Good Dispersion, Bad Dispersion[*,**]

Matthias Kehrig[†] and Nicolas Vincent[‡]

November 7, 2020

Abstract

We document that most dispersion in marginal revenue products of inputs occurs across plants *within* firms rather than between firms. This is commonly thought to reflect misallocation, i.e., dispersion is "bad." However, we show conceptually that eliminating frictions may *increase* productivity dispersion and raise overall output, i.e., dispersion is "good." In a model of multi-plant firms, we argue that good dispersion represents one quarter of the total variance of revenue products in U.S. manufacturing. In emerging economies, in contrast, we find much less scope for good dispersion. This implies that the gains from eliminating distortions are larger than previously thought.

Keywords: Misallocation, Productivity Dispersion, Multi-Plant Firms, Internal Capital Markets.
Jel Codes: E2, G3, L2, O4.

[*] We thank seminar participants at PSE, EIEF, Banco de Mexico, CREI, Birmingham, Wharton, Bonn, Duke, Oxford, Penn State, the Bank of Canada, the Dutch Central Bank, the Bank of Finland, the NBER EF&G meeting, the NBER Summer Institute, the Barcelona Summer Forum, the SED Annual Meeting, the 12th Macro-Finance Society Workshop and the BI Conference on Asset Pricing Research as well as Andrew Abel, Christian Bayer, David Berger, Ilan Cooper, Russ Cooper, Andrea Eisfeldt, Xavier Giroud, João Gomes, John Haltiwanger, Hugo Hopenhayn, Pete Klenow, Virgiliu Midrigan, Ben Moll, Gordon Phillips, Vincenzo Quadrini, Cian Ruane, Yongseok Shin, Michèle Tertilt and Venky Venkateswaran for helpful comments and discussions. Vincent gratefully acknowledges financial support from the Fondation HEC Montréal.

[**] Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. The latest version of this paper can be downloaded at https://sites.google.com/site/matthiaskehrig/research.

[†] Corresponding author. Duke University, NBER and CEPR matthias.kehrig@gmail.com.

[‡] Institute of Applied Economics, HEC Montréal, nicolas.vincent@hec.ca.

1 Introduction

A considerable body of recent research has documented a large, persistent and ubiquitous degree of productivity dispersion across production units, leading to a revival of interest in the causes and consequences of resource misallocation. In their seminal work, Hsieh and Klenow (2009) argue that resource reallocation in China or India that would render dispersion similar to that in the United States would yield substantial aggregate output gains. Multiple papers have since expanded and refined their methodology. Yet the common view in this literature remains that a high level of productivity dispersion is a sign of resource misallocation and therefore reduced welfare; i.e., dispersion is "bad."

We challenge this interpretation: we show that the relationship between frictions and dispersion is non-monotonic. Specifically, in the context of a simple conceptual model, we highlight situations in which *relaxing a friction* or constraint leads to increased output through a more efficient allocation of resources, despite generating *higher dispersion* of marginal products of inputs. In this context, more dispersion is "good." As a consequence, the level of fundamental frictions is non-monotonically related to empirically observable productivity dispersion, rendering the latter a potentially misleading indicator of misallocation. While this insight is broad and could be applied in other contexts, we focus on the role of multi-plant firms in shaping economy-wide productivity dispersion. This setting is particularly relevant for two reasons.

First, even if the literature has historically made little distinction between them, firms and plants are fundamentally different institutions. While firms compete for resources in markets, they act as planners in allocating these resources across their plants. This latter allocation activity is economically predominant: For instance, multi-plant firms account for most of aggregate output and investment in the U.S. manufacturing sector. As such, work studying the sources of productivity dispersion should take into account how firm-internal decisions differ from those across firms.

Second, using plant-level data from the U.S. Economic Census, we document that almost two-thirds of the overall dispersion in marginal revenue products of capital originates across plants operated by the same firm rather than between firms. This novel empirical finding highlights the importance of paying attention to the role played by the allocative decisions of firms. It also poses a theoretical challenge: Why do firms tolerate surprisingly high levels of dispersion in marginal revenue products across their plants and forgo seemingly large output gains? Why do they not reallocate resources across their plants in order to reduce within-firm productivity dispersion? Does the firm allocate resources less efficiently than markets? Understanding how firms work differently from markets may therefore shed new light on the causes of resource (mis)allocation.

To answer these questions, we build a quantitative model of capital allocation for a multi-plant firm that faces various constraints. Beyond its multi-plant nature, our framework is standard and embeds many frictions and imperfections that have been suggested in models of investment (see Caballero (1999) for an overview). It includes "technological" frictions, such as a fixed investment adjustment cost and capital irreversibility (see, among others, Abel and Eberly (1994), Caballero et al. (1995), Doms and Dunne (1998), Cooper and Haltiwanger (2006) and Gourio and Kashyap

(2007)), which are more relevant at the level of the plant, as well as external financing constraints (see, among others, Fazzari et al. (1988) or Gilchrist and Himmelberg (1995)), which affect the firm as a whole.[1] In our model, the firm organizes internal and costly external financing across plants that face both fixed and convex costs of adjustment. The model is calibrated using moments from the Annual Survey of Manufactures collected by the U.S. Census Bureau and standard parameter values in the literature.

Using our model simulations, we then address the key question of our study: Can higher dispersion be associated with a more efficient allocation? In our setting, the answer is unambiguously "yes." To assess the potential role of this *good dispersion*, we compare the dispersion of marginal products of capital as well as aggregate quantities in two economies: one in which multi-plant firms can pool resources and allocate them freely across their production plants, and one in which these internal capital markets (ICM) are shut down due to frictions, forcing the firm to effectively operate its plants as standalone units. As expected, we find that allowing for functional internal capital markets leads to higher aggregate capital and output, to the order of about 4% and 3%, respectively. More surprisingly, this more efficient allocation of resources is accompanied by a variance of the marginal products of capital that is 32% *higher* than in the economy with standalone plants.

While this non-standard relationship between economic activity and dispersion may seem surprising, the forces behind it are intuitive: When firms are constrained in their access to external funds, they leverage internal capital markets and focus investment on only a few plants *even if the expected rate of returns of all plants are identical*. In the next period, the firm will concentrate its internal financial resources on another set of plants. Such an internal reallocation of funds will continue until the firm has carried out investment projects in all its plants. This "staggered investment" policy therefore leads to a rise in the dispersion of both investment rates and marginal revenue products of capital within the firm. In the stochastic equilibrium with frequent shocks, this *good dispersion* is persistent and never vanishes. In addition, these firms exhibit less correlated investment rates across their plants than their unconstrained counterparts. Using plant-level data from the U.S. Annual Survey of Manufactures, we find empirical support for these predictions at the micro level.

Note that, while we illustrate our mechanism in the context of constraints in external financial markets, it can arise due to any firm-level factor in scarce supply: for instance, headquarter support such as engineers and implementation managers might be limited to only a few plants in a given year, or costly information acquisition may force the marketing team to focus on one market segment at a time, while remaining at first rationally inattentive to others. In Section 2.2 we illustrate a large set of factors that will result in the staggered investment policy and *good dispersion*.

Our quantitative exercises suggest two things. First, the interaction of plant- and firm-level frictions is quite powerful in generating substantial within-firm heterogeneity and dispersion in general. Second, the dispersion of marginal revenue products of capital is not necessarily an indicator of

[1] Gomes (2001), Khan and Thomas (2013) and Eisfeldt and Muir (2016) are examples of papers that combine real and financial frictions in a unified model of a firm operating a single plant.

misallocation or inefficient investment policies. In our context, *good dispersion* arises because the firm manages to partially overcome external financial frictions and reallocate resources to a limited extent. These constraint-efficient reallocation decisions result in a second-best allocation of the firm's resources, which increases output and welfare. Hence, we view our results as a cautionary tale about the risks of interpreting higher productivity dispersion as a sign of resource misallocation.

Our findings have potentially important implications for the literature on resource allocation. Starting with Hsieh and Klenow (2009), a large body of work has studied productivity dispersion across plants. This literature typically envisions frictions that by their nature increase dispersion in marginal revenue products, decrease output and therefore fall under our label of *bad dispersion*; it does not explicitly entertain the presence of *good dispersion* that we described above. Note that this distinction does not depend on the type of friction. Even if they are technological in nature, such as fixed investment cost and time-to-build (Asker et al. (2014)), a tightening of the frictions leads to more dispersion and lower output. In our setting, we show instead that tightening frictions may lead to less dispersion and vice versa. Bayer et al. (2015) distinguish the long-term from the temporary components of dispersion, while Buera et al. (2011), Khan and Thomas (2013) and Moll (2014) have studied the impact of financial frictions on misallocation, usually ignoring their effects within firms that operate several plants. In related work, Midrigan and Xu (2014) study the effects of financial frictions on firm entry and factor misallocation across firms. While we abstract from the entry channel, a modified version of the latter effect is present in our analysis, albeit in a framework in which financially constrained firms operate several plants and can overcome external financial frictions by internally reallocating financial resources.

Building on these insights, we argue that Hsieh and Klenow (2009) may in fact have *underestimated* the gains from reallocation in emerging economies such as India or China. When we make the U.S. economy more comparable to that of emerging economies by populating it with single-plant instead of multi-plant firms, dispersion in the U.S. is even smaller relative to that of India. This, in turn, implies that the distortions inferred to explain the observed dispersion and thus output losses from a distorted resource allocation are arguably higher than initially estimated. A quantitative exercise suggests that previous work may have missed between one-tenth and one-third of the output benefits from reallocation because it ignored the beneficial effects of reallocation within firms.

Ultimately, we also see our project as a first step toward modelling how the organizational structure of a firm may impact the micro-level adjustment of capital, as well as understanding the role of firms for efficiency. While most research ignores the within-firm dimension of decision making, some theoretical research has been done on the efficiency of internal versus external capital markets: Stein (1997) and Malenko (2016) study mostly principal-agent problems between a firm's owner and manager in a single-plant setup. Gertner et al. (1994); Scharfstein and Stein (2000) show that division managers may exploit imperfect monitoring by firm headquarters to build up "inefficient empires," resulting in lower firm values Rajan et al. (2000). Through the lens of these frameworks, diversified firms with complex internal capital markets will suffer from more acute agency problems, further misallocation and, consequentially, more *bad dispersion* within firms. While all these factors may be present, we are the first to explore the potential role of *good dispersion* in this context.

Eisfeldt and Papanikolaou (2013) stress the importance of organizational or intangible capital at the firm level in order to understand a firm's productivity, albeit without the multi-plant dimension we are interested in. With the exception of Lamont (1997), Schoar (2002), Giroud (2013), Matvos and Seru (2014) and Giroud and Mueller (2015), empirical research on within-firm dynamics is scarce and often limited to studying major business divisions of conglomerates. An exception is Giroud and Mueller (2019), whose work is closely related to ours. They show empirically how local shocks propagate through the firm's internal organization, and that the reaction of other establishments is only significant if the parent is financially constrained. Finally, while we take the organizational structure of the firm as given, Ševčík (2015) considers the endogenous formation of multi-plant firms (which he calls "business groups").

Our paper is organized as follows. In Section 2, we show evidence on the importance of the within-firm dimension for the dispersion of marginal revenue products of capital and investment; then we illustrate theoretically how relaxing frictions within a firm may increase rather than decrease dispersion. Section 3 describes our multi-plant model of a firm that faces an external financing constraint. Section 4 conducts various quantitative exercises geared toward understanding the nature of productivity dispersion and provides supporting evidence. Section 5 presents an application to the setting in Hsieh and Klenow (2009), and Section 6 concludes.

2 Motivation

In this section, we motivate both empirically and theoretically our subsequent quantitative work. As discussed in the introduction, many studies have documented the ubiquitous presence of a large and persistent dispersion of marginal revenue products of inputs across production units. We first show empirically that in U.S. manufacturing, the majority of the dispersion in both (log) marginal revenue products of capital ($mrpk$) and investment rates (i/k) occurs across plants within firms rather than across firms. In the standard economic models generally used in the literature on misallocation, reallocating capital from low-$mrpk$ to high-$mrpk$ plants through investment activity reduces $mrpk$ dispersion and increases aggregate output.[2] Yet as we show with the help of a simple framework in Section 2.2, the opposite may be true: Relaxing frictions within firms may lead to more dispersion of the marginal revenue products of capital. In the quantitative model of Section 3, we show that our empirical finding could be interpreted as the outcome of an improved allocation rather than as evidence of a suboptimal allocation of resources within the firm.

[2]For details on the assumptions underlying this, see Appendix A.1. Based on Hsieh and Klenow (2009), the misallocation literature usually postulates equalizing revenue total factor productivity ($TFPR$). Like Asker et al. (2014), we focus on the capital allocation problem and hence equalizing $mrpk$. In that context, investment should not necessarily flow toward units with the highest $TFPR$ if they already operate a large capital stock; it should flow to units with the highest expected capital return.

2.1 Empirical motivation: Dispersion within and across firms

Data sources and variables of interest Our data source is collected by the U.S. Census Bureau in the Annual Survey of Manufactures (ASM), which is an annual dataset covering manufacturing businesses described in detail in Appendix A.2. The Census Bureau collects its manufacturing data at the level of an "establishment," which is defined as a physical business unit at a single location for which the primary activity is production. In this paper, we generally refer to establishments as "plants." Each plant also carries information about its parent firm, which is defined by Census as a collection of plants under common ownership or control.

Following the literature, we assume a Cobb-Douglas production technology, which is common for all plants in a 4-digit NAICS industry, and approximate the marginal revenue product of plant n in year t with its real value added per unit of capital.[3] We study the variance of its logarithm, $V_t(mrpk_{nt}) \equiv V_t(\log MRPK_{nt}) = V_t(\log(y_{nt}/k_{nt}))$, within a 4-digit NAICS industry and aggregate industries using value-added weights, as detailed in Appendix A.5.

Dispersion in U.S. manufacturing First, overall dispersion in marginal revenue products across plants is large, as shown in Table 1. In the average industry and year, the standard deviation of its logarithm is 0.9. This means that a plant that is one standard deviation above the mean produces $e^{0.9} \approx 2.5$ times the value added as the average plant with the same capital stock; the difference between the plant at the 90th percentile and that at the 10th percentile even implies an $e^{2.12} \approx 8.3$-fold value-added difference given the same capital stock.

Table 1: Cross-sectional moments of capital and investment

Variable	*Cross-sectional moments*					
	Mean	StD	IDR	Skew-ness	Kelley Skewn.	Excess Kurtosis
$mrpk$		0.905	2.120	0.634	0.128	1.978
		(0.013)	(0.032)	(0.028)	(0.010)	(0.085)
i/k	0.112	0.362	0.175	6.113	0.479	57.204
	(0.015)	(0.093)	(0.008)	(0.099)	(0.008)	(2.378)

Note: Data consist of our benchmark panel comprising annual plant-level data from the ASM 1972-2009. Moments are computed in a given year and 4-digit NAICS industry first before being aggregated by industry and then averaged across years. For details see Appendix A.4.

Interestingly, Table 1 also indicates that the cross-sectional distribution of $mrpk$ is positively skewed. The standard coefficient of skewness is 0.634, while the quantile-based Kelley skewness measure is 0.128 on average.[4] The latter moment implies that the top half of the inter-decile range,

[3] Though we study average rather than marginal revenue products of capital, we consider their difference in Appendix A.6.2.

[4] Following Kelley (1947), p. 250, we define the Kelley skewness as $\gamma^{Kelley} = \frac{mrpk^{90}+mrpk^{10}-2mrpk^{50}}{mrpk^{90}-mrpk^{10}}$.

$mrpk^{90} - mrpk^{50}$, is about 29% more spread out than the bottom half, $mrpk^{50} - mrpk^{10}$. As we will later argue, this evidence is supportive of some of our modeling assumptions.

Investment rates also differ substantially across plants, which implies that the allocative activity of capital differs greatly across units within a typical year and industry. The cross-sectional standard deviation of 36% is large, given that the average plant in the economy has an investment rate of 11.2% – an indication of the well-known lumpy nature of investment. This also makes investment rates highly leptokurtic, which is reported in the last column of Table 1.

Under the standard interpretation of *mrpk* dispersion as evidence of misallocation, reallocating capital to high-*mrpk* plants in the same industry could hence result in a considerable boost in aggregate output. We show next that the majority of this dispersion occurs across plants within firms rather than across firms.

Dispersion within and across firms We decompose the total variance of marginal revenue products of capital, denoted by V_t, into two components: the variance between firms, denoted by V_t^B, and the average variance between plants within firms, denoted by V_t^W. To compare sufficiently similar units, we perform our analysis within 4-digit NAICS industries. This means we break up diversified conglomerates along industry lines, thus reducing the scope of actual within-firm dispersion. Our results should thus be regarded as a lower bound on within-firm dispersion. In a given 4-digit NAICS industry, our variance decomposition is then:

$$V_t\left(mrpk_{nt}\right) \equiv V_t = \underbrace{\sum_j \omega_{jt}\left(mrpk_{jt} - mrpk_t\right)^2}_{V_t^B \text{ average between-firm}} + \underbrace{\sum_j \omega_{jt} \underbrace{\sum_{n \in j}^{N_j} \omega_{nt}^j \left(mrpk_{njt} - mrpk_{jt}\right)^2}_{V_{jt}^W \text{ within firm } j}}_{V_t^W \text{ average within-firm}}. \tag{1}$$

The variable $mrpk_{njt}$ denotes the logarithm of the marginal revenue product of capital of plant n belonging to firm j in year t; $mrpk_{jt}$ the average *mrpk* in firm j in an industry; and $mrpk_t$ the average *mrpk* in a given industry. ω_{njt} is the weight of plant n at time t, ω_{jt} that of firm j and $\omega_{nt}^j = \omega_{njt}/\omega_{jt}$ that of plant n just inside firm j. While unweighted dispersion is our benchmark, we also consider capital-weighted dispersion to account for economic relevance. In the former case, we have $\omega_{njt} = 1/N_t$ (where N_t is the number of observations), while in the latter $\omega_{njt} = k_{njt}/k_t$ and accordingly for ω_{jt} and ω_{nt}^j. More details about this decomposition can be found in Appendix A.5, and the results are displayed in Table 2.

The main takeaway from our accounting exercise is that for the full sample (Panel A), about 60% of the dispersion of marginal revenue products of capital and 68% of the dispersion in investment rates in a typical industry occur within firms, with the remainder accounted for by between-firm variations.[5] The within-firm dispersion of *mrpk* is economically large: A plant that is one standard

[5]As Appendix A.3 shows, while multi-plant firms operate only 28% of all plants, they account for roughly 80% of aggregate economic activity.

Table 2: Cross-sectional moments of capital and investment

Variable	*A. Full panel*		*B. Homogeneous Products*	
	Share of variance		Share of variance	
	between firms	within firms	between firms	within firms
$mrpk$	0.399	0.601	0.441	0.559
i/k	0.321	0.679	0.311	0.689

Note: The data underlying Panel A are our benchmark panel comprising annual plant-level data from the ASM 1972-2009. Moments in Panel B are based on a subsample of homogeneous 7-digit SIC products as defined by Foster et al. (2008). Moments are computed for each industry and years first before being aggregated by industry and then averaged across years. For details, see Appendix A.4.

deviation ($0.702 = \sqrt{0.601 \times 0.905^2}$) above the firm's average produces twice the value added with the same capital stock as a plant that would reflect the firm average.

One might worry that this result is driven by residual product heterogeneity within 4-digit NAICS industries. To alleviate this concern, we repeat the decomposition, but this time focus on plants that produce only one physically homogeneous standardized good. We follow Foster et al. (2008) and consider industries that produce almost perfectly homogeneous goods such as cement, sugar, coffee beans, etc.[6] Even if we focus solely on these highly homogeneous industries, the within-firm share of dispersion in marginal revenue products of capital and investment rates displayed in Panel B amounts to 56% and 69%, respectively. The high within-firm share of dispersion does not reflect mechanical aggregation. This exercise, along with additional robustness checks, can be found in Appendix A.7.

In light of the misallocation literature (see Hsieh and Klenow (2009)), our finding that most dispersion occurs within firms may appear surprising: It seems to suggest the presence of particularly large frictions within firms, rendering them an inferior allocation mechanism. In the next subsection, however, we present a simple conceptual framework that shows how the opposite might be true: Relaxing frictions within the firm can increase the dispersion of marginal revenue products.

2.2 Theoretical motivation: *Bad dispersion* vs. *good dispersion*

Hsieh and Klenow's work on distortions and misallocation has been highly influential, spawning a myriad of studies on both the empirical and modeling fronts. Some have tried to map abstract distortions into empirically measurable market imperfections, often with the objective of quantifying potential output gains from eliminating specific imperfections. Others have attempted to clarify the distinction between imperfections and technological constraints (see, among others, Asker et al.

[6]More specifically, these "industries" are defined by the following SIC product codes: Sugar (2061011), Block and Processed Ice (2097011 and 2097051), Gasoline (2911131), Hardwood Flooring (2426111), Concrete (3273000), Whole Bean and Ground Coffee (2095111 and 2095117 & 2095118 – later merged into 2095115 – and 2095121), Carbon Black (2895011 and 2895000), Bread (2051111, later split into 2051121 and 2051122) and Plywood (2435100, later split into 2435101, 2435105, 2435107 and 2435147).

(2014), David and Venkateswaran (2019) and Haltiwanger et al. (2018)). What is common among all these papers is that reducing frictions spurs beneficial reallocation, brings down the dispersion of factor revenue products, and increases aggregate output. Crucially, dispersion is always assumed to be inversely related to aggregate output: Economies with higher dispersion in factor revenue products are thought to be worse off (see Syverson (2011) for a summary of the academic consensus and Dabla-Norris et al. (2015), Cirera and Maloney (2017) and Cusolito and Maloney (2018) for the importance of that view in global policy making). In other words, dispersion is "bad" because it is a symptom of misallocation.

In this section, we show that the opposite can be true: While reducing frictions always improves factor allocation, it may *increase* dispersion rather than reducing it.[7] In this context, dispersion is "good" because it reflects better resource allocation. In what follows, we lay out the key aspects of our framework and analyze the relationship between dispersion and misallocation.

Framework. Consider a firm that invests in N plants subject to two constraints. First, each plant provides only a limited amount of funds x to finance investment generated by, for example, past profits. Second, a fixed investment adjustment cost implies that τ units of funds are lost for every plant the firm invests in. Production in each plant is given by

$$y = \begin{cases} k^{\alpha} & \text{if no investment} \\ (k + i - \tau)^{\alpha} & \text{if investment} \end{cases} \tag{2}$$

where k is the existing capital stock in a plant that can be augmented by investment.[8] Returns to capital are positive and decreasing, which is reflected in $0 < \alpha < 1$. The firm uses the total funds available, Nx, for investment activity across its plants in order to maximize the sum of output subject to the fixed adjustment cost and the technology in Equation (2).

We now use this simple framework to illustrate the complex relationship between frictions, misallocation and dispersion. First, let us define our two concepts of dispersion:

Definition Dispersion in marginal revenue products, $V(mrpk)$, depends on the level of friction and is defined to be

- *good dispersion* if it decreases in the level of frictions,
- *bad dispersion* if it increases in the level of frictions.

Note that, while *good dispersion* and *bad dispersion* are local concepts, aggregate output is unambiguously decreasing in the level of frictions.

[7]Indeed, Bai et al. (2018) present a related finding: Reducing trade barriers in China has *worsened* misallocation.
[8]Though we assume the same k across all plants, our logic holds if we assume heterogeneous k.

Bad dispersion. Let us first focus on the role of the fixed investment adjustment cost. If $\tau = 0$, the firm effectively incurs no penalty from investing only small amounts in each plant. As a result, the optimal course of action is to invest equally across all plants to equate their marginal revenue products of capital. By definition, dispersion of *mrpk* is therefore nil. As the friction tightens (τ rises), the firm trades off the concavity of returns (pushing it to equalize investment across its plants) against the fixed adjustment cost (pushing it to concentrate on a few plants). The optimal action is to pick a share of plants, denoted by $n^* < 1$, in which the firm will invest equal amounts $i^* = x/n^*$. As a result, the variance of *mrpk* increases to

$$n^*(1-n^*)\left[\alpha \log(1+i^*/k)\right]^2 > 0, \tag{3}$$

and firm output is lower.

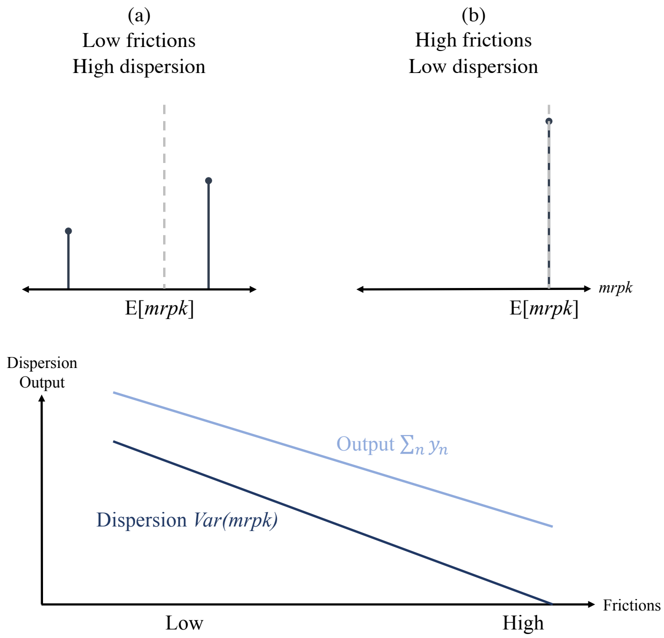
In sum, this example displays the standard relationship between dispersion and misallocation: As the friction is tightened, resource allocation moves further away from the unconstrained optimum and the dispersion of marginal revenue products of capital rises while output falls. In other words, more dispersion is *bad dispersion*. Note that it does not matter if τ represents a distortion as in Hsieh and Klenow (2009) or a fixed adjustment cost as in Asker et al. (2014).

Good dispersion. We now show that the interpretation of *mrpk* dispersion is, in reality, more complex and subtle than the impression given by the previous, standard argument. First, let us continue with the example of the fixed investment adjustment cost. Consider a starting level of τ so high that the firm decides to not invest at all, i.e., $n^* = 0$; the implication is that dispersion is again equal to zero, as was the case without any cost ($\tau = 0$). As τ is lowered, investing in a positive fraction of plants $n^* > 0$ becomes optimal. The outcome is a higher level of dispersion of *mrpk*, as can be seen from Equation (3), *as well as* more economic activity and output. In this case, as the friction is relaxed, resource allocation moves closer to the unconstrained optimum, and the dispersion of marginal revenue products of capital *mrpk* rises along with output. In other words, more dispersion is good.

Next, we turn our attention to the role of internal capital markets, holding $\tau > 0$ fixed. If frictions within the firm choke its ability to shift financial resources across its plants, effectively shutting down its internal capital market, only up to x funds can be used for investment in each plant. To simplify exposition, let's assume that $x < \tau < Nx$: The implication is that without the ability to pool funds, the firm cannot invest in any of its plants. Hence, without internal capital markets, the variance of (log) marginal revenue products is zero. This extreme case is depicted in Panel (b) of Figure 1.

What happens when we allow the firm to pool all Nx plant-level funds and reallocate them freely? Equipped with a functional internal capital market, the firm again trades off the concavity of returns (pushing it to equalize investment across its plants) against the fixed adjustment cost (pushing it to concentrate on a few plants). The optimal action is to pick a share of plants $n^* > 0$ in which the firm will invest equal amounts $i^* = x/n^*$. As a result, the variance of marginal revenue products increases to a positive value (see Equation (3)). The ex ante homogeneous plants, hit by the same shock, are now heterogeneous ex post *even though the allocation is more efficient*. This case is depicted in Panel (a) of Figure 1.

Figure 1: Frictions, output and dispersion



In this example, relaxing a friction and allowing for internal capital markets leads to a *higher* dispersion of marginal products of capital. In this sense, the nature of dispersion is again good.

More generally, we have shown that the relationship between frictions, misallocation and dispersion is non-monotonic.

Broader applications Even if our conceptual exercise were performed in an investment framework for a multi-plant firm, the underlying logic carries through in a broad range of settings. Hence, while the productive units in our example were plants, they could alternatively be business divisions, teams or even workers. The resources being pooled across productive units were financial funds, yet they could be interpreted as any firm-wide resource such as time, cognitive attention, technical knowledge or even managerial skills. Finally, the activity and its friction, in our case capital investment and the fixed adjustment cost, could instead be applied to a wide array of contexts:

- The introduction of new products subject to non-convex frictions, such as clinical trials for new drugs (see DiMasi et al. (2003)) or mandatory emissions regulations for new vehicles.
- Some innovation process that implies overhead costs of research and development as in Cohen and Klepper (1992), Cohen and Klepper (1996) and Aw et al. (2011) or fixed start-up costs for research labs.

- The decision to enter a new export market, subject to export rules and regulations that may render exporting small amounts of goods unprofitable (see Melitz (2003), Das et al. (2007) and Creusen et al. (2011)).
- Business restructuring that requires a minimal fixed amount of attention or time from managers (Caliendo et al. (2015)).
- Hiring activity subject to non-convex frictions τ such as costs related to job postings and interview procedures (Mortensen and Pissarides (1994)).

In sum, any tax, transaction cost, trade barrier, cost of doing business, or menu cost that is less than proportionally[9] related to an input will have the effect we described.

It should also be noted that while all of the examples above rely on some non-convex frictions, they are not a necessary ingredient for our result: the key to generating *good dispersion* is that the firm finds it optimal to focus the activity on a subset of units. For example, in our original multi-plant firm setting, it would suffice to assume locally increasing marginal returns to capital – say, up to a capacity constraint. In this case, eliminating internal capital market frictions would push the firm to pool resources across its plants and redistribute them toward a few units for investment purposes, generating more dispersion of $mrpk$ in the process. This means that the scope of our mechanism is broad, and so is the potential for *good dispersion*.

In the next section, we build a quantitative model of an economy in which firms operate several units and allocate capital across them. Capital is often pinned as the most distorted production factor (see Gopinath et al. (2017), among others) in explaining aggregate misallocation. We focus on multi-plant firms, as they account for the lion's share of economic activity in the U.S. economy. Importantly, that model features a rich set of frictions, as in Cooper and Haltiwanger (2006) and David and Venkateswaran (2019), which gives rise to the non-monotonic relationship between distortions and dispersion in marginal revenue products of capital.[10] Matching the rich model to establishment-level data from the ASM, we find that eliminating frictions increases the dispersion of marginal revenue products of capital.

3 A model of the multi-plant firm

In this section and the next, we describe, solve, simulate and analyze a simple model of a multi-plant firm. Our focus is on the role of within-firm frictions – more specifically, on those that regulate the functioning of internal capital markets – in shaping investment decisions and the dispersion of marginal revenue products of capital across plants within the firm. At one extreme, these frictions are so excessive that firms are merely a collection of disconnected productive units: Decisions are made on a plant-by-plant basis. At the other extreme, firm management can fully use its frictionless internal capital market to mitigate or offset the other frictions and constraints that it must cope

[9] Proportional distortions, as in Hsieh and Klenow (2009), can be added without changing our argument.

[10] Brown et al. (2016) make a similar point, in which the non-convex distortion affects the discrete entry decision.

with. In particular, the firm optimally alters the size and timing of plant-level investment projects, which in turn generates more dispersion across its plants.

3.1 The problem of the firm

We focus on the basic problem of a firm that operates two plants n, where $n = A, B$. We limit our model to only two plants in an effort to keep the numerical analysis of our model, which features non-differentiable investment policies, computationally feasible. A larger number of plants would exponentially increase the size of the state vector of the firm, which must include the capital stock and technology level of each of its plants, without adding insight into the underlying fundamental economic mechanisms. We start by describing the technology and constraints at the level of the individual plant before analyzing the problem of the firm. In what follows, lowercase letters refer to plant variables, uppercase letters to firm variables and bold uppercase letters to vectors of a firm's plant variables.

3.2 Technology and frictions at the plant level

The plant operates a Cobb-Douglas production function that combines the beginning-of-period capital stock, k_{nt}, and other variable inputs in order to produce output, y_{nt}. While capital is fixed throughout the period, we assume that plants can freely choose any other variable inputs in perfectly competitive markets.[11] This means we can substitute out any static first-order condition for variable inputs and write revenue net of variable factor costs for plant n as

$$y_{nt} = e^{z_{nt}} k_{nt}^{\alpha}. \tag{4}$$

z_{nt} contains plant (log) total factor productivity and prices of other statically chosen production factors, while α is the scaled production elasticity of capital. The productivity level of plant n in firm j consists of a component common to both plants in the firm and an idiosyncratic plant component; both evolve as follows:

$$z_{njt} = \rho_p z_{njt-1} + \eta_{njt} \tag{5}$$

$$z_{jt} = \rho_f z_{jt-1} + \eta_{jt}. \tag{6}$$

where η_{njt} and η_{jt} are both iid, mean zero and have variances σ_p^2 and σ_f^2, respectively.

The capital stock of plant n depreciates every period at rate δ and grows with investment i_{nt}, so it evolves over time according to the conventional expression

$$k_{nt+1} = (1 - \delta)k_{nt} + i_{nt}.$$

[11] Given our Cobb-Douglas production function, flexible factor markets will result in revenue products of flexible inputs that are completely equalized across plants and firms in the economy. This will not be the case, however, for marginal revenue products of capital because capital is chosen one period in advance and because of decreasing returns to scale as well as fixed investment adjustment cost.

As documented in a number of studies (see Cooper and Haltiwanger (1993), Cooper et al. (1999), Doms and Dunne (1998) and Caballero and Engel (1999), among others), investment dynamics at the plant level are characterized by lumpiness: Multiple periods of inactivity (no or only small amounts of maintenance investment) are followed by "investment spikes."[12] The traditional modeling feature used to reproduce this stylized fact is to introduce a fixed cost of investing: The firm must pay a certain cost, ψk_{nt}, if investment is greater than zero. Such costs can arise because investment activity – no matter how small or large – has a disruptive effect on production activities in the short run, for example. The parameter ψ regulates how much revenue is forgone when the plants needs to shut down production in order to install new capital. As a result of aggregation, firm-level investment activity will be less lumpy, as documented by Eberly et al. (2012).

In addition to this non-convex adjustment cost, we include a traditional quadratic adjustment cost. This convex adjustment cost captures the notion that larger investment projects become increasingly disruptive with size.[13] The parameter γ below captures the importance of this margin.

To summarize, frictions at the plant level are expressed as:

$$\theta\left(i_{nt}, k_{nt}\right) = \left[\psi \mathbb{I}\left\{\frac{i_{nt}}{k_{nt}} > \vartheta\right\} + \frac{\gamma}{2}\left(\frac{i_{nt}}{k_{nt}}\right)^2\right] k_{nt} \tag{7}$$

where $\mathbb{I}$ is an indicator function equal to 1 if the plant investment rate is above ϑ; ψ is a parameter regulating the forgone sales if the plant undergoes an investment; and γ regulates the impact of the quadratic adjustment cost. Everything is scaled by the plant's capital stock k_{nt} in order to eliminate size differences.

Combining equations (4) and (7) above, plant cash flow is given by

$$\pi_{nt} = z_{nt} k_{nt}^{\alpha} - \theta(i_{nt}, k_{nt}). \tag{8}$$

3.3 Technology and frictions at the firm level

What sets plants in a multi-unit firm apart from their identical counterparts in single-unit firms? What are the economic benefits the firm provides to its own plants? While these benefits are likely numerous, our focus is on the ability of firms to create internal capital markets by pooling and reallocating internal funds across its plants. This ability allows the firm to relax its external financing constraint.

A firm with frictionless internal capital markets collects the cash flow from all of its plants and decides how to allocate funds to finance investment projects across its plants. This means that firm cash flow is defined as

$$\Pi_t = \pi_{At} + \pi_{Bt}. \tag{9}$$

[12] Investment spikes are usually defined as investment rates exceeding 15% or 20%.

[13] This formulation is similar to assuming lower profitability during large capital adjustments, which has been documented by Power (1998) and Sakellaris (2004).

While all production and investment activities take place at the level of the individual plant, we assume that only the firm can organize external finance. In a dynamic model of investment, this seems the most sensible choice, but recall from the end of Section 2.2 that many other mechanisms would serve the same purpose and lead to similar results.

Our assumption that only the firm can organize external finance is realistic and sensible: While large and complex firms such as General Electric operate hundreds of plants, only the firm issues bonds, borrows from banks or raises equity. Typically, the firm then channels these funds to individual plants through its internal capital market. Consistent with this empirical pattern, we assume that it is the firm that coordinates investment plans across all of its plants, organizes financing of investment through either internal cash flow or external finance, and allocates funds to plants where investment is put in place. In the event that desired firm-wide investment exceeds firm cash flow, the firm attempts to raise external funds, denoted by E_t, so that all investment gets financed:

$$i_{At} + i_{Bt} = I_t \leq \Pi_t + E_t. \tag{10}$$

Organizing external finance, however, is an imperfect process. We assume that new equity issuances, E_t, are limited: Firm owners do not tolerate negative dividends beyond a certain fraction λ of the capital stock. This is consistent with the notion that it becomes increasingly costly to issue larger and larger amounts of equity, as is common in the finance literature. Hennessy and Whited (2007) and Altınkılıç and Hansen (2000) estimate the cost of raising external equity to be increasing and convex. Hennessy and Whited (2007) additionally find that these cost are significantly higher for small firms. We will capture this finding by making the financial constraint more binding for small firms.

One may also interpret our external finance constraint as applying to external debt. Such borrowing can be limited in various ways by moral hazard problems, as in Gertler and Kiyotaki (2010). Consider a borrowing firm that can divert a fraction $1/\eta$ of the loan for private benefit. Lenders will then require collateral, often a fraction ξ of the firm's capital stock, ξK, which they could seize in case of bankruptcy. Then, the divertable loan amount can never exceed the collateral. In this case, $\lambda = \eta\xi$ equals the maximum leverage the lender is willing to accept. Therefore, our argument continues to hold if external finance is debt instead of equity.[14] Ultimately, what is crucial is that the firm is constrained in its access to external funds.

We summarize the external financial constraint by the following function:

$$E_t \leq \lambda K_t. \tag{11}$$

The cost of investing in a given plant depends on the investment amount in that plant, the combined investment in the rest of the firm, whether the firm needs to raise external funds, and if it could be limited the financing constraint. The total cost of investment in Plant A then consists

[14] For computational reasons, we do not allow for savings by the firm except through the accumulation of capital. While this would represent an interesting extension, we argue there is no reason to believe that it would meaningfully alter our conclusions below.

of the fixed and quadratic adjustment costs (real costs $\theta(i_{At}/k_{At})$ in Equation (7)) as well as the financial constraint in Equation (11)). The latter part depends on how much the other plant in the firm, Plant B, invests, as this dictates how fast and how much the firm needs to borrow. Thus, investment in one plant imposes an externality on investment activity in the rest of the firm, because it depletes internal funds and imposes a financial cost that is shared by the entire firm.

3.4 Firm value and firm policy

We define the vectors of technology levels and capital stocks within the firm as $\mathbf{Z_t} = \{z_{At}, z_{Bt}\}$ and $\mathbf{K_t} = \{k_{At}, k_{Bt}\}$, respectively. Given the plant-level fixed adjustment cost, the firm's state consists of the distribution of capital stocks, $\mathbf{K_t}$, and technology levels, $\mathbf{Z_t}$, across plants within the firm. The firm chooses investment in either Plant A or B in order to maximize firm value, which corresponds to the net present value of discounted future gross profits net of investment and borrowing costs. When deciding the investment level of each plant, the firm takes into account the various adjustment costs and whether borrowing is required to finance the desired level of investment. The firm's problem can be written in recursive form as

$$
\begin{aligned}
V^{\text{ICM}}(\mathbf{Z_t}, \mathbf{K_t}) = \max_{i_{At}, i_{Bt}, E_t} & \left\{\Pi_t - I_t + \beta \mathbb{E} V(\mathbf{Z_{t+1}}, \mathbf{K_{t+1}})\right\} \\
s.t. \quad & k'_{nt} = (1-\delta)\, k_{nt} + i_{nt} \qquad \forall n = A, B \\
& E_t \leq \lambda K_t. \\
& I_t \leq \Pi_t + E_t.
\end{aligned}
$$

In the above problem, the superscript "ICM" on the value of the firm indicates that it can fully leverage its internal capital market. This has two important advantages. First, the firm can pool the cash flows generated by its two plants in order to finance investment projects where it sees fit. Second, it can combine plant-level capital stocks to increase its capacity to access external funds.

In the quantitative analysis that comes next, we will contrast this setup with one in which the firm is unable to pool plant-level resources. This could be due to a number of frictions and constraints. Stein (1997) and Scharfstein and Stein (2000), for example, emphasize how individual divisions within a firm compete for corporate resources in order to build "local empires," while Giroud (2013) has quantified the impact of imperfect information flow within a firm on the investment efficiency at the level of individual plants. When such frictions within a firm become extreme, internal capital markets cease to function, and the firm effectively operates its plants as standalone units: Maximizing firm-level profits then boils down to separately maximizing the value of each plant in isolation. Equations (10) and (11) now apply for each plant individually, and the value function of the firm becomes

$$
\begin{aligned}
V^{\text{No ICM}}(\mathbf{Z_t}, \mathbf{K_t}) = \max_{i_{At}, i_{Bt}, e_{At}, e_{Bt}} & \left\{ \Pi_t - I_t + \beta \mathbb{E} V(\mathbf{Z_{t+1}}, \mathbf{K_{t+1}}) \right\} \\
s.t. \quad & k'_{nt} = (1-\delta)\, k_{nt} + i_{nt} && \forall n = A, B \\
& e_{nt} \leq \lambda k_{nt} && \forall n = A, B \\
& i_{nt} \leq \pi_{nt} + e_{nt} && \forall n = A, B.
\end{aligned}
$$

In the next section, we study quantitatively the consequences of such a shutdown in internal capital markets.

4 Quantitative analysis

In what follows, we first perform a numerical analysis of the model of the previous section to illustrate quantitatively our main point: that eliminating a friction – in this case, the constraints to leveraging internal capital markets – can lead to both a more efficient allocation of resources *and* a rise in the dispersion of marginal products. Second, we provide empirical evidence that supports the main model mechanism.

4.1 Calibration

Table 3 summarizes the parameter values used to calibrate our model for the quantitative analysis. Most values are based on moments from the ASM dataset and are in line with calibrated parameters generally used in the investment literature.

Table 3: Model Calibration

Parameter	Meaning	Value	Target/Source
β	Discount rate	0.95	Long-run real interest rate
α	Production elasticity	0.627	estimated in ASM
ρ^p	TFP persistence plant	0.60	serial correlation of $mrpk^p$: 0.25 in ASM
ρ^f	TFP persistence firm	0.85	serial correlation of $mrpk^f$: 0.31 in ASM
σ^p	TFP shock plant	0.25	volatility of $mrpk^p$: 0.33 in ASM
σ^f	TFP shock firm	0.24	volatility of $mrpk^f$: 0.26 in ASM
δ	Depreciation rate	0.067	Mean investment rate in ASM
ψ	Fixed inv. adj. cost	0.039	Cooper and Haltiwanger (2006)
γ	Quadratic inv. adj. cost	0.049	Cooper and Haltiwanger (2006)
λ	External finance capacity	0.30	Li et al. (2016)

To inform us about the production function elasticity, α, we extend the structural framework of Cooper and Haltiwanger (2006) to accommodate multi-plant firms and re-estimate plant-level revenue functions. Our GMM estimate puts α at 0.627, which is fairly close to the value they find. The parameters governing persistence, ρ^p and ρ^f, and volatility, σ^p and σ^f, of the plant and firm shock processes are chosen to match the persistence and volatility of $mrpk$ at the plant and firm levels in two-plant firms in the ASM.

The depreciation rate δ is set to match the long-run investment rate in our ASM data. For the fixed and convex investment adjustment cost parameters, ψ and γ, we rely on the structural estimates of Cooper and Haltiwanger (2006), which are somewhat smaller than the analogous values estimated by Asker et al. (2014). Regarding the capacity for external finance, our benchmark case relies on $\lambda = 0.3$, while we will be experimenting with other values as well to illustrate the key driver of within-firm capital allocation. This value is in line with the evidence of Li et al. (2016), who estimate a dynamic model of the firm with a similar external financing constraint. Their estimates put λ between 0.22 and 0.32 across manufacturing industries.[15]

4.2 Frictions, allocation and *good dispersion*

Before analyzing the forces at play inside the model, we go straight to the main question of the paper: Does the model generate a quantitatively relevant increase in *both* dispersion *and* output once the firm is allowed to pool the cash flows and capacities for external finance (capital stock) from its two plants? Can relaxing a friction be welfare-improving, yet at the same time generate *more* dispersion in marginal revenue products?

To answer these questions, we compare quantitatively two distinct economies. The first one is composed of perfectly integrated two-plant firms, as described earlier: A firm can pool the cash flows and capacity for external finance from its two plants, and hence leverage its internal capital market (ICM). This is what we refer to as the "ICM" economy.

The second economy is composed of firms that face internal frictions and constraints that preclude their ability to pool resources (funds and external finance capacity) across their plants and take advantage of internal capital markets. In other words, "firms" are empty concepts in that economy, as plants effectively function as standalone units that are not part of an integrated firm. Consequently, a plant that is in need of funds to finance its investment activity must effectively obtain external financing on its own. Otherwise, it faces the exact same constraints and frictions as the plants in multi-unit firms. This environment is referred to as the "No ICM" economy.

In essence, the difference between the ICM and No ICM economies is that in the latter, we create a "wall" between plants that bars them from pooling financial resources when access to external financial markets is constrained. We allow plants to raise funds, but in doing so they are limited by their own capital stock. This exercise allows us to specifically isolate the role of the firm in creating internal capital markets.

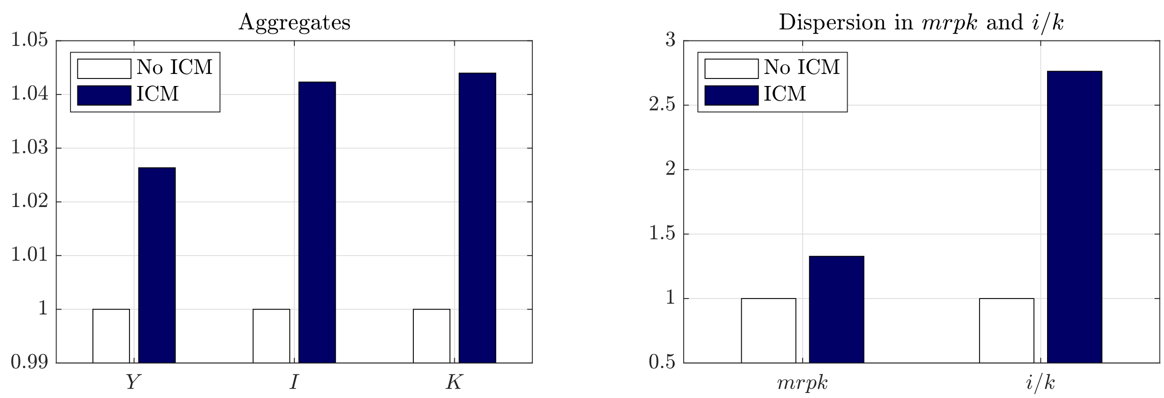
In each case, the problem of a representative firm is solved using a value function iteration procedure that is described in detail in Appendix C. Then, we simulate an economy composed of 1,000 two-plant firms for 1,000 periods to investigate two questions. First, do two-plant firms with

[15] In the case of Li et al. (2016), it is the non-depreciated value of the capital stock that is collateralizable. Also, while Rampini and Viswanathan (2013) and D'Acunto et al. (2016) also provide information on leverage, their approach is not model-based, which makes it difficult to transpose their findings into our framework.

internal capital markets produce more dispersion in marginal revenue products than firms without any internal ability to pool resources? Second, how much more aggregate output can the economy with internal capital markets produce? Naturally, two-plant firms with internal capital markets cannot do worse, since they convexify the choice set and can always reproduce the allocation of the single-plant firm economy. Our objective is to determine how large these gains can be.

Figure 2 provides answers to these questions. It displays dispersion measures and aggregate values for both economies, in which we normalize the values for the economy without internal capital markets to unity to facilitate the presentation and comparisons.

Figure 2: Quantitative effects in the multi-plant-firm vs. single-plant-firm economy



Note: The figure displays dispersion and aggregate values in an economy composed of firms that cannot pool resources across plants (No ICM, in white), as well as in an economy with internal capital markets (ICM, in black). The No ICM values are normalized to 1 to ease the presentation and comparisons. The left panel shows aggregate output, investment and capital. The right panel displays the dispersion of marginal revenue products of capital and investment rates.

The left panel shows the aggregate values for output, investment, and capital in both economies. As expected, allowing internal capital markets to play their role leads to higher aggregate values: Once firms are able to pool their plants' financial resources and capital stocks, aggregate capital and investment are about 4% higher, while output rises by just under 3%. Hence, not surprisingly, the elimination of this friction is welfare improving.

The findings in the right panel are more surprising. The first set of bars indicates that this extra aggregate output was generated by the ICM economy *despite* the fact that dispersion of the marginal revenue product of capital is 32% higher than when internal capital markets are not available. The difference is even more striking for investment: The economy-wide dispersion of investment rates is almost three times larger in the ICM version of the simulation.

In sum, we showed that a more efficient allocation can be accompanied by a *higher* dispersion of marginal revenue products. The difference between the blue and white bars in the right panel of Figure 2 thus reflects *good dispersion.*

4.3 Why do firms create more dispersion? The mechanism

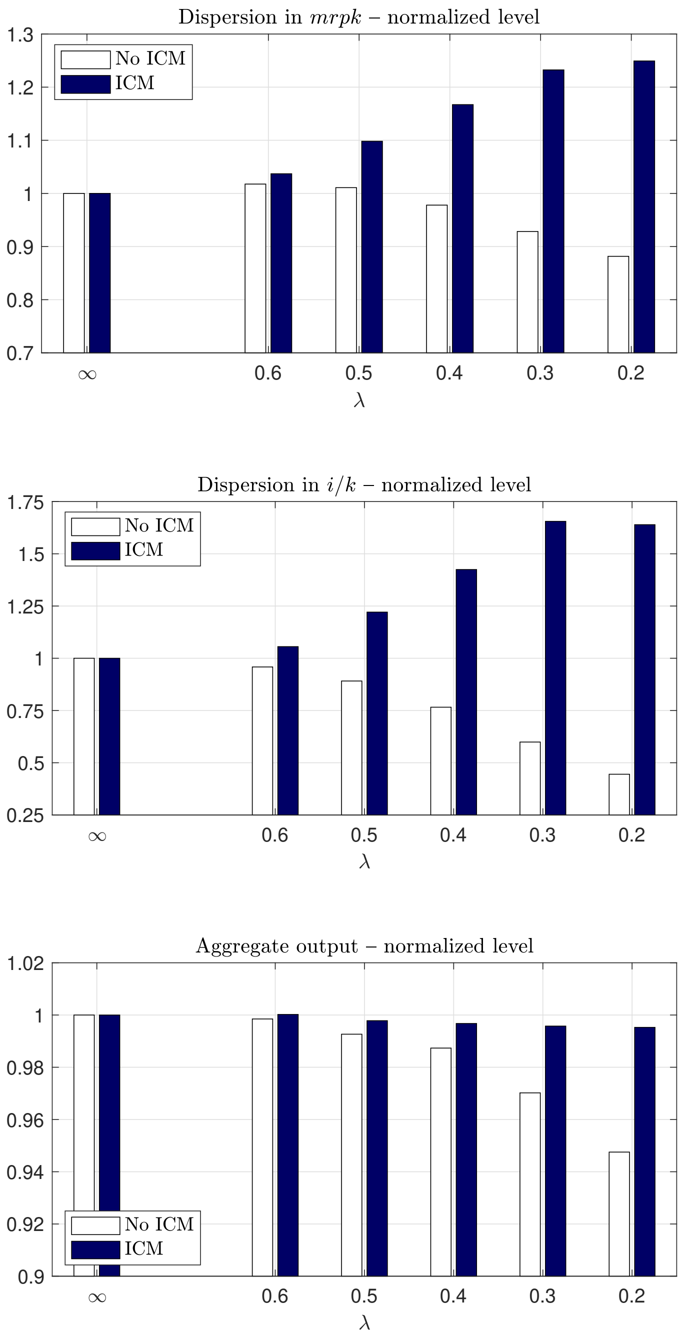
Next, we analyze the predictions of our quantitative model to gain better intuition about the *joint dynamics* of marginal revenue products of capital and the allocation of capital within the firm. Investigating these micro-level dynamics is helpful for understanding the mechanisms at the heart of the model. Our focus is on the interaction between the firm-level external financing constraint and the investment fixed cost, two frictions that are at the heart of the relationship between dispersion and allocation that we found in Section 4.2.

Internal capital markets and *good dispersion*. Figure 3 shows three moments we saw earlier, but this time for different values of the financial friction parameter λ. This parameter governs the value of internal capital markets to firms. When $\lambda = \infty$, the firm's ability to pool internal funds is worthless, as each plant can raise unlimited external finance on its own. Only when that external finance becomes limited ($\lambda < \infty$) does a firm's internal capital market become valuable, and the difference between the economies with and without internal capital markets plays out. In each panel, we plot the results for economies with (ICM) and without (No ICM) internal capital markets. All moments are normalized to 1 in the scenario in which the external financing constraint is not binding ($\lambda = \infty$). Unsurprisingly, the economies with and without internal capital markets coincide when $\lambda = \infty$ and firms do not matter.

The top and middle panels confirm our finding from the previous section: For a given value of λ, economy-wide dispersion of both the expected marginal revenue products of capital, $E_t[mrpk_{nt}]$, and the investment rate, i_{nt}/k_{nt}, is higher in the least-constrained environment that allows for internal capital markets (ICM) than when plants are constrained to function as standalone units (No ICM). This higher dispersion is accompanied by a higher level of output, as can be seen in the bottom panel of Figure 3.

Financial frictions and *good dispersion*. The simulation results of Figure 3 also indicate that the ICM friction is not the only margin along which the standard negative relationship between dispersion and output is reversed. As we tighten the external financing constraint by lowering the value of λ, the bottom panel indicates that aggregate output falls considerably in both economies. As expected, the drop is much more dramatic when internal capital markets are not functional: If the typical firm cannot pool cash flows from its productive units, investment in capital and production suffer heavily as its access to external funds becomes more limited. Yet the top panel shows that the dispersion of *mrpk falls* for this No ICM economy, despite the more stringent financial friction. In other words, there is a second dimension along which we can observe *good dispersion*: When the financial constraint is less restrictive (higher value of λ, toward the left in the figure), dispersion increases, even though the effect is not as dramatic as along the ICM friction dimension (dark blue vs. white bars).

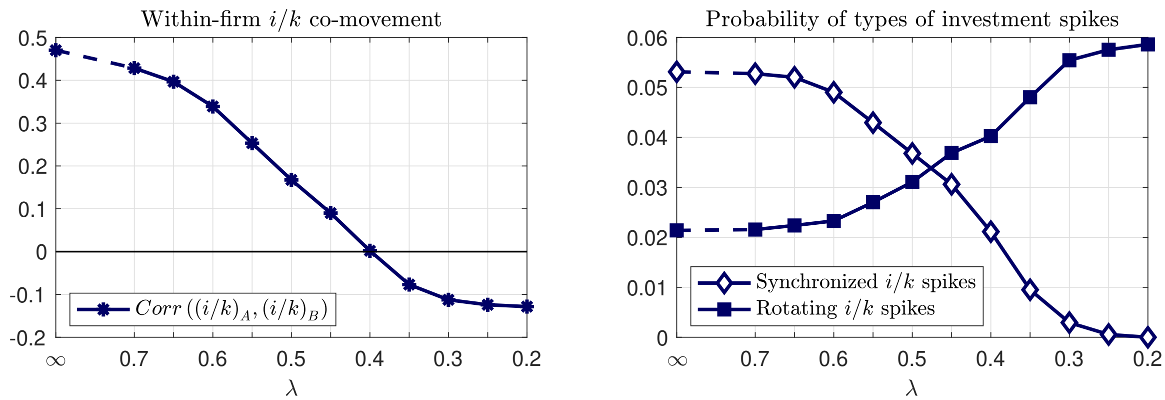
Figure 3: Quantitative effects of financial constraints on the multi-plant firm economy



Note: This figure plots various dispersion moments and aggregate output as a function of the capacity for external finance λ in economies with internal capital markets (ICM) and without (No ICM).

Within-firm investment dynamics. We obtain these non-standard relationships between dispersion and misallocation despite the fact that our model is quite standard. What is the mechanism that lies behind this result? The left panel of Figure 4 provides a hint: for the economy with internal capital markets, changes in the financing friction alter within-firm investment dynamics significantly. In this economy, we see that as the financing constraint increasingly limits the ability of the firm to borrow (i.e., λ falls), the correlation in investment activity across plants within the firm drops dramatically from its benchmark value of $+0.48$ in the unconstrained case. In fact, for low enough values of λ, the cross-plant investment correlation turns *negative*, reaching a trough of -0.12 with our parameterization. Given that plant-level TFP has a common firm component, z_{jt}, one would expect investment to comove positively as well. What force is driving plant-level investment apart?

Figure 4: Quantitative effects of financial constraints on the multi-plant firm economy



Note: This figure plots the correlation of investment rates across plants within the typical firm for various values of the external financing constraint λ, as well as the probabilities of observing rotating and synchronized investment spikes. All results are for the economy with internal capital markets (ICM).

The right panel of Figure 4 illustrates the mechanism that leads to this drop in correlation and rise in dispersion. Here, we plot two additional moments. The first represents the probability of observing synchronized investment spikes within the firm, i.e., investment rates above 15% in both plants at the same time. As the external financing constraint is tightened, the firm cannot allocate large amounts of new capital to both plants within the same period, even following a positive firm-level shock that increases internal funds. Conceivably, an option for the firm would be to invest smaller amounts in both plants to restrain how much it needs to borrow on capital markets. In the presence of investment fixed costs, however, the firm opts instead to alternate investment activity across its plants, as it is unable to gather the internal and external funds to finance investment projects in both units simultaneously. For example, in the event of a large positive firm-level shock, the firm allocates capital first to the plant with the highest (expected) return on capital, before doing the same for its second plant in the following period. Conversely, the frequency of "rotating investment spikes," i.e., an investment spike in one plant followed by a spike in the other plant the next period, rises as external financing becomes more limited. This process of staggering investment spikes over time, in turn, creates higher within-firm and economy-wide dispersion of not only investment rates but also *mrpk*. In an economy in which firms and plants are constantly buffeted by shocks, this effect occurs frequently: We find a very similar pattern if we instead compute the dispersion of 5-year moving averages of logged expected *mrpk*.

Conversely, in an economy without internal capital markets, the firm cannot pool resources to implement a staggered-investment strategy. As an illustration, consider that both plants are hit by the same firm- and plant-level positive shocks. In the face of these shocks, the plants, effectively functioning as standalone units with their own capacity for external finance and internal funds, will find it optimal to either invest or not. In both cases, there is no gain from desynchronizing investment activity across plants, and as a result dispersion is lower than in the economy with internal capital markets.

In sum, the presence of an external financial constraint pushes the multi-plant firm to leverage its internal capital market. Doing so requires rotating investment spikes across its plants.[16] This process not only leads to investment activity across individual plants within a firm that looks staggered and less correlated, but also creates higher within-firm and economy-wide dispersion in *mrpk* and investment rates.

Again, it is worth emphasizing that *this increased dispersion is the optimal response of the firm to the constraints it faces rather than a consequence of inefficient distortions.* The presence of such *good dispersion* is due to the interaction of the internal capital markets with two crucial frictions: the fixed investment adjustment cost ψ and the capacity for external finance λ. For example, in the absence of any external financing constraint, the firm would have unlimited access to financial resources and would invest to equalize *mrpk* across its plants. In other words, absent distortions à la Hsieh and Klenow (2009), any dispersion in *mrpk* would be detrimental and merely reflect the presence of the technological fixed investment adjustment cost ψ, as in Asker et al. (2014).[17] Conversely, without the presence of fixed adjustment costs, a financially constrained firm would invest suboptimally low amounts, but would still allocate investment to equalize marginal revenue products across its plants.

Multi-plant firms and investment flow adjustment costs. Finally, it is worth briefly discussing the link between the mechanism at the core of our multi-plant firm model and the investment flow adjustment cost of the form $\kappa(I_t/I_{t-1})^2$ that is commonly used in DSGE macro models. While this reduced-form friction has been found to fit the aggregate dynamics of investment much better than other types of adjustment costs (see Christiano et al. (2005)), the literature offers little in the way of micro-foundations to rationalize its use (an exception is Lucca (2007)). Our model of a multi-plant firm provides a natural interpretation: The optimal capital allocation it generates can

[16] Besley et al. (1993) study Rotating Savings and Credit Associations (ROSCA), a financial institution mostly found in developing economies that relies on a similar mechanism.

[17] Similar results could be obtained in the case of a constraint that would see the interest rate paid by the firm increase with the size of its borrowing. Under such conditions, the firm would again wish to stagger investment projects in order to minimize the cost of borrowing.

reconcile the lumpy investment dynamics at the plant level (Cooper and Haltiwanger (2006)) with the finding of Eberly et al. (2012) that investment for larger firms is smooth and compatible with an investment flow adjustment cost.[18]

4.4 Supporting evidence for the mechanism

We now turn our attention to the data and provide empirical support for the mechanisms that are central to the model. As in Section 2.1, our source is the Annual Survey of Manufactures (ASM), an annual Census dataset covering manufacturing plants and described in detail in Appendix A.2. To remain as close as possible to the model environment, we focus in this exercise on plants belonging to firms that operate exactly two plants.

As we saw in Figure 4, central to the model predictions is the effect of the external financing constraint on within-firm investment dynamics and, in turn, within-firm dispersion. We now aim at investigating whether the pattern in the right panel of Figure 4 is borne out in the data. We study how a firm responds to a firm-wide technology shock, defined analogous to Equation (6), that boosts both plants' productivity. In the previous section, we showed that such a firm-wide technology shock should lead to synchronized investment spikes if the firm has a large capacity for external finance and a functional internal capital market. On the other hand, the model predicts that a firm constrained in external financing but with a functioning internal capital market will react to the firm-wide shock by staggering investment spikes in order to circumvent the financial friction. Due to the lack of financial data in the Census, we proxy a firm's capacity for external finance – and therefore the tightness of the financial constraint – by its aggregate capital stock, as was the case in the model.

We first estimate the likelihood of synchronized investment spikes after a firm-wide technology shock using the following probit model:

$$Pr(Y_{jt}^{sync} = 1 | X_{jt}) = \Phi\left(X_{jt}\beta\right) \tag{12}$$

where X_{jt} is a vector of controls that includes the firm's level of the productivity shock, η_{jt}. The dummy variable Y_{jt}^{sync} is defined as follows:

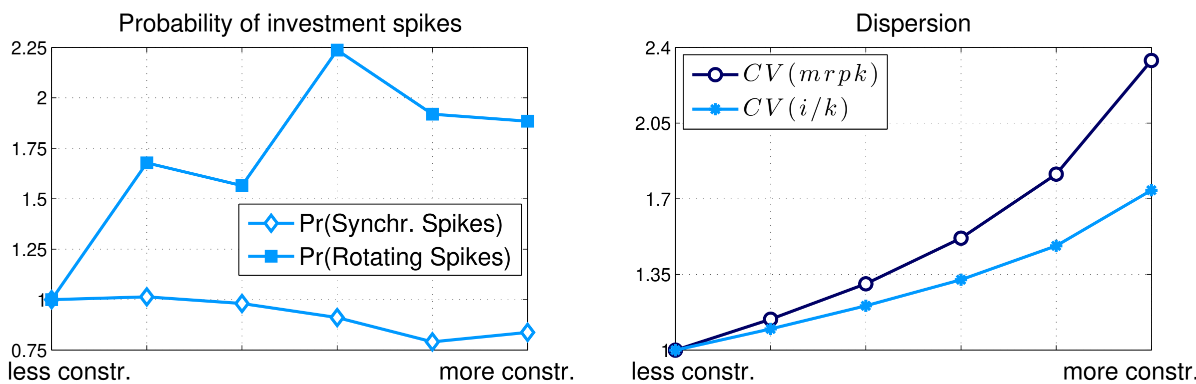
$$Y_{jt}^{sync} = \begin{cases} 1 & \text{if } \frac{i_{At}}{k_{At}} > 0.15 \text{ and } \frac{i_{Bt}}{k_{Bt}} > 0.15 \\ 0 & \text{otherwise.} \end{cases}$$

We focus on the firm's investment response to a firm shock, because this provides the cleanest example of the changing investment patterns as the financial constraint tightens. We estimate Equation (12) separately for 10 size deciles of two-plant firms, where size is defined as the value of the capital stock. We then evaluate the marginal likelihood of a synchronized investment spike in the

[18]In our model, investment at the firm level is smoother than at the plant level. Hence, owners of multi-plant firms will not be exposed to volatile consumption, thus alleviating the concern of Thomas (2002) that general equilibrium forces would render plant-level investment spikes irrelevant for aggregate dynamics.

wake of a firm-wide technology shock, i.e., a shock experienced by both plants. To ease disclosure requirements, we smooth the results of these deciles using a cross-sectional rolling-window average estimate of the five adjacent deciles. Normalizing the probability of a "synchronized investment spike" to unity for the least financially constrained firms, we plot the normalized probabilities in the left panel of Figure 5 (line with hollow diamonds). It shows that the most financially constrained firms are only 80% as likely as the least constrained firms to respond to a firm-wide technology shock with an investment spike in both plants.

Figure 5: Dispersion and the nature of investment spikes after firm TFP shocks



Note: Left panel: Normalized probabilities of "synchronized investment spikes" and "rotating investment spikes" after firm technology shocks by the size of the capital stock. Right panel: Dispersion of marginal revenue products of capital and investment rates by the size of the capital stock.

In a similar vein, we estimate probit models in which we regress a dummy variable equal to one if the firm experiences a "rotating investment spike," that is, if both plants feature an investment spike in the wake of a firm productivity shock, but in subsequent periods instead of concurrently. Specifically, we estimate

$$Pr(Y_{jt}^{rotate} = 1 | X_{jt}) = \Phi\left(X_{jt}\beta\right), \tag{13}$$

where X_{jt} is a vector of controls that includes the firm's level of the productivity shock, η_{jt}, and the dummy variable Y_{jt}^{rotate} is defined as follows:

$$Y_{jt}^{rotate} = \begin{cases} 1 & \text{if } \frac{i_{At}}{k_{At}} > 0.15 \text{ and } \frac{i_{Bt}}{k_{Bt}} < 0.15 \text{ and } \frac{i_{At+1}}{k_{At+1}} < 0.15 \text{ and } \frac{i_{Bt+1}}{k_{Bt+1}} > 0.15 \\ 1 & \text{if } \frac{i_{At}}{k_{At}} < 0.15 \text{ and } \frac{i_{Bt}}{k_{Bt}} > 0.15 \text{ and } \frac{i_{At+1}}{k_{At+1}} > 0.15 \text{ and } \frac{i_{Bt+1}}{k_{Bt+1}} < 0.15 \\ 0 & \text{otherwise.} \end{cases}$$

Again, we evaluate the marginal probabilities, smooth them across deciles and normalize the probability in the least constrained group to unity. The results are plotted in the left panel of Figure 5 (line with solid squares). In line with model predictions, the firms that are the most constrained in external financial markets are more likely to respond to firm shocks by rotating investment spikes. That is, the typical firm contemporaneously invests in only one plant, *then* invests in the other plant in the subsequent period. The relationship is economically significant: The most constrained firms are twice as likely to experience a rotating investment spike as financially unconstrained firms.[19] All in all, we view this evidence as strong support for the relevance of the key model mechanism. The empirical investment patterns documented in the left panel of Figure 5 mirror those produced by the model and displayed in the right panel of Figure 4.

In the top and middle panels of Figure 3, we saw that the within-firm dispersion of the marginal revenue product of capital and investment rate for ICM firms is generally rising as access to external funds becomes more limited. This *bad dispersion* came about because firms with a large capital stock should have an easier time equating marginal products across plants instead of having to rely heavily on internal capital markets by alternating capital expenditures. We compute the two dispersion measures by decile of capital stock and divide by the mean level in each decile to account for level differences. Unlike the standard deviation, the resulting coefficient of variation is dimensionless and can be easily compared across deciles. Again, we smooth out the results across deciles and plot the results in the right panel of Figure 5.

As shown in the right panel of Figure 5, we find empirical support for this prediction. First, investment dispersion monotonically increases in our proxy for external financing constraints: The most constrained firms have investment dispersion within firms that is about 1.7 times larger than that for the least constrained firms. The difference in marginal revenue products of capital is even stronger: The most constrained firms are about 2.5 times as dispersed as the least constrained firms.

To summarize, we showed in earlier sections that a higher dispersion of marginal revenue products of capital can be an outcome of constrained efficient behavior by multi-plant firms, instead of a sign of resource misallocation. In other words, if frictions do not impede its ability to pool resources across plants, the firm leverages its internal capital markets in order to circumvent external financial constraints. It does so by staggering investment activity, in the process optimally creating additional dispersion of marginal revenue products across its plants. We find evidence of this mechanism in the data.

5 Revisiting Hsieh/Klenow with internal capital markets

Next, we investigate the potential implications of *good dispersion* for quantifying aggregate losses from misallocation in emerging economies. In their seminal work, Hsieh and Klenow (2009) estimate that lowering distortions in the Chinese and Indian manufacturing sectors to match dispersion in revenue total factor productivity found in the U.S. would lead to aggregate TFP gains of 39% for China and 47% for India.[20] While these are already sizable numbers, our findings suggest that they may in fact be lower bounds on potential gains.

[19] This empirical pattern of plants in financially constrained U.S. manufacturing firms is more nuanced than what Midrigan and Xu (2014) find. This is likely due to the fact that data limitations force them to focus on short panels. It is therefore possible that these temporal patterns in investment spikes are confounded with plant fixed effects.

[20] To compute this number, we take the average gain across the three years analyzed by Hsieh and Klenow (2009). See their Table VI for details.

The reason lies in differences in the organizational complexity of firms in developed economies such as the United States relative to emerging economies such as China and India. The vast majority of economic activity in the former group is accounted for by multi-plant firms (see Table A1). As we showed with the help of our model, internal capital markets in these multi-plant firms generate *good dispersion*, which we quantified as about a quarter of total dispersion for U.S. manufacturing. Developing economies, on the other hand, are largely populated by single-plant firms.[21] With a limited role for internal capital markets, dispersion in marginal revenue products is predominantly caused by distortions or other frictions and thus reflect *bad dispersion*. Therefore, the differences in marginal revenue product dispersion due to distortions (*bad dispersion*) between developed and emerging economies should in fact be even greater than those found by Hsieh and Klenow (2009).

To summarize, the potential gains from eliminating misallocation through resource reallocation are twofold: First, reducing distortions increases output as shown by Hsieh and Klenow (2009); and second, introducing multi-plant firms with internal capital markets also renders the economy more efficient, even if it generates more dispersion. This second efficiency gain has been overlooked so far. In the preliminary exercise that follows, we show that its magnitude is likely not trivial.

Hsieh and Klenow (2009) use differences in $TFPR$ dispersion to infer potential aggregate TFP and output gains from eliminating inefficiencies. To maintain consistency with that framework, we adopt their main assumptions of constant returns to scale, monopolistic competition, and a joint log-normal distribution of physical and revenue total factor productivity, denoted by $TFPQ$ and $TFPR$, respectively. Although our model was silent on the potential causes for dispersion in marginal revenue products of labor, we assume further that the variance of $mrpl$ in the two-plant-firm economy relative to the one-plant-firm economy behaves similarly to that of $mrpk$, and that its covariance with $mrpk$ is unchanged.[22] Under these assumptions, aggregate (or sectoral) TFP can be decomposed into an efficiency and dispersion term, as in Equation (16) of Hsieh and Klenow (2009). Absent further sectoral information on the two terms and the capital share (α in their notation), as well as the sectoral weights in aggregate production (θ_s in their notation), we abstract from differences in these terms and parameters that are specific to the country and sector. Using their Equation (16), we can then write the conventionally computed output gain from reducing misallocation in emerging economies (EM) down to the level of the U.S. economy (US) as

$$\log(1\text{+conv. output gain}) = \frac{1}{\sigma-1}\log\left(\frac{\sum_i A_{n,US}^{\sigma-1}}{\sum_n A_{n,EM}^{\sigma-1}}\right) - \frac{\sigma}{2}\left[V(tfpr_{n,US}) - V(tfpr_{n,EM})\right] \quad (14)$$

where "conv. output gain" is the output gain in percent; $\sigma = 3$ the elasticity of substitution between product varieties within industries; A_n is the technological efficiency ($TFPQ$) of plant n; and $V(tfpr_n)$ is the variance of log revenue total factor productivity. In the next two subsections, we generate the counterfactual variances of $tfpr$ that would arise if emerging economies had the

[21] For example, very few plants in the Indian manufacturing data report another plant in the same firm. We are grateful to Pete Klenow and Cian Ruane for making that information available to us.

[22] As a conservative robustness check, we consider below the quantitative implications of assuming no change in $mrpl$ dispersion between the two economies.

same multi-plant firm structure as the United States. By plugging this information into Equation (14), we can ultimately quantify additional aggregate TFP gains from internal capital markets. We start with the case of India.

5.1 Quantifying the gains from internal capital markets in India

To account for the coexistence of multi- and single-plant firms, we write the total dispersion of $tfpr$ across Indian plants as follows:

$$\begin{aligned} V(tfpr_n) &= \sum_n \omega_n (tfpr_n - \overline{tfpr})^2 \\ &= \omega^M V(tfpr_n^M) + (1-\omega^M) V(tfpr_n^S) + \omega^M (1-\omega^M) \left(\overline{tfpr}^M - \overline{tfpr}^S\right)^2 \end{aligned}$$

where the superscript index M denotes the set of plants belonging to multi-plant firms; ω^M their share in the economy; $\overline{tfpr}^M$ their average level of revenue total factor productivity; and $V(tfpr^M)$ their dispersion. The superscript index S indicates the analogue variables for standalone plants.

Our counterfactual exercise consists in computing $V(tfpr)$ in India under the assumption that the fraction of plants belonging to multi-plant firms, ω^F, which is 8.9% in India, is the same as in the U.S., which is 21.9%. We know from our quantitative exercise that dispersion across plants in multi-plant firms is 32.7% higher than across standalone plants. As a result, the effect will be to raise total dispersion in India, an additional variance that reflects *good dispersion*. Quantitatively, this is given by

$$\begin{aligned} \Delta V(tfpr_{n,IN}) &= (\omega_{IN}^S - \omega_{US}^S) \times (1.327 - 1) \times V(tfpr_{IN}^S) \\ &= (0.92 - 0.719) \times 0.327 \times V(tfpr_{IN}^S) \approx 6.9\% \times V(tfpr_{IN}^S). \end{aligned}$$

Plugging this number into Equation (14), we quantify the additional output gain as follows:

$$\begin{aligned} \log(1{+}\text{true output gain}) &= \frac{1}{\sigma - 1} \log \left(\frac{\sum_i A_{n,US}^{\sigma-1}}{\sum_j A_{n,IN}^{\sigma-1}} \right) + \frac{\sigma}{2} \left[V(tfpr_{n,IN}) + \boldsymbol{\Delta V(tfpr_{n,IN})} - V(tfpr_{n,US}) \right] \\ &= \log(1{+}\text{conv. output gain}) + \frac{\sigma}{2} \times 0.069 \times 0.469 \\ &= 47\% + 4.9\% = 51.9\% \end{aligned}$$

The implication is that internal capital markets raise the aggregate TFP and output gains by an additional tenth of what Hsieh and Klenow (2009) computed. In addition to that, if the typical Indian multi-plant firm operates fewer plants than its U.S. counterpart, then this number seems to be a lower bound on the gains of internal capital markets, thus limiting the scope for *good dispersion*. Additionally, we note that our quantitative analysis was computed in a framework with fixed adjustment costs; this mitigates the argument raised by Asker et al. (2014) that fixed adjustment costs can explain a large portion of the *bad dispersion* documented by Hsieh and Klenow

(2009).

5.2 Quantifying the gains from internal capital markets in China

We cannot employ the same methodology to compute the benefits from *good dispersion* in China, because the data are sampled at the firm level. We can, however, still compute the aggregate gains from a comparable measure of dispersion by focusing on the portion that occurs *between* firms j. While this means that we must omit the within-firm portion, it allows us to carry out the analysis of Hsieh and Klenow (2009) at a comparable level of aggregation. In short, unlike the India exercise, we will be comparing *bad dispersion* in the U.S. to *bad dispersion* in China.

We showed in Section 2.1 that between-firm dispersion was 39.9% of overall dispersion in U.S. manufacturing. Consequently, we adjust the U.S. variance in Equation (14) as follows:

$$\begin{aligned}
\log(1{+}\text{true output gain}) &= \frac{1}{\sigma-1}\log\left(\frac{\sum_i A_{j,US}^{\sigma-1}}{\sum_j A_{j,CHI}^{\sigma-1}}\right) - \frac{\sigma}{2}\left[V(tfpr_{j,US}) - V(tfpr_{j,CHI})\right] \\
&= \log(1{+}\text{conv. output gain}) + \frac{\sigma}{2}\times 0.601 \times V(tfpr_{n,US}) \\
&= 39\% + 14.8\% = 53.8\%
\end{aligned}$$

This calculation implies that the aggregate TFP gains in China are more than one-third larger than previously thought.

While the above exercises provide a useful starting point to think about the role of *good dispersion*, they remain preliminary, in part due to data limitations. For example, a more detailed analysis would consider the exact empirical distribution of plants per firm in each economy, differences in production functions, the number of firms and plants in each sector, and how both $TFPR$ and $TFPQ$ are distributed. In addition, one strong assumption we made is that $V(mrpl)$ behaves similarly to $V(mrpk)$ in the two-plant-firm and single-plant-firm economies. While they likely exist in the real world, fixed adjustment costs of labor are arguably of lesser importance than those impeding the allocation of capital. As a result, our assumption that $V(mrpl)$ is affected similar to $V(mrpk)$ by the presence of firm-level internal markets may be too strong, and the extra output gains relative to Hsieh and Klenow (2009) may not be as large.[23] Yet despite its limitations, our development accounting exercise highlights the importance of taking into account the within-firm dimension for aggregate outcomes.

6 Conclusion

This paper shows that dispersion in marginal revenue products of capital need not indicate distortions. Motivated by evidence that dispersion mostly occurs within firms rather than across firms,

[23] According to the expression of $TFPR$ on p.1410 in Hsieh and Klenow (2009), our additional output gains would only be an α_s portion of what we computed in the benchmark case above. This still remains a significant adjustment, highlighting the potential impact of our mechanism on exercises that compute the gains from reducing distortions in emerging economies.

we build a model of a firm operating several plants. Such firms have at their disposal an internal capital market that helps ease external financial constraints, support investment activity and generate extra output. Most importantly, economies with multi-plant firms may well exhibit more dispersion in marginal revenue products of capital than economies with single-plant firms, but still produce more aggregate output with the same technologies. An implication is that output gains from capital reallocation may be higher than previously thought in emerging economies, where single-plant firms are relatively more prevalent.

References

Andrew B. Abel and Janice C. Eberly. A unified model of investment under uncertainty. *American Economic Review*, 84(5):1369–1384, December 1994. (Cited on page 2.)

Oya Altınkılıç and Robert S. Hansen. Are there economies of scale in underwriting fees? evidence of rising external financing costs. *Review of Financial Studies*, 13(1):191–218, January 2000. (Cited on page 15.)

John Asker, Allan Collard-Wexler, and Jan De Loecker. Dynamic inputs and resource (mis)allocation. *Journal of Political Economy*, 122(5), October 2014. (Cited on pages 4, 5, 8, 10, 18, 23, and 28.)

Bee Yan Aw, Mark J. Roberts, and Daniel Y. Xu. R&D investment, exporting, and productivity dynamics. *American Economic Review*, 102(4):1312–1344, June 2011. (Cited on page 11.)

Rüdiger Bachmann and Christian Bayer. Investment dispersion and the business cycle. *American Economic Review*, 104(4):1392–1416, April 2014. (Cited on page 45.)

Yan Bai, Keyu Jin, and Dan Lu. Misallocation under trade liberalization. *Working Paper*, 2018. (Cited on page 9.)

Eric J. Bartelsmann, John C. Haltiwanger, and Stefano Scarpetta. Cross-country differences in productivity: The role of allocation and selection. *American Economic Review*, 103(1):305–334, February 2013. (Cited on page 41.)

Christian Bayer, Ariel M. Mecikovsky, and Matthias Meier. Productivity dispersions: Could it simply be technology choice? *Working Paper*, 2015. (Cited on page 4.)

Timothy Besley, Stephen Coate, and Glenn Joury. The economics of rotating savings and credit associations. *American Economic Review*, 83(4):792–810, September 1993. (Cited on page 23.)

Nicholas Bloom. The impact of uncertainty shocks. *Econometrica*, 77(3):623–685, May 2009. (Cited on page 47.)

Nicholas Bloom, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen Terry. Really uncertain business cycles. *Econometrica*, 86(3):1031–1065, May 2018. (Cited on pages 45 and 47.)

J. David Brown, Emin Dinlersoz, and John S. Earle. Does higher productivity dispersion imply greater misallocation? a theoretical and empirical analysis. *Census Discussion Paper CES-WP-16-42*, September 2016. (Cited on page 12.)

Francisco J. Buera, Joseph P. Kaboski, and Yongseok Shin. Finance and development: A tale of two sectors. *American Economic Review*, 101(5):1964–2002, August 2011. (Cited on page 4.)

Ricardo J. Caballero. Aggregate investment. In John B. Taylor and Michael Woodford, editors, *Handbook of Macroeconomics*, volume 1, Part B, chapter 12, pages 813–862. Elsevier, Amsterdam, December 1999. (Cited on page 2.)

Ricardo J. Caballero and Eduardo M. R. A. Engel. Explaining investment dynamics in U.S. manufacturing: A generalized (*S,s*) approach. *Econometrica*, 67(4):783–826, July 1999. (Cited on page 14.)

Ricardo J. Caballero, Eduardo M. R. A. Engel, and John C. Haltiwanger. Plant-level adjustment and aggregate dynamics. *Brookings Papers on Economic Activity*, 1995(2):1–54, 1995. (Cited on page 2.)

Lorenzo Caliendo, Ferdinando Monte, and Esteban Rossi-Hansberg. The anatomy of French production hierarchies. *Journal of Political Economy*, 123(4):809–852, August 2015. (Cited on page 12.)

Lawrence J. Christiano, Martin Eichenbaum, and Charles L. Evans. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45, February 2005. (Cited on page 23.)

Lawrence J. Christiano, Roberto Motto, and Massimo Rostagno. Risk shocks. *American Economic Review*, 104(1):27–65, January 2014. (Cited on pages 45 and 47.)

Xavier Cirera and William F. Maloney. *The Innovation Paradox: Developing-Country Capabilities and the Unrealized Promise of Technological Catch-Up*. The World Bank, Washington, DC, 2017. (Cited on page 9.)

Wesley M. Cohen and Steven Klepper. The anatomy of industry R&D intensity distributions. *American Economic Review*, 82(4):773–799, September 1992. (Cited on page 11.)

Wesley M. Cohen and Steven Klepper. A reprise of size and R&D. *Economic Journal*, 106(437): 925–951, July 1996. (Cited on page 11.)

Russell Cooper and John C. Haltiwanger. The aggregate implications of machine replacement: Theory and evidence. *American Economic Review*, 83(3):181–186, June 1993. (Cited on page 14.)

Russell Cooper and John C. Haltiwanger. On the nature of capital adjustment costs. *Review of Economic Studies*, 73(3):611–633, July 2006. (Cited on pages 2, 12, 17, 18, and 24.)

Russell Cooper, John C. Haltiwanger, and Laura Power. Machine replacement and the business cycle: Lumps and bumps. *American Economic Review*, 89(4):921–946, September 1999. (Cited on page 14.)

Russell W. Cooper and Immo Schott. Capital reallocation and aggregate productivity. *Working Paper*, August 2018. (Cited on page 45.)

Harold Creusen, Henk Kox, Arjan Lejour, and Roger Smeets. Exploring the margins of Dutch exports: A firm-level analysis. *De Economist*, 159(4):413–434, December 2011. (Cited on page 12.)

Ana Paula Cusolito and William F. Maloney. *Productivity Revisited: Shifting Paradigms in Analysis and Policy*. World Bank, Washington, D.C., 2018. (Cited on page 9.)

Era Dabla-Norris, Si Guo, Vikram Haksar, Minsuk Kim, Kalpana Kochhar, Kevin Wiseman, and Aleksandra Zdzienicka. The new normal: A sector-level perspective on growth and productivity trends in advanced economies. *IMS Staff Discussion Note*, March 2015. (Cited on page 9.)

Francesco D'Acunto, Ryan Liu, Carolin Pflueger, and Michael Weber. Flexible prices and leverage. *Working Paper*, 2016. (Cited on page 18.)

Sanghamitra Das, Mark J. Roberts, and James R. Tybout. Market entry costs, producer heterogeneity, and export dynamics. *Econometrica*, 75(3):837–873, May 2007. (Cited on page 12.)

Joel M. David and Venky Venkateswaran. The sources of capital misallocation. *American Economic Review*, 109(7):2531–2567, July 2019. (Cited on pages 9 and 12.)

Joseph A. DiMasi, Ronald W. Hansen, and Henry G. Grabowski. The price of innovation: New estimates of drug development costs. *Journal of Health Economics*, 22(2):151–185, March 2003. (Cited on page 11.)

Mark Doms and Timothy Dunne. Capital adjustment patterns in manufacturing plants. *Review of Economic Dynamics*, 1(2):409–429, April 1998. (Cited on pages 2 and 14.)

Janice Eberly, Sergio Rebelo, and Nicolas Vincent. What explains the lagged investment effect? *Journal of Monetary Economics*, 59(4):370–380, May 2012. (Cited on pages 14 and 24.)

Andrea L. Eisfeldt and Tyler Muir. Aggregate external financing and savings waves. *Journal of Monetary Economics*, 84:116–133, December 2016. (Cited on page 3.)

Andrea L. Eisfeldt and Dimitris Papanikolaou. Organization capital and the cross-section of expected returns. *Journal of Finance*, 68(4):1365–1406, August 2013. (Cited on page 5.)

Andrea L. Eisfeldt and Adriano A. Rampini. Capital reallocation and liquidity. *Journal of Monetary Economics*, 53(3):369–399, April 2006. (Cited on page 45.)

Steven M. Fazzari, R. Glenn Hubbard, and Bruce C. Petersen. Financing constraints and corporate investment. *Brookings Papers on Economic Activity*, 1988(1):141–206, 1988. (Cited on page 3.)

Lucia Foster, John C. Haltiwanger, and Chad Syverson. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1):394–425, March 2008. (Cited on pages 8 and 44.)

Xavier Gabaix. The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772, May 2011. (Cited on page 45.)

Mark Gertler and Nobuhiro Kiyotaki. Financial intermediation and credit policy in business cycle analysis. In Benjamin M. Friedman and Mark Woodford, editors, *Handbook of Monetary Economics*, volume 3A, chapter 11, pages 547–599. Elsevier, Amsterdam, December 2010. (Cited on page 15.)

Robert Gertner, David Scharfstein, and Jeremy Stein. Internal versus external capital markets. *Quarterly Journal of Economics*, 109(4):1211–1230, November 1994. (Cited on page 4.)

Simon Gilchrist and Charles P. Himmelberg. The role of cash flow in reduced-form investment equations. *Journal of Monetary Economics*, 36(3):541–572, December 1995. (Cited on page 3.)

Xavier Giroud. Proximity and investment: Evidence from plant-level data. *Quarterly Journal of Economics*, 128(2):861–915, May 2013. (Cited on pages 5 and 16.)

Xavier Giroud and Holger M. Mueller. Capital and labor reallocation within firms. *Journal of Finance*, 70(4):1767–1804, August 2015. (Cited on page 5.)

Xavier Giroud and Holger M. Mueller. Firms' internal networks and local economic shocks. *American Economic Review*, 2019. (Cited on page 5.)

João F. Gomes. Financing investment. *American Economic Review*, 91(5):1263–1285, December 2001. (Cited on page 3.)

Gita Gopinath, Şebnem Kalemli-Özcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. Capital allocation and productivity in South Europe. *Working Paper*, 132(4):1915–1967, November 2017. (Cited on pages 12, 38, and 45.)

François Gourio and Anil K. Kashyap. Investment spikes: New facts and a general equilibrium exploration. *Journal of Monetary Economics*, 54:1–22, September 2007. (Cited on page 2.)

John Haltiwanger, Robert Kulick, and Chad Syverson. Misallocation measures: The distortion that ate the residual. *NBER Working Paper No. 24199*, October 2018. (Cited on page 9.)

Christopher A. Hennessy and Toni M. Whited. How costly is external financing? evidence from a structural estimation. *Journal of Finance*, 62(4):1705–1745, August 2007. (Cited on page 15.)

Chang-Tai Hsieh and Peter J. Klenow. Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics*, 124(4), November 2009. (Cited on pages 2, 4, 5, 8, 10, 12, 23, 26, 27, 28, and 29.)

Matthias Kehrig. The cyclical nature of the productivity distribution. *Working Paper*, 2015. (Cited on pages 37, 38, 45, and 46.)

Truman L. Kelley. *Fundamentals of Statistics*. Harvard University Press, Cambridge, Massachusetts, 1947. (Cited on pages 6 and 38.)

Maurice G. Kendall and Alan Stuart. *Advanced Theory of Statistics*, volume 1. Oxford University Press, New York, 5th edition, 1987. (Cited on page 38.)

Aubhik Khan and Julia K. Thomas. Credit shocks and aggregate fluctuations in an economy with production heterogeneity. *Journal of Political Economy*, 121(6):1055–1107, December 2013. (Cited on pages 3 and 4.)

Owen A. Lamont. Cash flow and investment: Evidence from internal capital markets. *Journal of Finance*, 52(1):83–109, March 1997. (Cited on page 5.)

Andrea Lanteri. The market for used capital: Endogenous irreversibility and reallocation over the business cycle. *Working Paper*, 2016. (Cited on page 46.)

Shaojin Li, Toni M. Whited, and Yufeng Wu. Collateral, taxes, and leverage. *Review of Financial Studies*, 29(6):1453–1500, June 2016. (Cited on pages 17 and 18.)

David O. Lucca. Resuscitating time-to-build. *Working Paper*, 2007. (Cited on page 23.)

Andrey Malenko. Optimal design of internal capital markets. *Working Paper*, February 2016. (Cited on page 4.)

Gregor Matvos and Amit Seru. Resource allocation within firms and financial market dislocation: Evidence from diversified conglomerates. *Review of Financial Studies*, 27(4):1143–1189, April 2014. (Cited on page 5.)

Marc J. Melitz. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725, November 2003. (Cited on page 12.)

Virgiliu Midrigan and Daniel Y. Xu. Finance and misallocation: Evidence from plant-level data. *American Economic Review*, 104(2):422–458, February 2014. (Cited on pages 4 and 26.)

Benjamin Moll. Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review*, 104(10):3186–3221, October 2014. (Cited on page 4.)

Dale T. Mortensen and Christopher A. Pissarides. Job creation and job destruction in the theory of unemployment. *Review of Economic Studies*, 61(3):397–415, July 1994. (Cited on page 12.)

Ezra Oberfield and Devesh Raval. Micro data and macro technology. *Econometrica*, forthcoming. (Cited on page 42.)

Laura Power. The missing link: Technology, investment, and productivity. *Review of Economics and Statistics*, 80(2):300–313, May 1998. (Cited on page 14.)

Raghuram Rajan, Henri Servaes, and Luigi Zingales. The cost of diversity: The diversification discount and inefficient investment. *Journal of Finance*, 55(1):35–80, February 2000. URL `http://www.jstor.org/stable/222550`. (Cited on page 4.)

Adriano A. Rampini and S. Viswanathan. Collateral and capital structure. *Journal of Financial Economics*, 109(2):466–492, August 2013. (Cited on page 18.)

Plutarchos Sakellaris. Patterns of plant adjustment. *Journal of Monetary Economics*, 51(2):425–450, March 2004. (Cited on page 14.)

David S. Scharfstein and Jeremy C. Stein. The dark side of internal capital markets: Divisional rent-seeking and inefficient investment. *Journal of Finance*, 55(6):2537–2564, December 2000. (Cited on pages 4 and 16.)

Antoinette Schoar. Effects of corporate diversification on productivity. *Journal of Finance*, 57(6): 2379–2403, 2002. (Cited on page 5.)

Jae Song, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. Firming up inequality. *Quarterly Journal of Economics*, 134(1):1–50, February 2019. (Cited on pages 37 and 40.)

Jeremy C. Stein. Internal capital markets and the competition for corporate resources. *Journal of Finance*, 52(1):111–133, March 1997. (Cited on pages 4 and 16.)

Chad Syverson. What determines productivity? *Journal of Economic Literature*, 49(2):326–365, June 2011. (Cited on page 9.)

Julia K. Thomas. Is lumpy investment relevant for the business cycle? *Journal of Political Economy*, 110(3):508–534, June 2002. (Cited on page 24.)

Pavel Ševčík. Financial frictions, internal capital markets, and the organization of production. *Review of Economic Dynamics*, 18(3):502–522, July 2015. (Cited on page 5.)

Appendix

— for online publication —

A Additional empirical findings

A.1 Investment and equalizing the marginal revenue product of capital

The assumption that investment should ideally be undertaken to equalize marginal revenue products of capital rests on some assumptions that we work out in this appendix. In a standard frictionless economy with decreasing returns to scale, agents should choose investment to equate the expected capital return. To understand how equating capital returns relates to equating marginal revenue products, consider the expression of the expected capital return. This return is defined as the proceeds of one unit of capital at the end of next period – the value of undepreciated capital plus its marginal revenue product – divided by the cost of next period's capital in the current period. Defining industry output as the numéraire, we denote the price of next period's capital k' in terms of this period's numéraire by $P_t^{k'}$, the industry-wide depreciation rate by δ_t, the marginal revenue product of capital in plant n and year t by $MRPK_{nt}$, real value added by y_{nt}, and the capital stock by k_{nt}. Then the expected gross return, $\mathbb{E}\mathcal{R}_{nt+1}$, in a given year and industry is

$$\mathbb{E}\mathcal{R}_{nt+1} = \mathbb{E}\frac{P_{t+1}^{k}(1-\delta_t) + MRPK_{nt+1}}{P_t^{k'}}.$$

We assume all units in an industry face the same price of capital, $P_t^{k'}$, and the same depreciation rate δ_t. Then the only source of heterogeneity in returns stems from differences in expected marginal revenue products of capital, $MRPK_{nt+1}$. In a large set of models with Cobb-Douglas technology, this object is proportional to the expected average product of capital, $\mathbb{E}\frac{y_{t+1}}{k_{t+1}}$. Since we do not measure the expected marginal revenue product of capital, we approximate it by the realized marginal revenue product of capital. This is a good approximation if capital is chosen one period in advance, all other inputs are chosen statically, and total factor productivity is sufficiently persistent. Only unexpected innovations to profitability will then render the realized and the expected marginal revenue product of capital different. All of these assumptions are plausible and widely used in the macroeconomic and investment literature. From now on we study the logarithm of marginal revenue products of capital which is denoted by lower-case letters: $mrpk_{nt} \equiv \log(MRPK_{nt})$. Given our above assumptions, we measure dispersion in marginal revenue products of capital as $V_t\left(mrpk_{nt}\right) = V_t\left(\log(y_{nt}/k_{nt})\right)$.

A.2 Data

We mainly use confidential data on manufacturing establishments collected by the U.S. Census Bureau that comprise the 1972-2009 Annual Survey of Manufactures (ASM), the Census of Manufactures (CMF) from 1972-2007, and the Longitudinal Business Database (LBD) from 1976-2009. These data inform us about age, output, capital stocks, investment expenditures, and other inputs at the level of the individual establishment. In the manufacturing sector, the CMF defines an "establishment" as a business location where the principal activity is production; we hence think of an "establishment" as a production plant. The CMF also contains information about the ownership of each plant (denoted by the variable `FIRMID`), which allows us to construct the hierarchical plant structure of "firms" necessary for our main object of interest, the within-firm and between-firm component of heterogeneity in returns, productivity and reallocation.

From the CMF and the ASM, we construct a large dataset of plants in the U.S. manufacturing sector. In order to obtain a consistent longitudinal panel, we limit attention to the ASM and the ASM portion of the CMF data (identified by establishment type `ET`=0). We prefer the ASM over the CMF as our benchmark dataset because we want to test the dynamic implications of our model of investment in multi-plant firms at the highest possible frequency. Many aspects of our mechanism would disappear at the quinquennial frequency of the CMF. By focusing on the ASM portion in all years, we automatically eliminate all administrative observations (identified by `AR`=1), which are imputed mainly off industry means and would thus corrupt moments of the distribution we are interested in. Our resulting panel spans the years 1972-2009, which allows us to study the long-run features of the dispersion of marginal revenue products of capital and reallocation. Every year, we observe about 55,000 plants which total to 2.1 million observations.

We combine the Census data with industry-level data from several publicly available sources: input and output price deflators from the NBER-CES Manufacturing Industry Database (NBER-CES), various asset data from the the Capital Tables published by the Bureau of Labor Statistics (BLS), and the Fixed Asset Tables published by the Bureau of Economic Analysis (BEA). Unless otherwise noted, all datasets are at annual frequency. Most of the information contained in the non-Census datasets (BEA, BLS, NBER-CES) other than manufacturing data, are only needed to estimate productivity and the replacement value of capital at current market conditions.

For each plant in these data, we construct real value added, the real capital stock, and real investment. To obtain real value added, y_{nt}, we first compute nominal value added as sales less intermediate and energy inputs, correct for inventory changes and resales,[24] and deflate the resulting measure by the 6-digit NAICS shipment price deflator from the NBER-CES manufacturing database. The real capital stock, k_{nt}, is the sum of structure and equipment capital each of which are expressed as real replacement values at current market conditions. These replacement values are computed individually for structure and equipment capital with the perpetual inventory method, using investment expenditures and depreciation rates. When a plant is observed for the first time, we initialize its capital stock at its book value, which is transformed as follows. First, we convert nominal book values into nominal market values and then deflate this measure using the BLS's price deflators for capital goods at the 3-digit NAICS industry level.[25] Like capital, we compute real investment, i_{nt}, as the sum of real structure and equipment investment by deflating the respective nominal investment expenditures by the 3-digit NAICS industry investment price deflators from the BLS. Our capital measure denotes beginning-of-year stock values, while our investment and value added measures refer to flow values during the year. To avoid outliers driving our results about dispersion and the investment-productivity link, we drop the 1% tails of the productivity and investment rate distributions in a given 4-digit NAICS industry.

A firm is defined as all manufacturing plants within the same `FIRMID`[26] in a given year and 4-digit NAICS industry. The `FIRMID` defines the collection of plants under common ownership or control. All plants of subsidiary firms are included as part of the owning or controlling firm. If the same firm is active in several industries, we define each subset of plants belonging to the

[24]Resales are goods purchased from another producer and resold in an unchanged condition. Correcting for them means we assess productivity of the plant as a producer rather than its productivity as a trader.

[25]For more details about the primary data and the transformation needed to obtain measures of the real capital stock and estimate productivity, see the description in the appendix to Kehrig (2015).

[26]Song et al. (2019) identify firms off the `EIN`, the employer identification number, which comes from tax records. Since we are interested in organizational control rather than tax liability and because the same `FIRMID` may operate hundreds of `EIN`s for tax purposes, we prefer `FIRMID` to indicate firms.

same industry as separate firms. Our within-firm dispersion measures are hence an understatement because we ignore the between-industry component of within-firm dispersion.

A.3 The economic importance of multi-plant firms

Our between-firm/within-firm analysis is economically relevant if a significant portion of aggregate economic activity is accounted for by multi-plant firms. Table A1 shows that while single-plant firms dominate in numbers, multi-plant firms operate the majority of the capital stock, produce most output, and generate most investment. In fact, firms that consist of 20 or more plants operate almost one-half of all the capital stock in U.S. manufacturing.

Table A1: Economic activity by firm type in U.S. manufacturing

	Share of ...			
	plants	value added	capital stock	investment
Single-plant firms	0.719	0.220	0.178	0.215
Multi-plant firms	0.281	0.780	0.822	0.785
Firms with at least...				
... 10 plants	0.131	0.513	0.602	0.548
... 20 plants	0.095	0.398	0.470	0.421
... 40 plants	0.060	0.252	0.296	0.261

Note: The sample underlying this table comprises all establishments in the Census of Manufactures 1972-2007 less administrative records. The share of each variable in multi-plant vs. single-plant firms is computed for each Census year and then averaged across Census years. Non-manufacturing operations of firms are ignored.

A.4 Empirics of cross-sectional moments

This appendix details how the cross-sectional moments underlying Table 2, Panel B were computed. First, we compute cross-plant moments $\mathcal{M}_{it}$ and their standard errors in a given industry i and year t. $\mathcal{M}_{it}$ stands for the cross-sectional standard deviation, inter-decile range, skewness, Kelley skewness, and excess kurtosis. We adopt the formulae for the first four moments, the inter-quantile range and their standard errors from Kendall and Stuart (1987). Kelley skewness is a quantile-based measure of skewness whose predecessor was proposed by Kelley (1947).

Every cross-sectional moment is computed by industry and by year. To get long-run industry-specific moments, we first aggregate over years in order to exclude any industry-specific trends. As do Kehrig (2015); Gopinath et al. (2017), we note an upward trend in dispersion and – to a lesser extent – in skewness and a downward trend in kurtosis. Notice that the cross-plant standard deviation increases about 10 log points per decade; both between-firm and within-firm dispersion increases evenly, so there is no discernible trend in the within-firm share of the overall industry variance. The cross-plant skewness becomes more positive over time: Kelley skewness increases from around zero (unskewed) to 0.25 (right tail about 1.66 times as wide at the bottom tail) in 2007. We compute the typical cross-sectional moment in a given NAICS-4 industry in 1990 that corresponds to the middle of our sample.

Then, we aggregate across industries using that industry's average share in value added: $\mathcal{M}_t = \sum_i \omega_{it}\mathcal{M}_{it}$. Standard errors are computed according to this aggregation: $SE_{\mathcal{M}_t} = \sqrt{\sum_i(\omega_{it}SE_{\mathcal{M}_{it}})^2}$.

This yields the moments within the average industry in the middle of our sample.

A.5 Empirics of between-firm and within-firm moments

In this section, we detail how we compute the within-firm and between-firm dispersion in marginal revenue products of capital and capital reallocation that underlie Table 2 and the robustness exercises in Section A.7.

First, we decompose the overall variance in marginal revenue products of capital into three components: one between industries (reflecting differences in measurement and the definition of capital and value added), one between firms in a given industry, and one across plants within a firm and industry. We define firms that operate plants in separate industries as different firms, thus biasing the true within-firm component of dispersion downward.

$$
\begin{aligned}
V_t &= \sum_n \omega_{njit} \left(mrpk_{njit} - mrpk_t\right)^2 \\
&= \underbrace{\sum_i \omega_{it} \left(mrpk_{it} - mrpk_t\right)^2}_{V_t^{Ind} \text{ between-industry}} + \underbrace{\sum_i \omega_{it} \underbrace{\sum_{j \in i} \omega_{jt}^i \left(mrpk_{jit} - mrpk_{it}\right)^2}_{V_{it}^B \text{ between firms within ind. } i}}_{V_t^B \text{ average between-firm}} + \underbrace{\sum_i \omega_{it} \sum_{j\in i} \omega_{jt}^i \underbrace{\sum_{n \in j,i}^{N_j} \omega_{nt}^{ji} \left(mrpk_{njit} - mrpk_{jit}\right)^2}_{V_{jit}^W \text{ within firm } j \text{ and industry } i}}_{V_t^W \text{ average within-firm}}
\end{aligned}
\tag{A1}
$$

where n indicates the plant, j the firm, i the 4-digit NAICS industry and t the year. $mrpk_{njit}$ denotes the marginal revenue product of capital of plant n belonging to firm j and industry i in year t, $mrpk_{jit}$ the average return in firm j in industry i, $mrpk_{it}$ the average return in industry i, and $mrpk_t$ the average level of returns in the economy.

An industry's level of marginal revenue product of capital is determined by the level of P_t^k and the asset bundle it typically reflects in that industry. This and other industry specificities in measurement will artificially drive V^{Ind} – an object we ignore for its lack of economic meaning. In our empirical analysis in Section 2.1, we focus only on the V_i^B and V_i^W of firms with at least two plants, because it is meaningful to compare them and how much of the dispersion in marginal revenue products of capital within an industry originates within firms as opposed to between firms in that same industry: $\mathcal{W}_i \equiv \frac{V_i^W}{V^W i + V_i^B}$. When computing an "aggregate" number for $\mathcal{W}$, we compute the average of industry ratios, which is weighted by ω_i, i.e., that industry's share in plants or capital, depending on whether we are looking at unweighted or capital-weighted dispersion.

Although investment rates do not suffer from industry-specific measurement issues, such as the marginal revenue product of capital, we proceed in a similar way to assess between-firm and within-firm investment-rate dispersion.

A.6 Robustness

A.6.1 Accounting for measurement error

In Section A.7 we dealt with some measurement error. If plant-level variables are measured with noise, then firm-level averages will be measured more precisely and artificially inflate the within-firm variance. Time aggregation should filter out this type of measurement error. Because time aggregation cannot deal with persistent measurement error, we now consider that type. To do that, we consider marginal revenue products of capital that are computed using separate measures of capital and values added. Our alternative measures come from different datasets or are separately measured variables in our baseline dataset. We have:

- K^{TAB} — we use appropriately deflated values of variable `TAB` instead of the perpetual inventory method;
- Y^{IRS} — we use administrative data on sales from the IRS instead of `TVS` from CMF/ASM;
- Y^{PCU} — we use collected data on actual production from the Plant Capacity Utilization Survey (PCU) instead of `TVS`.

These alternative measures should be correlated with our original measures of K and Y in the ASM (since they measure the same underlying object), but they should still be different due to different coverage or handling by the statistical agency. We recompute marginal revenue products of capital using the three alternative measures and redo the cross-sectional within-firm between-firm decomposition on these alternative measures. If the dominance of within-firm share is true, then this should show up in all of these measures. Song et al. (2019) follow a similar procedure.

Table A2: Accounting for measurement error

	Alt. Measure	$Corr\left(\log\left(\frac{y}{k}\right)^{bench}, \log\left(\frac{y}{k}\right)^{alt}\right)$	$\left(\frac{V^W}{V^W+V^B}\right)^{bench}$	$\left(\frac{V^W}{V^W+V^B}\right)^{alt}$
I: CMF 1972-2007	K^{TAB}	0.979	0.563	0.556
			(0.008)	(0.005)
II: CMF 2002-2007	Y^{IRS}	0.990	0.538	0.542
			(0.005)	(0.005)
III: ASM 1974-2007	Y^{PCU}	0.494	0.581	0.626
			(0.015)	(0.017)

Note: This table displays the within-firm share of overall dispersion for alternative measures of value added Y – collected either from tax records or separately measured in the Plant Capacity Utilization Survey (PCU) – and capital K (real replacement value at current market prices directly computed from book values instead of from the perpetual inventory method). Correlation of the computed marginal revenue products of capital measures are positive, some are high, and the within-firm share of overall marginal revenue products of capital dispersion is not statistically different at the 95% level except when using value added from the PCU, which yields an even higher within-firm share. Error bands constructed from averaging across 86 NAICS-4 industries.

Since using these alternative measures limits our sample at times, we also recompute the within-firm/between-firm decomposition using our original data so that we are comparing moments for the same underlying sample, for which we have both our benchmark measure as well as the alternative. It turns out that the differences in the within-firm share are marginal and almost always lie in the 95% error bands of the other measure. Only when using value added from the PCU does the benchmark differ from the alternative, which yields an even higher within-firm share. Error bands constructed from averaging across 86 NAICS-4 industries. We conclude from this exercise that our main result of the within-firm dispersion accounting for the largest portion in overall dispersion does not go away when using alternative measures of output and capital.

A.6.2 Marginal vs. average revenue products

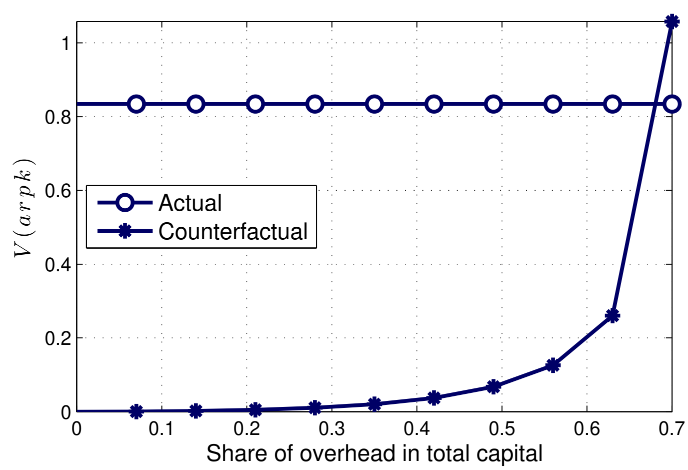
Our empirical work in Section 2.1 aimed at measuring marginal revenue products of capital, which are the relevant measure of what should be equalized across production units. But in the data, one can measure only average revenue products of capital; as in the literature, we approximate the dispersion of marginal revenue products with average revenue products. This approximation is usually justified with a Cobb-Douglas production function, since in this framework average revenue products are proportional to marginal ones. But they are not if technology is not multiplicative or if it is Cobb-Douglas with overhead inputs. Bartelsmann et al. (2013) document that overheads in production are quite powerful in explaining differences between micro production units, and they find significant aggregate consequences. Most overhead inputs are likely at the headquarters level of a firm rather than the plant level. Though this casts some doubt that all our results could be driven by constant inputs, we cannot dismiss this possibility.

Overheads Any constant input requirements at the plant level are hard to identify empirically. We therefore carry out a quantitative exercise to examine how large overheads would have to be in order to explain all or a portion of the empirically observed dispersion. Suppose the true technology is $y_n = z_n(k_n - \overline{k})^\alpha$. In that case, the average revenue product of capital can be written as $arpk_n = mrpk_n - \log\alpha + \log\left(1 - \overline{k}/k_n\right)$. Further suppose that marginal revenue products – which we cannot measure – are completely equalized. Then the entire variance of average revenue products would reflect the differential share of overheads across firms of different capital size:

$$V(arpk_n) = V\left(\log\left(1 - \frac{\overline{k}}{k_n}\right)\right). \tag{A2}$$

We simulate a firm-size distribution realistically, assuming that capital – unlike employment – is distributed log-normally. We consider how large the right-hand-side variance in Equation (A2) is for different levels of $\overline{k}$.

Figure A1: How much overheads is necessary to explain the observed $V(arpk)$?



Note: Simulation of the right-hand-side of Equation (A2) against the share of overhead in total inputs.

Figure A1 plots the RHS of Equation (A2) as a function of $\mathbb{E}\left[\overline{k}/k_n\right]$. Naturally, when $\overline{k} = 0$, this variance will be zero and the observed variance of average revenue products must be caused by the variance of marginal revenue products. This does not change much for low and moderate levels of overheads. Even if half of all inputs are overhead, less than one-tenth of the empirically observed variance in average revenue products can be explained by overhead. Clearly, this amount of overhead inputs at the level of the plant is unreasonable. Only if the average share of overhead in total inputs approaches 70% can the observed variance be explained by overhead. We conclude that overhead may only play a limited role in explaining the long-run dispersion of average revenue products of capital.

Non-unitary elasticity of substitution Another empirically plausible alternative to a simple Cobb-Douglas production function would be a constant elasticity of substitution production function. Suppose $y_n = \left[\alpha k_n^{\frac{\sigma-1}{\sigma}} + (1-\alpha) x_n^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$ where σ is the elasticity of substitution and x_n the variable inputs of plant n. In that case, the average revenue product of capital can be written as $arpk_n = \sigma[mrpk_n - \log \alpha]$ and the variance of average and marginal revenue products as

$$V(arpk_n) = \sigma^2 V(mrpk_n). \tag{A3}$$

The true dispersion of marginal revenue products could then be much lower if the elasticity of substitution is larger than unity. However, Oberfield and Raval (forthcoming), who estimate the elasticity of substitution at the plant level, put that number significantly smaller than 1, suggesting that the empirically measured dispersion of average revenue products would be a lower bound on that of marginal revenue products. We conclude that a non-unitary elasticity of substitution would most plausibly measure only a portion of the true dispersion of marginal revenue products.

A.7 Further dimensions of within-firm dispersion

In the previous section, we documented that most dispersion in marginal revenue products of capital and investment rates originates within firms rather than between firms. We now study the between-firm/within-firm decomposition for a number of subsamples. The objective is to confirm the robustness of our main empirical result and identify possible causes behind the importance of within-firm dispersion. The details of these exercises can be found in Appendix A.6.1. Table A3 provides an overview of our robustness checks.

The first row in Table A3 reiterates the baseline result of within-firm vs. between-firm dispersion of both marginal revenue products of capital (navy blue on the left) and investment rates (light blue on the right): About 60% of the variance of *mrpk* and 68% of the variance in i/k arise within firms.

Ruling out ASM sampling specificities We use the ASM as our benchmark panel. This panel is known to overrepresent large plants, which in turn are more likely to be part of a multi-plant firm. Since we do not know whether within-firm dispersion is larger in firms with few or many plants, we repeat our between-firm/within-firm decomposition for the entire Census of Manufactures (every five years). This allows us to study the full sample of manufacturing plants in the economy. Row (2) illustrates our findings: While the within-firm share of dispersion is slightly lower, it still remains dominant at 57% and 66% for *mrpk* and i/k, respectively.

Ruling out life-cycle dynamics Next, we examine whether our result is driven by entry, exit, or other life-cycle dynamics. Young, presumably more productive plants will be characterized by higher revenue products and will hence attract higher investment rates. The opposite may be true of older plants the firm keeps operational until the capital stock depreciates away. We therefore redo the decomposition using only "mid-age firms," which we define as plants that are at least three years old and at least three years away from exit. As Row (3) shows, the within-firm share of

Table A3: Dispersion of *mrpk* within and between firms

	Sample	Share of $V(mrpk)$		Share of $V(i/k)$	
		b/w plants within firms	b/w firms	b/w plants within firms	b/w firms
(1)	Full panel	0.601	0.399	0.679	0.321
(2)	Census sample	0.566	0.434	0.663	0.337
(3)	Mid-age plants	0.595	0.405	0.685	0.315
(4)	Balanced panel	0.809	0.191	0.895	0.105
(5)	5-year averages	0.554	0.446	0.656	0.344
(6)	5-plant firms	0.668	0.332	0.786	0.214
(7)	Homog. Industries	0.559	0.441	0.689	0.311
(8)	y is physical output	0.625	0.376		
(9)	K-weighted	0.540	0.460	0.680	0.320
(10)	Equipment	0.591	0.409	0.631	0.369
(11)	Private firms	0.713	0.287	0.795	0.205
(12)	Counterfactual firms	0.482	0.518		
		(0.030)	(0.043)		

dispersion in both variables is almost unchanged, and the same is true when we consider a five-year distance to entry and exit. If we turn to a strongly balanced panel (Row (4)), the share of within-firm dispersion is even larger.

Ruling out measurement error In our next robustness check, we want to address the possibility that the high within-firm share primarily reflects measurement error at the plant level. Indeed, transitory plant-level noise would "wash out" at the firm level, and thus artificially increase the within-firm share of dispersion. To rule out that this effect drives our result, we construct rolling 5-year windows of averaged *mrpk* and i/k for each plant. This time aggregation should filter out most high-frequency noise at the plant level. Performing the between-firm/within-firm decomposition on this averaged data shows that the importance of the within-firm share of dispersion persists. As shown in Row (5) of Table A3, the within-firm share of overall dispersion in an industry is now 55% and 66%, respectively. This suggests that the presence of plant-level measurement error is unlikely to be driving our findings. Further investigations of measurement error can be found in Appendix A.6.1.

Ruling out mechanical aggregation If firms were mere random collections of plants, one would expect the within-firm variance of *mrpk* to merely reflect random noise and not contain any meaningful economic information. To check how much dispersion within firms would arise from such randomness, we construct counterfactual firms in the following way: Within a 3-digit NAICS industry, we maintain the firms and how many plants each firm operates. Then, we randomly assign plants to all firms and repeat the between-within decomposition on this set of counterfactual firms. We repeat this exercise 1,000 times. Row (12) in Table A3 shows the average and standard deviations of these draws. The within-firm share of the variance drops from about 0.6 in the actual full panel to 0.48 in the bootstrapped sample of counterfactual firms. This difference is statistically significant, as the standard error of 0.03 makes clear.

Another way to look at the effects of mechanical aggregation is to study only samples of firms with the same number of plants per firm. In a world composed of many plants owned and operated by a single firm, the within-firm share of dispersion would be 100% by definition. As such, one could worry that the large share of dispersion occurring within the firm is driven by large entities. This is unlikely, since in our benchmark decomposition each firm receives equal weight, irrespective of its size. Because in our sample small 2-plant and 3-plant firms are much more numerous than large, complex firms, this bias probably does not play a large role. Still, in order to determine whether this mechanical aggregation could be an issue, we recompute the between-/within-firm decomposition for the set of firms that operate exactly five plants.[27] Row (6) shows that for this set of firms, the within-firm share of dispersion is indeed higher than for the whole manufacturing sector (67% for $mrpk$ and 79% for i/k), but not dramatically so. This suggests that our main result is not merely driven by the statistical importance of large firms.

Ruling out multi-product-firm bias Even though our benchmark definition of an industry is fairly fine, products within 4-digit NAICS industries are still heterogeneous. This could potentially lead to spurious differences of marginal revenue products of capital arising from differences in product composition within 4-digit NAICS industries. Whether such within-industry product differences are most likely to occur between or within firms is ambiguous. For robustness purposes, we repeat the decomposition but this time focus on plants that produce only one homogeneous standardized good. We follow Foster et al. (2008) and consider industries that produce almost perfectly homogeneous goods such as cement, sugar, coffee beans, etc.[28] Naturally, we expect the within-firm share of dispersion to be smaller then in our benchmark, since many firms within a given 4-digit NAICS industry are spread out across several of these narrowly-defined product codes. But even in these homogeneous industries, the within-firm share of dispersion in marginal revenue products of capital and investment rates displayed in Row (7) amounts to 56% and 69%, respectively.

Ruling out markups Limiting our attention to homogeneous goods has another advantage: It allows us to derive real value added in two ways. In addition to the standard approach of deflating sales, we can also use the measured physical quantity of production, a meaningful object for these homogeneous product groups. This makes it possible to study how much of dispersion in capital revenue products reflects price differences – due to differential markups or transfer prices – rather than physical productivity differences. Row (8) shows that the within-firm share of marginal revenue products of capital is slightly higher, at 63%, when using physical output to compute y rather than deflated sales. This suggests that if anything, prices impact the within-firm and between-firm variances in a way that stacks the odds against our main empirical finding. It is also consistent with the fact that plant-level prices and physical productivity are negatively correlated, as documented by Foster et al. (2008).

Demonstrating economic relevance Next, we wish to confirm that our findings are of economic relevance. Instead of decomposing the unweighted variance, we now consider capital weights

[27] According to our more restrictive definition of a firm as all plants operated by the same firm within a 4-digit NAICS industry, half of the capital stock is operated by firms with five plants or more.

[28] More specifically, these industries are defined as the following SIC product codes: Sugar (2061011), Block and Processed Ice (2097011 and 2097051), Gasoline (2911131), Hardwood flooring (2426111), Concrete (3273000), Whole Bean and Ground Coffee (2095111 and 2095117 & 2095118 – later merged into 2095115 – and 2095121), Carbon Black (2895011 and 2895000), Bread (2051111, later split into 2051121 and 2051122) and Plywood (2435100, later split into 2435101, 2435105, 2435107, and 2435147).

for the ω's in Equation (1) and redo the decomposition of the dispersion in $mrpk$ and i/k. Row (9) shows that while the within-firm share of capital-weighted dispersion is slightly lower, it still remains dominant at 54% and 68%, respectively.

Examining different capital types In Row (10), we display our decomposition results when focusing only on equipment capital when computing both revenue products and investment rates. Arguably, equipment can be more easily reallocated across production units than structures, which would lower dispersion. Again, results of the unweighted between-firm/within-firm decomposition are almost unchanged at 59% and 63%, respectively.

These exercises have confirmed that the importance of the within-firm share of dispersion in revenue products of capital and investment rates is robust to changing the sampling frame in order to account for measurement and aggregation problems, life-cycle dynamics, multi-product firms, markups and transfer prices, the predominance of multi-plant firms with little capital, the type of capital, or the sampling of the ASM. In many cases, the within-firm share of overall dispersion is even higher, suggesting that our baseline results may in fact represent lower bounds on the actual within-firm share of dispersion.

A.8 Cyclicality of dispersion between and within firms

So far, we have focused on time-series averages of between-firm and within-firm dispersion in marginal revenue products and investment rates. At the aggregate level, cyclical movements in either dispersion measure are well known: The countercyclical nature of productivity dispersion has been documented empirically using marginal revenue products of capital in Compustat data by Eisfeldt and Rampini (2006), TFP levels by Kehrig (2015), and TFP innovations by Bloom et al. (2018), while Bachmann and Bayer (2014) have shown that the dispersion in investment rates is procyclical. These findings have important implications for the literatures on Schumpeterian creative destruction, misallocation, development or uncertainty-driven business cycles. For example, Cooper and Schott (2018) study the effects of cyclical capital reallocation on aggregate productivity. Yet to our knowledge, no one has investigated separately the cyclicality of dispersion within firms. We close this gap by studying the time-series properties of the various components of dispersion. We first compute detrended measures of the between-firm and within-firm variance,[29] which we then use for time-series analysis. In addition, we study the cyclical properties of the lower and upper portions of the distribution as measured by the distance between the 90th percentile and the median as well as that between the median and the 10th percentile. Table A4 displays properties of the long-run averages, autocorrelations, and time-series standard deviations for each measure.

Consistent with the evidence from Section 2.1, Panel A.1 in Table A4 shows that the within-firm portion of the variance in $mrpk$ is larger than that between firms. When studying fluctuations of the two variances over time, we find that the volatility of the within-firm portion is also twice as strong as that of the between-firm portion. This is true even if one compares the time-series coefficient of variations instead of the time-series standard deviation, and similar patterns are observed in the between-firm and within-firm dispersion of investment rates.

As can be seen from Panel A.2. of Table A4, both $V^B(mrpk)$ and $V^W(mrpk)$ are countercyclical. This result could have important implications for the uncertainty literature. If one interprets dispersion as a cause of cycles, as do Bloom et al. (2018) and Christiano et al. (2014), then our within-firm result suggests that cycles manifest themselves within as well as between firms. This means that looking at the granular level of the firm, as Gabaix (2011) and Eisfeldt and Rampini

[29] As do Kehrig (2015) and Gopinath et al. (2017), we find an upward trend in cross-sectional dispersion.

Table A4: Dynamic properties of the *mrpk* and *i/k* distributions

	$V^B(mrpk)$	$V^W(mrpk)$	$V^B(i/k)$	$V^W(i/k)$
A.1 Dispersion: Time-series moments				
Average	0.315	0.476	0.016	0.037
Autocorrelation	0.732	0.657	0.667	0.679
Volatility	0.042	0.084	0.006	0.016
A.2 Dispersion: Cyclicality				
$Corr(\Delta Y_{t+1}^{mfg}, ...)$	-0.252	-0.147	0.306	0.330
$Corr(\Delta Y_{t}^{mfg}, ...)$	-0.582	-0.304	0.389	0.394
$Corr(\Delta Y_{t-1}^{mfg}, ...)$	-0.413	-0.229	0.222	0.241
	$mrpk^{50} - mrpk^{10}$	$mrpk^{90} - mrpk^{50}$	$i/k^{50} - i/k^{10}$	$i/k^{90} - i/k^{50}$
B. Skewness: Cyclicality				
$Corr(\Delta Y_{t+1}^{mfg}, ...)$	-0.026	-0.071	-0.155	0.079
$Corr(\Delta Y_{t}^{mfg}, ...)$	-0.293	-0.152	0.148	0.197
$Corr(\Delta Y_{t-1}^{mfg}, ...)$	-0.226	-0.109	0.225	0.244

Note: The table reports time-series moments of between-firm and within-firm variance for both marginal revenue products of capital and capital reallocation. "Average" denotes the long-run average of each variance term; "Autocorrelation" the annual persistence, $Corr(V_t, V_{t-1})$; "Volatility" the time-series standard deviation, $StD(V_t)$; and cyclicality is with respect to the growth rate of aggregate manufacturing value added, denoted by ΔY_t^{mfg}.

(2006) do, may underestimate the role of dispersion for aggregate fluctuations. Since fluctuations in heterogeneity at the subgranular level of plants within firms are larger and countercyclical, this suggests tha one should not discard the dynamics inside multi-unit firms when studying business cycles.

Lastly, we study which tail of the marginal revenue product and investment rate distribution is more cyclically sensitive. This is motivated by recent work on cyclical capital reallocation and the various frictions governing that process. Lanteri (2016), for example, develops an endogenous process of capital reallocation based on capital resales that posits that the left, less productive tail of the distribution of marginal revenue products is more cyclically sensitive. Panel B in Table A4 confirms this hypothesis empirically: The distance between the median and the 10th percentile is more countercyclical than its corresponding portion in the upper tail of the marginal revenue product of capital distribution. Such a cyclical pattern of the lower tail of the productivity distribution has been shown to hold as well for TFP levels by Kehrig (2015).

A.9 Discussion

More generally, our findings in this section imply that welfare gains from a more efficient allocation of resources would not only stem from reallocation across firms, but also within. This highlights the importance of developing a better understanding of the factors that impede capital from flowing to its most productive use inside the firm. As such, our findings have implications for micro-founded macroeconomic models and their calibration. In much of the literature, the concepts of plants and firms are used interchangeably, with little discussion of their respective roles and constraints. For example, in the empirical uncertainty literature, plants are almost always interpreted as independent decision makers facing various frictions that impede the reallocation of productive capital. Arguably, some frictions, such as technological ones as in Bloom (2009) and Bloom et al. (2018), are indeed most relevant at the level of the plant. Yet others are more likely to impact the decisions of firms. This is, for example, the case of external financing constraints, which affect interactions between firms and their lenders, as in Christiano et al. (2014).

What is the link between firm-level financial frictions and *mrpk* dispersion? To investigate this issue, we repeat our between-firm/within-firm decomposition on the sample of privately held firms only. We find that the share of within-firm dispersion for private firms is *higher* than for the whole sample, at 71% and 80% for *mrpk* and i/k, respectively. This is despite the fact that privately held firms tend to operate fewer plants than their publicly traded counterparts, which, as discussed earlier, leaves less room for within-firm dispersion in the first place.

This result may appear surprising. After all, it could be expected that by impeding the efficient allocation of capital, financial frictions would increase the dispersion of *mrpk* across firms. Yet our results indicate that firm-level borrowing constraints may in fact shape the allocation of capital across plants *within* the firm. This suggests that the internal capital market of a multi-plant firm could play an important role in overcoming external financial frictions. In the next section, we aim to gain insight into this channel by building a model of a multi-plant firm facing various types of frictions, including financial ones.

B Investment and financing policies

Here we present an illustration of the total cost of investing within the firm subject to the various adjustment costs and financing constraints. In addition to the external financing constraint, this illustration features a fixed cost of accessing external financial markets, a friction that we abandon in our quantitative model because it plays little role.

We plot the total cost of investment in Figure B1 to illustrate the multiple non-convexities and how the interaction of investment across plants shapes the cost of investment for the firm. Note that in these plots we assume that ψ, the parameter that regulates the fixed costs of investing in a plant, is "small" in the sense that the minimum investment in one plant can be financed using internal funds of the firm. If they were excessive, even the minimum investment to justify the fixed investment adjustment costs would require borrowing. In that case, the effective fixed cost of investing in any plant would be $(\psi + \zeta)k_{nt}$.

C Numerical solution of model

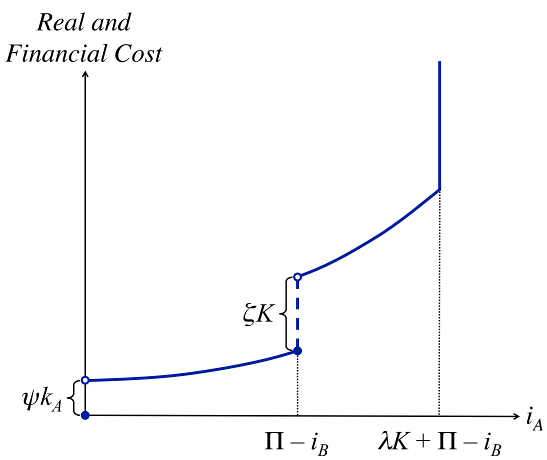
To solve the model, we discretize plant-level capital stock using an N_k-point grid, where $N_k = 100$ to produce the results in this paper. This implies that the two-plant capital grid contains a total of 10,000 points. In addition, the shock process is approximated by an 8-point grid, from the combination of plant-specific and firm-specific Markov chain processes.

We use a hybrid iterative method to solve the model. First, we iterate and maximize over the (k'_A, k'_B) pairs of plant-specific capital until convergence of the policy function. Then, we continue iterating until changes in the value of the firm between iteration steps is below a given threshold for all states.[30] We then ensure that the policy function is indeed stable. This method, while not particularly computationally efficient, allows us to handle the numerous non-convexities of our model. We also tested to verify that lowering or increasing N_k did not have any meaningful impact on our results.

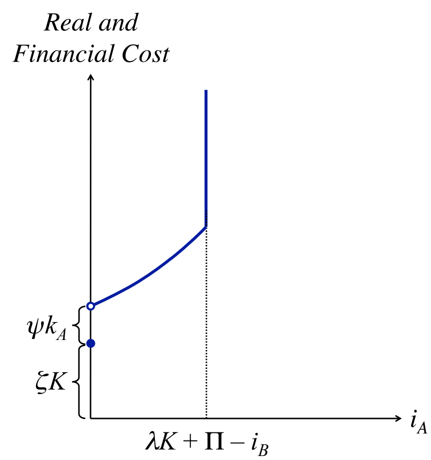
[30] Because we report the value of firms under various economic environments, we cannot solely rely on the convergence of the policy function.

Figure B1: Total cost of investment

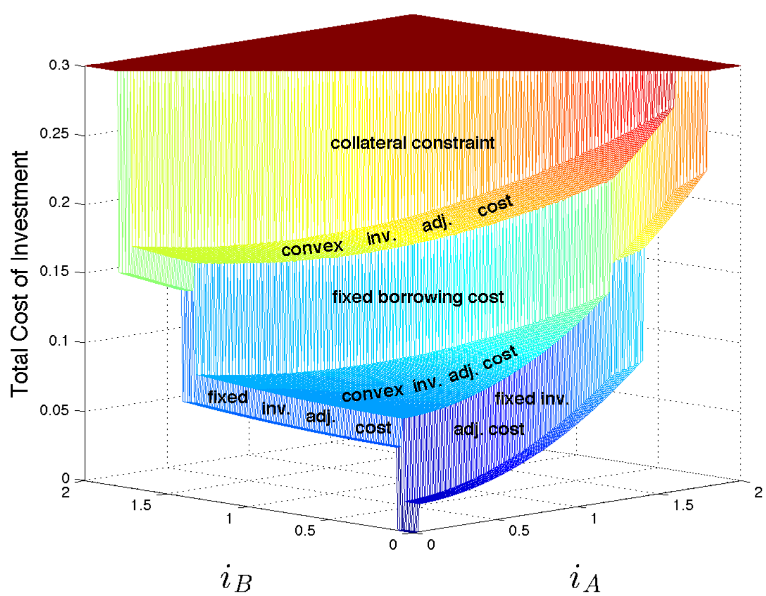
(a) Total cost of investing in plant A when i_B can be fully financed internally



(b) Total cost of investing in plant A when i_B cannot be financed internally



(c) Total investment costs for the firm



Note: Panel (a) on the left displays total cost of investing in plant A when investment in the rest of the firm does not exceed internal funds $0 \leq i_{Bt} \leq \Pi_t$. Then, small amounts of i_{At} can be financed with left over internal funds $(\Pi_t - i_{Bt})$ without incurring borrowing costs. Any investment exceeding that amount makes the cost level jump due an additional fixed borrowing cost, denoted ζK_t. Panel (b) on the right displays the case when the firm already needs to borrow to finance investment in the rest of the firm. Even zero investment in plant A means fixed and linear borrowing costs $\zeta K_t + R(i_{Bt} - \Pi_t)$. Investment in either case is always limited by the external financial constraint: $E_t \leq \lambda K \Leftrightarrow i_{At} \leq \lambda K + \Pi_t - i_{Bt}$. Panel (c) shows the the total cost jointly for $k_{At} = 1$, $k_{Bt} = 3$, $\zeta = \psi = 0.02$, $\gamma = 0.04$, $\vartheta = 0.03$, $\lambda = 0.1$ and $\Pi_t = 1.4$.

Next, we simulate a single two-plant firm over 100,500 periods, throw out the first 500 observations to allow for burn-in, and then create a panel of 500 two-plant firms with the simulations that were kept. This approach is appropriate because there are no aggregate shocks in our setup: With uncorrelated firm-level shocks, we are not required to simulate a panel of firms period-by-period. Simulated moments are computed on this firm panel.

However, since shocks are completely uncorrelated across firms in our panel, dispersion *across* firms is mechanically higher than it would be if we allowed for aggregate disturbances. Because our focus is not on aggregate time-series properties, we follow a different route and instead adjust the between-firm dispersion measures in order to match the within-firm share of *mrpk* dispersion found in the data (equal to 0.6; see Table 2). While this allows for more meaningful dispersion comparisons across various scenarios, no other moments are affected by this adjustment.