}}{\sqrt{\sigma^2_{\widetilde{arpk}} - \sigma^2_{\log\hat{\alpha}}}\sqrt{\sigma^2_{\widetilde{arpn}} - \left(\frac{\bar{\alpha}}{\hat{\zeta}-\bar{\alpha}}\right)^2\sigma^2_{\log\hat{\alpha}}}}$$

which is increasing in $\sigma^2_{\log\hat{\alpha}}$. An upper bound for $\sigma^2_{\log\hat{\alpha}}$, denoted $\bar{\sigma}^2_{\log\hat{\alpha}}$, is where $\rho\left(\tau_{it}^K, \tau_{it}^N\right) = 1$,

and substituting and rearranging gives

$$\bar{\sigma}^2_{\hat{\alpha}} = \frac{\sigma^2_{\widetilde{arpk}}\sigma^2_{\widetilde{arpn}} - \text{cov}\left(\widetilde{arpk}_{it}, \widetilde{arpn}_{it}\right)^2}{2\frac{\bar{\alpha}}{\hat{\zeta}-\bar{\alpha}}\text{cov}\left(\widetilde{arpk}_{it}, \widetilde{arpn}_{it}\right) + \left(\frac{\bar{\alpha}}{\hat{\zeta}-\bar{\alpha}}\right)^2 \sigma^2_{\widetilde{arpk}} + \sigma^2_{\widetilde{arpn}}}$$

□

Heterogeneous materials elasticities. We now allow for heterogeneity in $\hat{\zeta}_{it}$, so the cost minimization problem becomes

$$\min_{K_{it},N_{it},M_{it}} \; R_t T^K_{it} K_{it} + W_t T^N_{it} N_{it} + P^M_{it} M_{it} \quad s.t. \quad Y_{it} \leq K^{\hat{\alpha}_{it}}_{it} N^{\hat{\zeta}_{it}-\hat{\alpha}_{it}}_{it} M^{1-\hat{\zeta}_{it}}_{it}$$

The first order conditions give the optimal average products of inputs (after some rearranging):

$$\begin{aligned}
ARPK_{it} &\equiv \frac{P_{it}Y_{it}}{K_{it}} = \frac{P_{it}}{MC_{it}}\frac{1}{\hat{\alpha}_{it}}T^K_{it}R_t \\
ARPN_{it} &\equiv \frac{P_{it}Y_{it}}{N_{it}} = \frac{P_{it}}{MC_{it}}\frac{1}{\hat{\zeta}_{it}-\hat{\alpha}_{it}}T^N_{it}W_t \\
ARPM_{it} &\equiv \frac{P_{it}Y_{it}}{P^M_{it}M_{it}} = \frac{P_{it}}{MC_{it}}\frac{1}{1-\hat{\zeta}_{it}}
\end{aligned}$$

or in logs:

$$\begin{aligned}
arpk_{it} &= \varphi_{it} - \log\hat{\alpha}_{it} + \tau^K_{it} + \text{Constant} \\
arpn_{it} &= \varphi_{it} - \log\left(\hat{\zeta}_{it} - \hat{\alpha}_{it}\right) + \tau^N_{it} + \text{Constant} \\
&\approx \varphi_{it} - \frac{\bar{\zeta}}{\bar{\zeta}-\bar{\alpha}}\log\hat{\zeta}_{it} + \frac{\bar{\alpha}}{\bar{\zeta}-\bar{\alpha}}\log\hat{\alpha}_{it} + \tau^N_{it} + \text{Constant} \\
arpm_{it} &= \varphi_{it} - \log\left(1-\hat{\zeta}_{it}\right) \\
&\approx \varphi_{it} + \frac{\bar{\zeta}}{1-\bar{\zeta}}\log\hat{\zeta}_{it} + \text{Constant}
\end{aligned}$$

where, to ease notation, we define $\varphi_{it} \equiv \log\frac{P_{it}}{MC_{it}}$ as the log markup and the approximations reflect log-linearizations around the average elasticities, denoted $\bar{\alpha}$ and $\bar{\zeta}$.

There are three categories of firm-specific variation in the average product of inputs: markups, φ_{it}, input elasticities in production, $\hat{\alpha}_{it}$ and $\hat{\zeta}_{it}$, and distortions, τ^K_{it} and τ^N_{it}. We make the following key assumption: this variation is independent across categories. Within categories, however, we allow for arbitrary correlations. In other words, the covariances $\sigma_{\log\hat{\alpha},\log\hat{\zeta}}$ and σ_{τ^K,τ^N} are

unrestricted but the other covariances are set to zero.[52] We can derive the following expressions for the second moments of the average products of the three inputs:

$$\sigma^2_{arpk} = \sigma^2_{\varphi} + \sigma^2_{\log\hat{\alpha}} + \sigma^2_{\tau^K} \tag{36}$$

$$\sigma^2_{arpn} = \sigma^2_{\varphi} + \left(\frac{\bar{\zeta}}{\bar{\zeta}-\bar{\alpha}}\right)^2 \sigma^2_{\log\hat{\zeta}} + \left(\frac{\bar{\alpha}}{\bar{\zeta}-\bar{\alpha}}\right)^2 \sigma^2_{\log\hat{\alpha}} - \frac{2\bar{\alpha}\bar{\zeta}}{\left(\bar{\zeta}-\bar{\alpha}\right)^2}\sigma_{\log\hat{\alpha},\log\hat{\zeta}} + \sigma^2_{\tau^N} \tag{37}$$

$$\sigma_{arpk,arpn} = \sigma^2_{\varphi} + \frac{\bar{\zeta}}{\bar{\zeta}-\bar{\alpha}}\sigma_{\log\hat{\alpha},\log\hat{\zeta}} - \frac{\bar{\alpha}}{\bar{\zeta}-\bar{\alpha}}\sigma^2_{\log\hat{\alpha}} + \sigma_{\tau^K,\tau^N} \tag{38}$$

$$\sigma^2_{arpm} = \sigma^2_{\varphi} + \left(\frac{\bar{\zeta}}{1-\bar{\zeta}}\right)^2 \sigma^2_{\log\hat{\zeta}} \tag{39}$$

$$\sigma_{arpk,arpm} = \sigma^2_{\varphi} - \frac{\bar{\zeta}}{1-\bar{\zeta}}\sigma_{\log\hat{\alpha},\log\hat{\zeta}} \tag{40}$$

$$\sigma_{arpn,arpm} = \sigma^2_{\varphi} - \frac{\bar{\zeta}^2}{\left(\bar{\zeta}-\bar{\alpha}\right)\left(1-\bar{\zeta}\right)}\sigma^2_{\log\hat{\zeta}} + \frac{\bar{\alpha}\bar{\zeta}}{\left(\bar{\zeta}-\bar{\alpha}\right)\left(1-\bar{\zeta}\right)}\sigma_{\log\hat{\alpha},\log\hat{\zeta}} \tag{41}$$

The following result states that we can identify the dispersion in (log) markups and materials elasticities from the second moments of *arpm*.

Lemma 1. *The parameters $\sigma^2_{\varphi}, \sigma^2_{\log\hat{\zeta}}$ and $\sigma_{\log\hat{\alpha},\log\hat{\zeta}}$ are uniquely identified by $\sigma^2_{arpm}, \sigma_{arpn,arpm}$ and $\sigma_{arpk,arpm}$.*

Proof. Rearrange (41) to derive

$$\frac{\bar{\zeta}}{1-\bar{\zeta}}\sigma_{\log\hat{\alpha},\log\hat{\zeta}} = \frac{\bar{\zeta}-\bar{\alpha}}{\bar{\alpha}}\left(\sigma_{arpn,arpm} - \sigma^2_{\varphi}\right) + \frac{1-\bar{\zeta}}{\bar{\alpha}}\left(\frac{\bar{\zeta}}{1-\bar{\zeta}}\right)^2 \sigma^2_{\log\hat{\zeta}}$$

Substituting into (40) gives:

$$\sigma^2_{\varphi} = \frac{\bar{\alpha}}{\bar{\zeta}}\sigma_{arpk,arpm} + \frac{\bar{\zeta}-\bar{\alpha}}{\bar{\zeta}}\sigma_{arpn,arpm} + \frac{1-\bar{\zeta}}{\bar{\zeta}}\left(\frac{\bar{\zeta}}{1-\bar{\zeta}}\right)^2 \sigma^2_{\log\hat{\zeta}}$$

and then into (39) yields:

$$\left(\frac{\bar{\zeta}}{1-\bar{\zeta}}\right)^2 \sigma^2_{\log\hat{\zeta}} = \bar{\zeta}\sigma^2_{arpm} - \bar{\alpha}\sigma_{arpk,arpm} - \left(\bar{\zeta}-\bar{\alpha}\right)\sigma_{arpn,arpm}$$

This equation pins down $\sigma^2_{\log\hat{\zeta}}$. Given this, the other two equations yield σ^2_{φ} and $\sigma_{\log\hat{\alpha},\log\hat{\zeta}}$. □

Intuitively, the greater the positive covariation in the average revenue products of the three inputs, the lower (higher) the variation in the output elasticity of materials (markups). Using

[52] The fact that materials shares were relatively uncorrelated with size provides some justification for this assumption.

these estimates, we can appropriately adjust the second moments of *arpk* and *arpn* and apply the logic of Proposition 2 to derive an upper bound for the variation in $\hat{\alpha}_{it}$. We state this result formally in the following proposition:

Proposition 3. *The dispersion in* $\log \hat{\alpha}_{it}$ *satisfies*

$$\sigma^2(\log \hat{\alpha}_{it}) \leq \frac{\tilde{\sigma}^2_{arpk}\tilde{\sigma}^2_{arpn} - \tilde{cov}\left(arpk, arpn\right)^2}{2\frac{\bar{\alpha}}{\bar{\zeta}-\bar{\alpha}}\tilde{cov}\left(arpk, arpn\right) + \left(\frac{\bar{\alpha}}{\bar{\zeta}-\bar{\alpha}}\right)^2 \tilde{\sigma}^2_{arpk} + \tilde{\sigma}^2_{arpn}} \,. \tag{42}$$

where

$$\begin{aligned}
\tilde{\sigma}^2_{arpk} &\equiv \sigma^2_{arpk} - \sigma^2_{\varphi} \\
\tilde{\sigma}^2_{arpn} &\equiv \sigma^2_{arpn} - \sigma^2_{\varphi} - \left(\frac{\bar{\zeta}}{\bar{\zeta}-\bar{\alpha}}\right)^2 \sigma^2_{\log\hat{\zeta}} + \frac{2\bar{\alpha}\bar{\zeta}}{\left(\bar{\zeta}-\bar{\alpha}\right)^2}\sigma_{\log\hat{\alpha},\log\hat{\zeta}} \\
\tilde{cov}\left(arpk, arpn\right) &\equiv \sigma_{arpk,arpn} - \sigma^2_{\varphi} - \frac{\bar{\zeta}}{\bar{\zeta}-\bar{\alpha}}\sigma_{\log\hat{\alpha},\log\hat{\zeta}} \,.
\end{aligned}$$

Proof. Note that

$$\begin{aligned}
\tilde{\sigma}^2_{arpk} &= \sigma^2_{\log\hat{\alpha}} + \sigma^2_{\tau^K} \\
\tilde{\sigma}^2_{arpn} &= \left(\frac{\bar{\alpha}}{\bar{\zeta}-\bar{\alpha}}\right)^2 \sigma^2_{\log\hat{\alpha}} + \sigma^2_{\tau^N} \\
\tilde{cov}\left(arpk, arpn\right) &= -\frac{\bar{\alpha}}{\bar{\zeta}-\bar{\alpha}}\sigma^2_{\log\hat{\alpha}} + \mathrm{cov}\left(\tau^K_{it}, \tau^N_{it}\right) \,.
\end{aligned}$$

which are identical to expressions (33), (34) and (35). The proof is then the same as for Proposition 2. □

In Table 11, we report the results from applying this methodology to our datasets. Comparing them to the results in Table 4 reveals that allowing for unobserved variation in the output elasticity of materials slightly attenuates the contribution of markup dispersion and raises the upper bound for the effects of technology heterogeneity, but the values are extremely close. One reason why the two approaches are so similar is that the variation in the materials elasticity (across firms within the same industry), $\sigma^2_{\log\hat{\zeta}}$, is estimated to be very small in both countries: 0.007 in the US and 0.018 in China.

Table 11: Heterogeneous Markups and Technologies with Firm-Specific Materials Elasticities

	China			US		
Covariance matrix	$arpk_{it}$	$arpn_{it}$	$arpm_{it}$	$arpk_{it}$	$arpn_{it}$	$arpm_{it}$
$arpk_{it}$	1.30			0.41		
$arpn_{it}$	0.33	0.69		0.10	0.20	
$arpm_{it}$	0.01	0.01	0.05	0.03	0.05	0.06
Estimated $\Delta\sigma^2_{arpk}$	Level	Share		Level	Share	
Dispersion in Markups	0.03	(2.4%)		0.05	(11.8%)	
Dispersion in $\log\hat{\alpha}_{it}$	0.31	(24.1%)		0.21	(51.6%)	
Total	0.34	(26.5%)		0.26	(63.4%)	

G Size-Dependent Policies

In this appendix, we explore the relationship between size- and productivity-dependent factors. First, note that our empirical strategy can be thought of as essentially recovering the law of motion for k_{it} – in particular, the coefficients ψ_1, $\psi_2(1+\gamma)$, ψ_3 and ψ_4. Importantly, these estimates are invariant to assumptions about γ_k, which only affects the mapping from these coefficients to the underlying structural parameters. For example, suppose we assume $\gamma_k = 0$. Then, given our values for (α, β, δ), the estimated ψ_1 identifies the adjustment cost parameter ξ. Next, the value of ξ can be used to pin down ψ_2, allowing us to recover γ from the estimated $\psi_2(1+\gamma)$. This procedure can be applied for any given γ_k as well. Since the estimated ψ_1 and $\psi_2(1+\gamma)$ do not change, for any γ_k, the adjustment cost parameter becomes, from (8),

$$\xi = \psi_1 \frac{1-\alpha-\gamma_k}{\beta\psi_1^2 + 1 - \psi_1(1+\beta)}.$$

The next step is the same as before: the estimated ξ implies a value for ψ_2, which then allows us to back out γ from the estimated $\psi_2(1+\gamma)$. Table 5 applies this procedure for various values of γ_k to trace out a set of parameters that are observationally equivalent, i.e., that cannot be distinguished using only data on capital and value-added.

H Financial Frictions

Including the liquidity cost, the firm's problem can be written as

$$\begin{aligned}\mathcal{V}(K_{it}, B_{it}, \mathcal{I}_{it}) = \max_{B_{it+1}, K_{it+1}} \quad & \mathbb{E}_{it}\left[\Pi(K_{it}, A_{it}) + RB_{it} - B_{it+1} - T^K_{it+1}K_{it+1}(1-\beta(1-\delta))\right] \\ - \quad & \Phi(K_{it+1}, K_{it}) - \Upsilon(K_{it+1}, B_{it+1}) + \beta\mathbb{E}_{it}\left[\mathcal{V}(K_{it+1}, B_{it+1}, \mathcal{I}_{it+1})\right]\end{aligned}$$

The first order conditions are given by

$$\begin{aligned}\mathbb{E}_{it}\left[\beta\Pi_1\left(K_{it+1},A_{it+1}\right)-\beta\Phi_2\left(K_{it+2},K_{it+1}\right)\right] &= T^K_{it+1}\left(1-\beta\left(1-\delta\right)\right)+\Phi_1\left(K_{it+1},K_{it}\right)+\Upsilon_1\left(K_{it+1},B_{it+1}\right)\\ -\Upsilon_2\left(K_{it+1},B_{it+1}\right)+\beta R &= 1\end{aligned}$$

Note that

$$\Upsilon_2\left(K_{it+1},B_{it+1}\right)=-\hat{\nu}\omega_2\frac{K^{\omega_1}_{it+1}}{B^{\omega_2+1}_{it+1}},\qquad \Upsilon_1\left(K_{it+1},B_{it+1}\right)=\hat{\nu}\omega_1\frac{K^{\omega_1-1}_{it+1}}{B^{\omega_2}_{it+1}}$$

Using the FOC for B_{it+1}

$$\begin{aligned}1 &= \hat{\nu}\omega_2\frac{K^{\omega_1}_{it+1}}{B^{\omega_2+1}_{it+1}}+\beta R \quad\Rightarrow\quad B_{it+1}=\left(\frac{\hat{v}\omega_2}{1-\beta R}\right)^{\frac{1}{\omega_2+1}}K^{\frac{\omega_1}{\omega_2+1}}_{it+1}\\ \Upsilon_1\left(K_{it+1},B_{it+1}\right) &= \hat{\nu}\omega_1\frac{K^{\omega_1-1}_{it+1}}{B^{\omega_2}_{it+1}}=\hat{\nu}\omega_1\frac{K^{\omega_1-1}_{it+1}}{\left(\frac{\hat{v}\omega_2}{1-\beta R}\right)^{\frac{\omega_2}{\omega_2+1}}K^{\frac{\omega_2\omega_1}{\omega_2+1}}_{it+1}}=\left(\frac{\hat{\nu}}{\omega_2^{\omega_2}}\right)^{\frac{1}{\omega_2+1}}\omega_1\left(1-\beta R\right)^{\frac{\omega_2}{\omega_2+1}}K^{\frac{\omega_1-(\omega_2+1)}{\omega_2+1}}_{it+1}\\ &= \nu\left(1-\beta R\right)^{\frac{\omega_2}{\omega_2+1}}K^{\omega}_{it+1}\,,\end{aligned}$$

where

$$\begin{aligned}\nu &\equiv \left(\frac{\hat{\nu}}{\omega_2^{\omega_2}}\right)^{\frac{1}{\omega_2+1}}\omega_1\\ \omega &\equiv \frac{\omega_1-(\omega_2+1)}{\omega_2+1}.\end{aligned}$$

Log-linearizing,

$$\begin{aligned}\bar{\Upsilon}_1+\bar{\Upsilon}_1\upsilon_{1t+1} &\approx \nu\left(1-\beta R\right)^{\frac{\omega_2}{\omega_2+1}}\bar{K}^{\omega}+\nu\left(1-\beta R\right)^{\frac{\omega_2}{\omega_2+1}}\bar{K}^{\omega}\omega k_{it+1}\\ \bar{\Upsilon}_1\upsilon_{1t+1} &\approx \nu\left(1-\beta R\right)^{\frac{\omega_2}{\omega_2+1}}\bar{K}^{\omega}\omega k_{it+1}\,.\end{aligned}$$

Substituting into the FOC,

$$\begin{aligned}&\mathbb{E}_{it}\left[\alpha\beta\bar{G}\bar{A}\bar{K}^{\alpha-1}\left(a_{it+1}+\left(\alpha-1\right)k_{it+1}\right)+\beta\hat{\xi}\left(k_{it+2}-k_{it+1}\right)-\tau^K_{it+1}\left(1-\beta\left(1-\delta\right)\right)\right]\\ =\;&\hat{\xi}\left(k_{it+1}-k_{it}\right)+\nu\left(1-\beta R\right)^{\frac{\omega_2}{\omega_2+1}}\bar{K}^{\omega}\omega k_{it+1}\,,\end{aligned}$$

or

$$k_{it+1}\left(\left(1+\beta\right)\xi+1-\alpha-\gamma_k\right)=\mathbb{E}_{it}\left[a_{it+1}+\tau_{it+1}\right]+\beta\xi\mathbb{E}_{it}\left[k_{it+2}\right]+\xi k_{it}\,,$$

where

$$\begin{aligned}\gamma_k &= -\frac{\nu\left(1-\beta R\right)^{\frac{\omega_2}{\omega_2+1}}\omega\bar{K}^{\omega}}{\alpha\beta\bar{G}\bar{A}\bar{K}^{\alpha-1}} = -\frac{\nu\left(1-\beta R\right)^{\frac{\omega_2}{\omega_2+1}}\omega\bar{K}^{\omega}}{\nu\left(1-\beta R\right)^{\frac{\omega_2}{\omega_2+1}}\bar{K}^{\omega}+1-\beta\left(1-\delta\right)+\hat{\xi}\delta\left(1-\beta\left(1-\frac{\delta}{2}\right)\right)} \\ &= -\frac{\omega\bar{\Upsilon}_1}{\bar{\Upsilon}_1+\kappa}\end{aligned}$$

where we have substituted in from the steady state Euler equations and $\kappa \equiv 1-\beta\left(1-\delta\right)+\hat{\xi}\delta\left(1-\beta\left(1-\frac{\delta}{2}\right)\right)$.

I Robustness

I.1 Computation and Estimation of Non-Convex Model

This appendix provides details of our analysis of non-convex adjustment costs from Section 6.1. Since we can no longer rely on the perturbation approach, we solve the model non-linearly using value function iteration and estimate the parameters via simulated method of moments.

Estimation details. Our estimation uses the following procedure. For a given parameter vector $\left(\hat{\xi},\hat{\xi}_f,\mathbb{V},\gamma.\sigma_\varepsilon^2,\sigma_\chi^2\right)$, we solve for the value and policy functions using a standard iterative procedure and discretized grids for the state variables.[53] We then use these solutions to simulate time paths (10,000 periods) for firm-level capital and productivity. We discard the first 5,000 periods and compute the moments of interest using the remaining observations. We then search over the parameter vector to minimize the equally-weighted sum of squared deviations between the simulated values of the six target moments and their empirical counterparts.

Model fit. We report the fit of the estimated model in row 'All factors baseline' of Table 12. The left panel displays the parameter estimates and the right panel the simulated moments (the top row labeled 'Data' reports the empirical values of the moments). The set of targeted moments is marked in bold italics. The table shows that the model matches the targeted moments quite well (the fit is almost exact in the US; the model slightly undershoots the variance of investment growth and the correlation of *arpk* with *a* in China, but is still fairly close on those dimensions).

The last four columns of the table contain four additional moments not explicitly targeted in the estimation – the autocorrelation of investment (in levels), denoted $\rho_{k,k_{-1}}$, the correlation

[53] We use a relatively fine grid for capital (at intervals of 0.025 log points). For the productivity process, we use 21 grid points, spanning 6 standard deviations (±3 standard deviations on either side of the mean). We have verified that our results are not particularly sensitive to these choices.

of investment with productivity, $\rho_{k,a}$, and investment 'spikes,' defined as the fraction of observations with (gross) investment rates above 20%, $spike^+$, or less than -20%, $spike^-$. These are the moments targeted in Cooper and Haltiwanger (2006). The set of moments examined in the table were broadly chosen to encompass those from our estimation and additional moments considered in previous influential studies of adjustment costs, namely, Cooper and Haltiwanger (2006) and Asker et al. (2014) (the latter paper targets the variability of investment, inaction and spikes, where the latter two moments are defined in the same way as here).

Turning to the non-targeted moments, the model somewhat over-predicts the serial correlation of investment as well as its correlation with productivity. The model also overshoots a bit on the fraction of positive investment spikes. To explore the extent to which these deviations matter for our main conclusions, we estimated two alternative versions of the model. The first replaces inaction as a target with $spike^+$ and $spike^-$.[54] The results, reported in row 'All factors spikes', yield estimates of the adjustment costs that are only slightly higher than the baseline values in both countries. In the second exercise, we targeted the serial correlation of investment in levels (rather than growth rates). As with the first exercise, the parameter estimates (reported in row 'All factors $\rho_{k,k_{-1}}$') change only slightly. Importantly, across both exercises, the contribution of the various factors to *arpk* dispersion (not reported in the table) are almost unchanged. In sum, these exercises reveal that (i) while our relatively simple specification of adjustment costs and other distortionary factors can reconcile an extremely broad set of investment moments, it struggles to exactly match all moments simultaneously, but (ii) despite this, our conclusions about the sources of dispersion in *arpk* are quite robust to the precise choice of moments.[55]

The role of other factors. Explicitly allowing for other distortionary factors plays a key role in our analysis. It significantly contributes to our ability to simultaneously match various data moments and is the primary reason for the difference between our estimates of adjustment costs and those in previous studies.

To show this more clearly, we estimated three alternative versions of our model with only adjustment costs.[56] The first is estimated by targeting the same moments as do Cooper and Haltiwanger (2006), namely, the serial correlation of investment, its correlation with productiv-

[54]This is in line with Cooper and Haltiwanger (2006), who argue that inaction may be poorly measured in the micro-data and use these moments instead.

[55]We also estimated a version where all 10 moments are targeted together. Unsurprisingly, the model cannot exactly match all of them, but the best-fit parameter estimates are quite similar to the baseline.

[56]Formally, all the parameters except the two governing adjustment costs ($\hat{\xi}$ and $\hat{\xi}_f$) are set to 0. We then search over $\hat{\xi}$ and $\hat{\xi}_f$ to minimize the equally-weighted distance between the model-implied and data values for the target moments.

Table 12: Non-Convex Adjustment Costs - Alternative Approaches

	$\hat{\xi}$	$\hat{\xi}_f$	$\mathbb{V}$	γ	σ_ε^2	σ_χ^2	$\rho_{\iota,\iota_{-1}}$	σ_ι^2	$\rho_{arpk,a}$	$\rho_{\iota,a_{-1}}$	σ_{arpk}^2	$inact$	$\rho_{k,k_{-1}}$	$\rho_{k,a}$	$spike^+$	$spike^-$
US																
Data							−0.30	0.06	0.55	0.13	0.45	0.18	0.25	0.06	0.26	0.10
All factors baseline	0.25	0.002	0.03	−0.30	0.02	0.29	***-0.30***	***0.05***	***0.55***	***0.13***	***0.45***	***0.18***	0.46	0.17	0.31	0.08
All factors spikes	0.31	0.007	0.03	−0.27	0.02	0.29	***-0.29***	***0.04***	***0.54***	***0.13***	***0.45***	0.30	0.50	0.20	***0.30***	***0.07***
All factors $\rho_{k,k_{-1}}$	0.34	0.003	0.04	−0.32	0.16	0.26	−0.42	***0.09***	***0.59***	***0.13***	***0.45***	***0.18***	***0.26***	0.17	0.33	0.10
AC only I	0.11	0.130	-	-	-	-	−0.42	0.26	0.57	−0.26	0.06	0.70	***0.24***	***0.03***	***0.23***	***0.07***
AC only II	0.10	0.120	-	-	-	-	−0.40	0.24	0.57	−0.27	0.05	0.34	***0.27***	***0.03***	***0.21***	***0.06***
AC only III	0.65	0.003	-	-	-	-	−0.18	***0.03***	0.75	−0.22	0.11	***0.18***	0.66	0.20	***0.30***	***0.07***
China																
Data							−0.36	0.14	0.76	0.29	0.92	0.20	0.04	−0.06	0.27	0.15
All factors baseline	0.07	0.002	0.09	−0.64	0.00	0.44	***-0.38***	***0.11***	***0.71***	***0.28***	***0.92***	***0.21***	0.24	0.17	0.35	0.13
All factors spikes	0.09	0.001	0.09	−0.66	0.00	0.41	***-0.38***	***0.09***	***0.73***	***0.28***	***0.92***	0.18	0.25	0.17	***0.33***	***0.11***
All factors $\rho_{k,k_{-1}}$	0.05	0.002	0.10	−0.70	0.01	0.38	−0.47	***0.16***	***0.76***	***0.26***	***0.92***	***0.23***	***0.05***	0.14	0.36	0.15
AC only I	0.01	0.200	-	-	-	-	−0.47	2.23	0.51	−0.39	0.05	0.60	***0.05***	***-0.09***	***0.28***	***0.12***
AC only II	0.01	0.190	-	-	-	-	−0.47	2.28	0.51	−0.40	0.04	0.24	***0.05***	***-0.10***	***0.27***	***0.11***
AC only III	0.80	0.012	-	-	-	-	−0.17	***0.05***	0.88	−0.20	0.28	***0.20***	0.67	0.26	***0.34***	***0.13***

Notes: Rows labeled 'All factors' contain estimations of the full model. Rows labeled 'AC only' contain estimations of adjustment cost-only models. Targeted moments are in bold italics. 'All factors baseline' replicates our baseline analysis from Section 6.1. 'All factors spikes' targets investment spikes. 'All factors $\rho_{k,k_{-1}}$' targets the serial correlation of investment in levels. 'AC only I' targets the moments from Cooper and Haltiwanger (2006). 'AC only II' targets the same moments, but only subjects investment rates greater than than 5% to the fixed adjustment cost. 'AC only III' targets moments along the lines of Asker et al. (2014).

ity and the fractions of positive and negative spikes.[57] The results, presented in row 'AC Only I' in Table 12, show much larger fixed adjustment costs (and lower convex costs) compared to all the variants of our estimation in the first three rows. Intuitively, as the only offsetting force, large non-convex costs are necessary to match the relatively modest serial correlation of investment observed in the data. However, this comes at the expense of counterfactually high values for inaction and investment variability. For example, in the US, this version of the model predicts an inaction rate of 70% and a variance of investment growth of 0.26, compared to their empirical values of 18% and 0.06, respectively. In other words, an adjustment cost-only model estimated to match only the serial correlation of investment and investment spikes struggles to match the modest degrees of both inaction and investment variability observed in the data.

It is possible to partly fix some of these counterfactual implications using a more complicated specification of the adjustment cost function. For example, assuming that only large investments are subject to the fixed cost helps reduce the degree of inaction.[58] The row labeled 'AC Only II' shows results from such a modification, where the fixed cost is incurred only for investment rates greater than 5% in absolute value. As expected, this brings the predicted value for inaction much closer to the data, particularly in China (predicted inaction remains excessively high in the US, 34% compared to 18% in the data), but has little effect on the variability of investment growth, which remains counterfactually high. More importantly for our purposes, neither version of the adjustment cost-only model generates significant dispersion in *arpk* (indeed the second specification that better fits the data actually *reduces* the implied dispersion from adjustment costs) – in the US, the predicted σ^2_{arpk} is only about 13% of the observed level in the data. This fraction is even lower in China.

The final exercise, displayed in row 'AC Only III', targets the variability of investment growth along with inaction and spikes together. This is similar to the strategy in Asker et al. (2014) (with the caveat that they target the variance of investment in levels). This produces substantially higher estimates for the convex cost in both countries:[59] Intuitively, a large convex component is necessary to match the extremely low variability of investment. However, these estimates imply a counterfactually high serial correlation. In China, for example, the predicted $\rho_{k,k_{-1}}$ is 0.67, compared to the empirical value of only 0.04 (the corresponding values in the US are 0.66 compared to 0.25). These findings are precisely in line with the logic presented in

[57] See section 4.1.2 of that paper. The moments and resulting parameter estimates are not directly comparable since the set of firms is quite different – Cooper and Haltiwanger (2006) work with data on US manufacturing firms from the Longitudinal Research Database.

[58] This is along the lines of the specification in Khan and Thomas (2008), who also point out the inability of standard adjustment cost models to simultaneously match both inaction and spikes in firm/establishment-level data.

[59] The results for the fixed component are more mixed – the estimate is very close to our baseline value in the US and is significantly higher in China, though well below the previous two adjustment cost-only estimations.

Section 3 and are analogous to our discussion of the adjustment cost results in Section 4.4 – the fact that the data show only modest serial correlations of investment/investment growth rates limits the potential for convex adjustment frictions; an estimation strategy ignoring this moment (and targeting the volatility of investment) can lead to substantial upward bias in the estimates of convex costs.

In contrast to these adjustment cost-only specifications, our baseline model is able to capture a broader set of data patterns precisely because of the inclusion of other factors influencing investment. To gain some intuition for how they help, we turn to the formulae for the variance and serial correlation of investment derived in Section 3 for the random walk case:

$$\begin{aligned} \sigma_k^2 &= \left(\frac{\psi_2^2}{1-\psi_1^2}\right)(1+\gamma)^2\,\sigma_\mu^2 + \frac{2}{1+\psi_1}\psi_3^2\sigma_\varepsilon^2 \\ \rho_{k,k_{-1}} &= \psi_1 - \frac{\psi_3^2\sigma_\varepsilon^2}{\sigma_k^2}\ , \end{aligned}$$

where ψ_1, ψ_2, ψ_3 are composite parameters independent of distortions. These expressions show that, *ceteris paribus*, more severe correlated distortions (i.e., more negative γ) reduce both the volatility and serial correlation of investment (the latter through the effects on σ_k^2). Intuitively, correlated distortions lessen the influence of the persistent productivity process on investment, reducing the serial correlation. Uncorrelated factors (higher σ_ε^2) also make investment less serially correlated, but more volatile. Quantitatively, the first effect is much larger.[60] As a result, the model can match both of these moments without resorting to large non-convex costs (and the associated counterfactual implications).

In sum, the exercises in this appendix emphasize one of the main messages of our analysis: examining a broad set of investment moments imposes additional discipline on the magnitude of the various forces (including adjustment costs). In both countries, the data show that investment/investment growth is (i) neither particularly volatile (ii) nor highly autocorrelated, but (iii) there are large and extremely extremely persistent deviations of firm-level capital from its 'efficient' level. These patterns seem hard to rationalize with standard specifications of adjustment costs alone and lead us to find a significant role for other factors, particularly when it comes to explaining the dispersion in σ_{arpk}^2.

I.2 Alternative Stochastic Processes

In this section, we analyze the implications of alternative, richer stochastic processes for firm-level productivity and distortions.

[60] For example, at our baseline estimates in the US, the coefficient on $\psi_3^2\sigma_\varepsilon^2$ in the expression for σ_k^2 is $\frac{2}{1+\psi_1} \approx 1.2$ while the coefficient in the expression for $\rho_{k,k_{-1}}$ is $\frac{1}{\sigma_k^2} \approx 25$.

Fixed-effects in productivity. First, we generalize the process on productivity in equation (5) to include firm-level fixed-effects. Specifically, we assume:

$$\begin{aligned} a_{it} &= \bar{a}_i + \hat{a}_{it}, \qquad \bar{a}_{it} \sim \mathcal{N}\left(0, \sigma^2_{\bar{a}}\right) \\ \hat{a}_{it} &= \rho \hat{a}_{it-1} + \mu_{it}, \quad \mu_{it} \sim \mathcal{N}\left(0, \sigma^2_{\mu}\right) \end{aligned} \tag{43}$$

Now, productivity is composed of both an AR(1) component, $\hat{a}_{it}$, as in equation (5) and a firm fixed-effect, $\bar{a}_i$, with cross-sectional variance $\sigma^2_{\bar{a}}$.

We can show that the three parameters $(\rho, \sigma^2_{\mu}, \sigma^2_{\bar{a}})$ are uniquely identified by (i) the serial correlation of productivity growth along with (ii) the coefficient from a simple autoregression of a_{it} on a_{it-1} and (iii) the residuals from that regression, which we denote $\sigma^2_{\hat{\mu}}$ to distinguish it from σ^2_{μ}, which is the true, unobserved variance of the innovations in the process (the latter two moments are the same that were used to identify the baseline process). Specifically, we derive:

$$\begin{aligned} \rho_{\Delta a, \Delta a_{-1}} &= \frac{\rho - 1}{2} \\ \rho_{a, a_{-1}} &= \frac{\sigma^2_{\bar{a}} + \rho \sigma^2_{\hat{a}}}{\sigma^2_{\bar{a}} + \sigma^2_{\hat{a}}} \\ \sigma^2_{\hat{\mu}} &= \left(1 - \rho_{a,a_{-1}}\right)^2 \sigma^2_{\bar{a}} + \left(\rho - \rho_{a,a_{-1}}\right)^2 \sigma^2_{\hat{a}} + \sigma^2_{\mu} \end{aligned}$$

where $\sigma^2_{\hat{a}} = \frac{\sigma^2_{\mu}}{1-\rho^2}$. The first equation identifies ρ directly. The second two represent two equations in two unknowns, which can be solved for $\sigma^2_{\bar{a}}$ and σ^2_{μ}. We can then re-estimate the other parameters of the model using this richer process for a_{it} (the remaining moments are unchanged).

We report the results of this estimation in Table 13. The first three columns display the parameters governing the process on productivity. The estimates for the fixed-effect, $\sigma^2_{\bar{a}}$, are significant in both countries – 0.29 and 0.33 in China and the US, respectively. Comparing the parameter estimates with those in Table 2 shows that (i) the persistence of the AR(1) component here is somewhat lower than under the baseline specification – 0.87 vs. 0.91 for China and 0.84 vs. 0.91 for the US and (ii) the volatility of the shocks, σ^2_{μ}, is almost unchanged in both countries.

We report the results for the other parameters in the remaining columns of Table 13. The top panel shows the parameter estimates and the bottom panel the contribution of each factor to observed *arpk* dispersion. In both countries, the estimated adjustment costs, ξ, are slightly lower and the correlated distortion slightly higher (in absolute value) than under the baseline specification in Table 3. The values are almost unchanged for the other factors. The bottom panel of the table shows that our main conclusions regarding the sources of *arpk* dispersion

Table 13: Estimates with Firm Fixed-Effects

Parameters	ρ	σ^2_μ	σ^2_a	ξ	$\mathbb{V}$	γ	σ^2_ε	σ^2_χ
China	0.87	0.14	0.29	0.12	0.10	-0.71	0.00	0.42
US	0.84	0.08	0.33	0.76	0.04	-0.38	0.00	0.30
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$								
China				1.1%	10.3%	47.8%	0.0%	46.1%
US				6.7%	8.1%	18.9%	1.1%	65.9%

continue to hold.

Persistence in distortions. In our baseline setup, the transitory uncorrelated distortion was assumed to be an iid draw in each period. Here, we generalize that formulation and assume that it follows an AR(1) process:

$$\begin{aligned} \tau_{it} &= \gamma a_{it} + \hat{\tau}_{it} + \chi_i \\ \hat{\tau}_{it} &= \rho_\tau \hat{\tau}_{it-1} + \varepsilon_{it} \end{aligned} \tag{44}$$

Compared to the baseline case, there is now an additional parameter, ρ_τ, and therefore the estimation requires more moments. At the end of this appendix, we extend our analytical approach from Section 3 and prove identification of all the parameters under this more general process.[61] In particular, we show that adding the second-order serial correlation of *arpk* (in changes) is sufficient for identification of the new parameter, ρ_τ.[62] Guided by this result, we re-estimated the model adding both the first- and second-order serial correlations of *arpk* as target moments.

We report the results from this estimation in Table 14. The estimates for ρ_τ are essentially zero in both countries, providing support for the baseline iid assumption. In other words, conditional on the fixed and correlated components, the remaining transitory piece of the distortion is extremely short-lived. The remaining parameters are quite close to their baseline values.[63]

The remainder of this appendix proves identification of the model parameters in the case

[61] As before, we cannot identify the fixed-effect, σ^2_χ since second moments in levels are not well defined.

[62] The first-order serial correlation turns out to be a simple transformation of other target moments and thus does not contain new information.

[63] We also estimated another version of the model, where we assumed the uncorrelated component follows an AR(1) without fixed-effects, i.e., we set $\sigma^2_\chi = 0$, and estimated ρ_τ along with the other parameters by targeting the same set of moments as in the baseline analysis. This yielded values of ρ_τ very close to one, again pointing to an extremely persistent component. Further, the estimated magnitude of the uncorrelated component in this case, i.e., $\frac{\sigma^2_\epsilon}{1-\rho^2_\tau}$, was quite close to the baseline estimate for σ^2_χ. The remaining parameters were almost identical in the two versions.

Table 14: Persistence in Distortions

	New Moments		Parameter Estimates					
	$\rho_{arpk,arpk_{-1}}$	$\rho_{arpk,arpk_{-2}}$	ξ	$\mathbb{V}$	γ	σ^2_ε	ρ_τ	σ^2_χ
China	0.90	0.81	0.24	0.09	-0.69	0.00	0.00	0.40
US	0.91	0.83	1.00	0.03	-0.34	0.01	0.00	0.28

that productivity follows a random walk and distortions follow the process in (44).

Following the same steps as in Section 2 of the text, we can derive the firm's investment policy function under this more general structure:

$$k_{it+1} = \psi_1 k_{it} + \psi_2 (1+\gamma) \mathbb{E}_{it} [a_{it+1}] + \psi_3 \hat{\tau}_{it+1} + \psi_4 \chi_i$$

where ψ_1 solves the quadratic equation

$$\xi \left(\beta \psi_1^2 + 1\right) = \psi_1 \left((1+\beta)\xi + 1 - \alpha\right)$$

and

$$\begin{aligned} \psi_2 &= \frac{1}{1-\alpha-\beta\xi\psi_1+\xi} \\ \psi_3 &= \frac{1}{1-\alpha+\beta\xi(1-\psi_1-\rho_\tau)+\xi} \\ \psi_4 &= \frac{1}{1-\alpha+\xi(1-\beta\psi_1)} \end{aligned}$$

This law of motion is similar to (7) and (8) with a few modifications to the coefficients: ψ_1 and ψ_4 are the same as before, but ψ_2 here corresponds to the case where $\rho = 1$ and ψ_3 is generalized to allow for $\rho_\tau \neq 0$.

Investment is given by:

$$\begin{aligned} \Delta k_{it+1} &= \Delta k_{it} + \psi_2 (1+\gamma) \Delta \mathbb{E}_{it} [a_{it+1}] + \psi_3 \Delta \hat{\tau}_{it+1} \\ &= \psi_1 \Delta k_{it} + \psi_2 (1+\gamma) \left((1-\phi)\mu_{it} + \phi \mu_{it+1} + \phi (e_{it+1} - e_{it})\right) + \psi_3 \left((\rho_\tau - 1)\hat{\tau}_{it} + \varepsilon_{it+1}\right) \end{aligned}$$

where $1-\phi = \frac{\mathbb{V}}{\sigma^2_\mu}$. From here, we can derive the following four moments: the variance of investment, σ^2_k, the autocovariance of investment, $\sigma_{k,k_{-1}}$, the coefficient from a regression of $\Delta arpk_{it}$ on Δa_{it} ($\lambda_{arpk,a}$), and the covariance of investment with lagged innovations in productivity,

$\sigma_{k,a_{-1}}$:

$$\begin{aligned}
\sigma_k^2 &= \psi_1^2\sigma_k^2 + (1+\gamma)^2\psi_2^2\sigma_\mu^2 + \frac{2(1-\psi_1)\psi_3^2\sigma_\varepsilon^2}{(1+\rho_\tau)(1-\psi_1\rho_\tau)} && (45)\\
\sigma_{k,k_{-1}} &= \psi_1\sigma_k^2 - \frac{(1-\rho_\tau)\psi_3^2\sigma_\varepsilon^2}{(1+\rho_\tau)(1-\psi_1\rho_\tau)} && (46)\\
\lambda_{arpk,a} &= 1-(1-\alpha)(1+\gamma)\psi_2\phi && (47)\\
\sigma_{k,a_{-1}} &= (1-\phi(1-\psi_1))(1+\gamma)\psi_2\sigma_\mu^2 && (48)
\end{aligned}$$

Here, we have one new parameter, ρ_τ, and so will need an additional moment. It turns out that the first-order serial correlation of *arpk* does not contain any additional information. To see this, we can derive

$$\begin{aligned}
\sigma_{arpk,arpk_{-1}} &\equiv \operatorname{cov}(\Delta arpk_{it}, \Delta arpk_{it-1})\\
&= \operatorname{cov}(\mu_{it} - (1-\alpha)\Delta k_{it}, \mu_{it-1} - (1-\alpha)\Delta k_{it-1})\\
&= (1-\alpha)^2\sigma_{k,k_{-1}} - (1-\alpha)\sigma_{k,a_{-1}}
\end{aligned}$$

which shows that the moment is a simple combination of two moments we have previously used.

Similarly, the variance of $\Delta arpk_{it}$ is

$$\begin{aligned}
\sigma_{arpk}^2 &\equiv \operatorname{var}(\Delta arpk_{it}) = \operatorname{var}(\mu_{it} - (1-\alpha)\Delta k_{it})\\
&= \sigma_\mu^2 + (1-\alpha)^2\sigma_k^2 - (1-\alpha)\lambda_{arpk,a}\sigma_\mu^2
\end{aligned}$$

which, again, is simply a combination of other moments we have already used.

However, the second-order serial correlation, $\sigma_{arpk,arpk_{-2}}$, does contain new information:

$$\begin{aligned}
\sigma_{arpk,arpk_{-2}} &\equiv \operatorname{cov}(\Delta arpk_{it}, \Delta arpk_{it-2})\\
&= \operatorname{cov}(\mu_{it} - (1-\alpha)\Delta k_{it}, \mu_{it-2} - (1-\alpha)\Delta k_{it-2})\\
&= (1-\alpha)^2\operatorname{cov}(\Delta k_{it}, \Delta k_{it-2})\\
&= (1-\alpha)^2\left(\psi_1\sigma_{k,k_{-1}} - \frac{\rho_\tau\psi_3^2\sigma_\varepsilon^2}{(1+\rho_\tau)(1-\psi_1\rho_\tau)}\right) && (49)
\end{aligned}$$

Substituting expressions (46)-(48) into (45), we obtain:

$$\sigma_k^2 = \psi_1^2\sigma_k^2 + \left(\frac{\sigma_{k,a_{-1}}}{\sigma_\mu^2} + \left(\frac{1-\lambda_{arpk,a}}{1-\alpha}\right)(1-\psi_1)\right)^2\sigma_\mu^2 + \frac{2\left(\psi_1\sigma_k^2 - \sigma_{k,k_{-1}}\right)(1-\psi_1)}{1-\rho_\tau}$$

This is one equation in two unknowns, ψ_1 and ρ_τ. Next, substitute (46) into (49):

$$\sigma_{arpk,arpk_{-2}} = (1-\alpha)^2 \left(\psi_1 \sigma_{k,k_{-1}} - \frac{\rho_\tau}{1-\rho_\tau} \left(\psi_1 \sigma_k^2 - \sigma_{k,k_{-1}} \right) \right)$$

which gives a second equation in the two unknowns. The solution to the two equations yields ψ_1 and ρ_τ. From ψ_1, we can compute ξ and ψ_2. Along with ρ_τ, these give ψ_3. The remaining moment conditions yield the remaining parameters, γ, σ_ε^2 and $\mathbb{V}$.

I.3 Lower Elasticity of Substitution

There is no clear consensus on the appropriate value for the elasticity of substitution parameter, θ, which is set to 6 in our analysis. Estimates generally range between 3 and 10 (see, e.g., Broda and Weinstein (2006)). Studies on firm dynamics tend to use values closer to 6. For example, Cooper and Haltiwanger (2006) estimate a demand elasticity among US manufacturing firms of just about 6; the curvature parameter in Atkeson and Kehoe (2005) is 0.85, which corresponds to $\theta = 7$ in our setup. The literature on misallocation, following Hsieh and Klenow (2009), often uses a lower value, $\theta = 3$.

To investigate the robustness of our conclusions to this parameter, we re-did our analysis using $\theta = 3$. In conjunction with the production function elasticities, $\hat{\alpha}_1$ and $\hat{\alpha}_2$, reported in Table 1, this yields values of α of 0.4 in the US and 0.5 in China (compared to 0.62 and 0.71 in the baseline analysis). We have recomputed the target moments under these new values (recall that moments in productivity depend on the curvature parameter) – Table 15 – and re-estimated the model targeting these moments – Table 16.

The moments in Table 15 show largely the same patterns as those in Table 2 and many of the point estimates change little – for example, investment growth in China is more correlated with lagged shocks, is more volatile and less serially correlated and shows a higher correlation between *arpk* and productivity (although this figure is somewhat lower in both countries than under the baseline α). The extent of the overall dispersion in the *arpk* is almost identical to the baseline, since this figure is independent of α (there is a negligible difference in this value for the US due to the trimming of outliers).

Table 15: Moments – $\theta = 3$

	ρ	σ_μ^2	$\rho_{\iota,a_{-1}}$	$\rho_{\iota,\iota_{-1}}$	$\rho_{arpk,a}$	σ_ι^2	σ_{arpk}^2
China	0.92	0.13	0.21	−0.36	0.56	0.15	0.92
US	0.95	0.11	0.03	−0.30	0.28	0.06	0.46

The estimation results in Table 16 also point to very similar patterns regarding the sources

of *arpk* dispersion: adjustment/information frictions explain only a modest share, leaving a large role for other factors. The estimated adjustment costs are slightly higher in China and lower in the US. The opposite is true for the level of uncertainty. The estimates for correlated (permanent) factors are slightly smaller (larger) in both countries. Importantly, correlated factors are estimated to be much more severe in China than the US. Of course, θ also plays a significant role in determining the magnitude of aggregate productivity losses from a given amount of *arpk* dispersion, as expression (9) reveals. Thus, with $\theta = 3$, the implied TFP losses from all of the factors are smaller than the baseline. However, these losses remain substantial and differ across the two countries, totaling about 50% in China and 12% in the US.

Table 16: Contributions to 'Misallocation' – $\theta = 3$

			Other Factors		
	Adjustment Costs	Uncertainty	Correlated	Transitory	Permanent
Parameters	ξ	$\mathbb{V}$	γ	σ^2_ε	σ^2_χ
China	0.21	0.08	−0.49	0.00	0.64
US	0.66	0.04	−0.10	0.00	0.38
$\Delta\sigma^2_{arpk}$					
China	0.01	0.08	0.24	0.00	0.64
US	0.02	0.04	0.01	0.00	0.38
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$					
China	1.1%	8.3%	26.4%	0.0%	69.2%
US	4.3%	9.4%	1.4%	0.0%	82.5%
Δa					
China	0.01	0.04	0.12	0.00	0.32
US	0.01	0.01	0.00	0.00	0.10

Taken together, our findings in Table 16 confirm that our main conclusions regarding the sources of *arpk* dispersion are not overly sensitive to the value of the elasticity of substitution. While the exact productivity costs of that dispersion does depend on this parameter (and more generally, on the extent of curvature), all the cases we have examined suggest they can be substantial.

I.4 Alternative Targets: Investment Moments

In this appendix, we re-estimate our model targeting the autocorrelation and variance of investment in levels, rather than growth rates. The values of these moments are 0.25 and 0.04, respectively, in the US and 0.04 and 0.08 in China. The other target moments are the same as

in Table 2. Table 17 reports the results. A comparison to Table 3 shows that the parameter estimates are quite close to the baseline, as are the contributions to *arpk* dispersion – adjustment costs and uncertainty account for between 15% and 20% of σ^2_{arpk} in the two countries, correlated factors play a large role in China and less so in the US, while fixed factors are quite significant in both countries.

Table 17: Using Moments from Investment in Levels

			Other Factors		
	Adjustment Costs	Uncertainty	Correlated	Transitory	Permanent
Parameters	ξ	$\mathbb{V}$	γ	σ^2_ε	σ^2_χ
China	0.37	0.11	-0.72	0.02	0.38
US	1.77	0.04	-0.31	0.19	0.28
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$					
China	4.3%	11.9%	48.9%	2.5%	40.8%
US	12.1%	8.1%	13.2%	42.4%	62.8%

I.5 Measurement of Capital

Our baseline analysis uses reported book values of firm-level capital stocks. Here, we use the perpetual inventory method to construct an alternative measure of capital for the US firms. To do this, we follow the approach in Eberly et al. (2012). Here, we briefly describe the procedure and refer the reader to that paper for more details. We use the book value of capital in the first year of our data as the starting value of the capital stock and use the recursion:

$$K_{it} = \left(K_{it-1} \frac{P_{Kt}}{P_{Kt-1}} + I_{it} \right) (1 - \delta_j)$$

to estimate the capital stock in the following years, where I_t is measured as expenditures on property, plant and equipment, P_K is the implicit price deflator for nonresidential investment, obtained from the 2013 Economic Report of the President, Table 7, and δ_j is a four-digit industry-specific estimate of the depreciation rate. We calculate the useful life of capital goods in industry j as $L_j = \frac{1}{N_j} \sum_{N_j} \frac{PPENT_{it-1} + DEPR_{it-1} + I_{it}}{DEPR_{it}}$ where N_j is the number of firms in industry j, $PPENT$ is property, plant and equipment net of depreciation and $DEPR$ is depreciation and amortization. The implied depreciation rate for industry j is $\delta_j = \frac{2}{L_j}$. We use the average value for each industry over the sample period.

Table 18 reports the estimation results. The parameters governing firm productivity, ρ and σ^2_μ, are quite close to the baseline values, as is the total amount of observed *arpk* dispersion,

σ^2_{arpk}.[64] The autocorrelation of investment growth is somewhat higher and its volatility somewhat lower, which together lead to a higher estimate of the adjustment cost parameter, ξ. This is reflected in the higher contribution of these costs to *arpk* dispersion, which is about 27% of the total (compared to 11% in the baseline). The estimated degree of uncertainty is close to the baseline value. Together, these two forces account for about 33% of the observed *arpk* dispersion, compared to about 18% under our baseline calculations. Thus, our finding of a key role for other firm-specific factors continues to hold – these factors account for roughly two-thirds of σ^2_{arpk}. The largest component shows up as a permanent factor that is orthogonal to firm productivity. The time-varying correlated and uncorrelated components contribute only modestly.

Table 18: Perpetual Inventory Method for Capital - US firms

Moments	ρ	σ^2_μ	$\rho_{\iota,a_{-1}}$	$\rho_{\iota,\iota_{-1}}$	$\rho_{arpk,a}$	σ^2_ι	σ^2_{arpk}
	0.94	0.07	0.15	−0.18	0.55	0.01	0.43
Parameters			ξ	$\mathbb{V}$	γ	σ^2_ε	σ^2_χ
			5.80	0.02	−0.17	0.05	0.26
Aggregate Effects							
$\Delta\sigma^2_{arpk}$			0.12	0.02	0.02	0.05	0.26
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$			27.5%	5.7%	4.3%	12.8%	59.9%
Δa			0.05	0.01	0.01	0.02	0.11

Similar to the exercise in Appendix I.4, we have also re-estimated the model using this alternative measure of firm-level capital stocks and targeting the autocorrelation and variability of investment in levels, rather than growth rates. The results are reported in Table 19. The estimates are broadly in line with those in Table 18 and are extremely close to the baseline ones in Table 3. To see why, we have also computed the implied values of the autocorrelation and variance of investment using the parameter estimates from Table 18. This gives values of 0.69 and 0.02, respectively, compared to the empirical values of 0.57 and 0.02. Because the estimation in Table 18 already matches these (non-targeted) moments fairly closely, explicitly targeting them does not have a large effect.

Table 19: Perpetual Inventory Capital and Investment in Levels - US

	ξ	$\mathbb{V}$	γ	σ^2_ε	σ^2_χ
Parameters	1.65	0.03	−0.32	0.00	0.28
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$	12.0%	6.9%	14.0%	0.7%	64.3%

[64] Even in the last year of the sample, the correlation of the two capital stock measures exceeds 0.95.

I.6 Sectoral Analysis

In this appendix, we repeat our analysis for US firms at a disaggregated sectoral level, allowing for sector-specific structural parameters.

We begin by computing sector-specific α's (curvature in the profit function) using data on value-added and compensation of labor by sector from the Bureau of Economic Analysis, Annual Industry Accounts.[65] To match the SIC (or NAICS) classifications in Compustat, we compute labor's share of value-added for the 9 major sectors of the industrial classification – Agriculture, Forestry and Fishing; Mining; Construction; Manufacturing; Transportation, Communications and Utilities; Wholesale Trade; Retail Trade; Finance, Insurance and Real Estate; Services.[66] To translate these shares into a value of α, note that under our assumptions of monopolistic competition and constant returns to scale in production, labor's share of value-added is equal to $LS = \frac{\theta-1}{\theta}(1-\hat{\alpha}_1)$ where $1-\hat{\alpha}_1$ is the labor elasticity in the production function. Then, solving for $\hat{\alpha}_1$ and substituting into the definition of α, we have

$$\alpha = \frac{\alpha_1}{1-\alpha_2} = \frac{\frac{\theta-1}{\theta} - LS}{1-(1-\hat{\alpha}_1)\frac{\theta-1}{\theta}} = \frac{\frac{\theta-1}{\theta} - LS}{1-LS}$$

Implementing this procedure yields the values of α in the top panel of Table 20.[67]

Next, we re-compute our cross-sectional moments for each sector, using the values of α to estimate firm-level productivities. We continue to control for time and industry fixed-effects to extract the firm-specific components of the series (there are multiple four-digit industries within each sector). We report the target moments in the first panel of Table 20. We then estimate the model separately for each sector, allowing the structural parameters governing the various sources of *arpk* dispersion to vary across sectors. The resulting parameter estimates are presented in the second panel of the table and the implied contribution of each factor to *arpk* dispersion in the last two panels.

There is some heterogeneity across the sectors, both in the overall extent of *arpk* dispersion as well as in the estimates for the underlying factors. For example, adjustment costs are largest in manufacturing, where they account for as much as 20% of the observed dispersion and are smallest in FIRE. But, overall, the main message from our baseline analysis continues to hold – adjustment and information frictions, although significant, do not create a lot of *arpk*

[65] The data are available at `https://www.bea.gov/industry/iedguide.htm`.

[66] Most of these correspond one-for-one with sectors reported by the BEA data. There, Transportation and Utilities are reported separately, as are several subcategories of services, which we aggregate. The only sector we were unable to include from the BEA data was Information, as it does not line up one-for-one with an SIC or NAICS category. The shares are calculated as the average over the most recent period available, 1998-2011 (which roughly lines up with the period of the firm-level data, 1998-2009).

[67] We have also calculated this value for the entire US economy by summing across all the sectors reported by the BEA. This gives an aggregate labor share of 0.56 and an implied α of 0.62, exactly our baseline value.

dispersion, leaving a substantial role for other firm-specific factors. While the results point to some heterogeneity in the correlation structure of these factors, the permanent component seems to play a key role across all sectors.

J Estimates for Other Countries/Firms

In this appendix, we apply our empirical methodology to two additional countries for which we have firm-level data - Colombia and Mexico - as well as to publicly traded firms in China.

The Colombian data come from the Annual Manufacturers Survey (AMS) and span the years 1982-1998. The AMS contains plant-level data and covers plants with more than 10 employees, or sales above a certain threshold (around $35,000 in 1998, the last year of the data). We use data on output and capital, which includes buildings, structures, machinery and equipment. The construction of these variables is described in detail in Eslava et al. (2004). Plants are classified into industries defined at a 4-digit level. The Mexican data are from the Annual Industrial Survey over the years 1984-1990, which covers plants of the 3200 largest manufacturing firms. They are also at the plant-level. We use data on output and capital, which includes machinery and equipment, the value of current construction, land, transportation equipment and other fixed capital assets. A detailed description is in Tybout and Westbrook (1995). Plants are again classified into industries defined at a 4-digit level. Data on publicly traded Chinese firms are from Compustat Global. Due to a lack of a sufficient time-series for most firms, we focus on single cross-section for 2015 (the moments use data going back to 2012). Similarly, due to the sparse representation of many industries, we focus on those with at least 20 firms. For all the datasets, we compute the target moments following the same methodology as outlined in the main text of the paper. Our final samples consist of 44,909 and 3,208 plant-year observations for Colombia and Mexico, respectively, and 1,055 firms in China.

Table 21 reports the moments and estimated parameter values for these sets of firms, as well as the share of *arpk* dispersion arising from each factor and the effects on aggregate productivity. The results are quite similar to those for Chinese manufacturing firms in Table 3 in the main text. The contribution of adjustment costs and uncertainty to observed *arpk* dispersion is rather limited, and that of uncorrelated transitory factors negligible - across these sets of firms, a large portion of the observed dispersion stems from correlated and permanent firm-specific factors.

Table 20: Sector-Level Results

Moments	α	ρ	σ^2_μ	$\rho_{\iota,a_{-1}}$	$\rho_{\iota,\iota_{-1}}$	$\rho_{arpk,a}$	σ^2_ι	σ^2_{arpk}
Agr., Forestry and Fishing	0.77	0.92	0.11	0.13	−0.37	0.92	0.03	0.61
Mining	0.76	0.91	0.10	0.16	−0.29	0.74	0.07	0.35
Construction	0.49	0.93	0.15	0.17	−0.28	0.71	0.07	0.69
Manufacturing	0.59	0.94	0.08	0.10	−0.32	0.50	0.05	0.43
Trans., Comm. and Utilities	0.67	0.94	0.04	0.13	−0.32	0.58	0.03	0.38
Wholesale Trade	0.65	0.94	0.08	0.18	−0.31	0.67	0.05	0.57
Retail Trade	0.61	0.96	0.02	0.20	−0.30	0.25	0.02	0.20
Finance, Insurance and Real Estate	0.78	0.90	0.09	0.28	−0.32	0.77	0.07	0.61
Services	0.38	0.95	0.10	0.03	−0.28	0.31	0.08	0.53
Parameters				ξ	$\mathbb{V}$	γ	σ^2_ε	σ^2_χ
Agr., Forestry and Fishing				0.83	0.05	−0.78	0.01	0.09
Mining				0.49	0.04	−0.56	0.00	0.13
Construction				0.65	0.08	−0.50	0.00	0.32
Manufacturing				3.35	0.03	−0.17	0.18	0.28
Trans., Comm. and Utilities				0.55	0.02	−0.55	0.00	0.25
Wholesale Trade				0.55	0.04	−0.54	0.00	0.30
Retail Trade				1.97	0.01	−0.07	0.03	0.17
Finance, Insurance and Real Estate				0.18	0.06	−0.80	0.00	0.26
Services				0.81	0.04	−0.14	0.00	0.44
$\Delta\sigma^2_{arpk}$								
Agr., Forestry and Fishing				0.07	0.05	0.45	0.01	0.09
Mining				0.06	0.04	0.19	0.00	0.13
Construction				0.04	0.08	0.26	0.00	0.32
Manufacturing				0.09	0.03	0.02	0.18	0.28
Trans., Comm. and Utilities				0.01	0.02	0.10	0.00	0.25
Wholesale Trade				0.02	0.04	0.20	0.00	0.30
Retail Trade				0.02	0.01	0.00	0.03	0.17
Finance, Insurance and Real Estate				0.01	0.06	0.31	0.00	0.26
Services				0.02	0.04	0.02	0.00	0.44
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$								
Agr., Forestry and Fishing				0.11	0.08	0.74	0.02	0.15
Mining				0.18	0.10	0.54	0.00	0.37
Construction				0.05	0.11	0.37	0.00	0.47
Manufacturing				0.21	0.07	0.05	0.41	0.63
Trans., Comm. and Utilities				0.03	0.05	0.26	0.01	0.65
Wholesale Trade				0.04	0.07	0.35	0.00	0.54
Retail Trade				0.08	0.05	0.01	0.14	0.85
Finance, Insurance and Real Estate				0.02	0.09	0.51	0.00	0.42
Services				0.05	0.07	0.04	0.00	0.83

Table 21: Additional Countries/Firms

Moments	ρ	σ^2_μ	$\rho_{\iota,a_{-1}}$	$\rho_{\iota,\iota_{-1}}$	$\rho_{arpk,a}$	σ^2_ι	σ^2_{arpk}
Colombia	0.95	0.09	0.28	−0.35	0.61	0.07	0.98
Mexico	0.93	0.07	0.17	−0.39	0.69	0.02	0.79
China Compustat	0.96	0.04	0.30	−0.42	0.76	0.04	0.41
Parameters			ξ	$\mathbb{V}$	γ	σ^2_ε	σ^2_χ
Colombia			0.54	0.05	−0.55	0.01	0.60
Mexico			0.13	0.04	−0.82	0.00	0.42
China Compustat			0.15	0.03	−0.69	0.00	0.18
$\Delta\sigma^2_{arpk}$							
Colombia			0.02	0.05	0.30	0.01	0.60
Mexico			0.00	0.04	0.36	0.00	0.42
China Compustat			0.00	0.03	0.22	0.00	0.18
$\frac{\Delta\sigma^2_{arpk}}{\sigma^2_{arpk}}$							
Colombia			2.5%	5.6%	30.9%	0.7%	61.3%
Mexico			0.5%	4.9%	44.9%	0.0%	52.8%
China Compustat			0.8%	6.3%	54.0%	0.2%	43.7%
Δa							
Colombia			0.01	0.02	0.13	0.00	0.26
Mexico			0.00	0.02	0.16	0.00	0.18
China Compustat			0.00	0.02	0.19	0.00	0.16