Do Firms Mitigate or Magnify Capital Misallocation? Evidence from Plant-Level Data*,**

Matthias Kehrig† and Nicolas Vincent‡

September 14, 2016

Abstract

More than half of the cross-plant dispersion in average capital returns occurs across plants within the same firm rather than between firms. Even though firms allocate investment very differently across their plants, they do not equalize average returns across their plants. We reconcile these findings in a model of multi-plant firms, physical adjustment costs and credit constraints. When multi-plant firms are credit constrained, they can utilize internal capital markets by concentrating internal funds on investment in only a few of their plants in a given period and rotating funds to another set of plants in the future. The resulting increase in within-firm dispersion of capital returns is hence not a symptom of misallocation within the firm, but rather indicates the mitigating of external credit constraints. Economies with multi-plant firms produce more aggregate output despite *higher* dispersion in capital returns compared to economies with single-unit firms. The efficiency gains from bringing down misallocation in emerging single-plant-firm economies to the level of developed multi-plant-firm economies are even larger than previously thought.

Keywords: Misallocation, Productivity Dispersion, Investment, Plants vs. Firms, Internal vs. External Capital Markets.
Jel Codes: E2, G3

*This draft is partly based on our earlier paper "Financial Frictions and Investment Dynamics in Multi-Plant Firms." We thank seminar participants at Duke, Oxford, Penn State, Iowa, UW Madison, Arizona State, Ente Einaudi, Boston College, UT Austin, the Federal Reserve, the Dutch Central Bank, the Barcelona Summer Forum, the NBER Summer Institute, the Ifo Conference on Macroeconomics and Survey Data, the Midwest Macro Meetings, the Empirical Macro Workshop and the Annual Meeting of German Economists Abroad as well as Klaus Adam, Gadi Barlevy, Marco Bassetto, Christian Bayer, David Berger, Jeff Campbell, John Cochrane, Russ Cooper, François Gourio, John Haltiwanger, Hugo Hopenhayn, Ben Moll, Gordon Phillips, Michèle Tertilt and Mark Wright for helpful discussions.

**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 http://papers.ssrn.com/abstract=2731594

†University of Texas at Austin, Department of Economics, 2225 Speedway Stop C3100, Austin, TX 78712, matthias.kehrig@austin.utexas.edu.

‡Institute of Applied Economics, HEC Montréal, 3000, chemin de la Côte-Sainte-Catherine, Montréal (Québec) H3T 2A7, 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 in turn to a revival of interest in the causes and consequences of resource misallocation. For instance, Hsieh and Klenow (2009) have found substantial aggregate output increases if one managed to lower this dispersion by improving the allocation of resources across production units. Alternatively interpreting dispersion as a consequence of exogenous increases in uncertainty, Bloom (2009) has shown the recessionary effects of higher productivity dispersion. The implicit assumption in the literatures growing out of these seminal papers is that a high level of productivity dispersion is a sign of resource misallocation in an economy and therefore reduces welfare. But does higher productivity dispersion always depress output and welfare? In this paper, we explore the opposite hypothesis: we show theoretically that optimal behavior of agents may lead to increased productivity dispersion. Novel empirical evidence from data on multi-plant firms supports our hypothesis. We then build a model of investment in multi-plant firms which, calibrated to U.S. manufacturing, attributes up to half of the empirically observed productivity dispersion to optimal firm behavior rather than misallocation. As a consequence, our theoretical and empirical work suggests that the gains from reducing inefficient distortions in developing economies to the level of the U.S. are considerably larger than previously thought.

To illustrate our theoretical argument, consider an economy with many production units which are perfectly identical in terms of technology, productivity and capital. Given their state, each unit would find it optimal to undertake an investment project of a fixed size, but cannot do so for a lack of access to external finance. With this extreme financing constraint, no investment activity takes place and productivity dispersion across units – conceptualized as the dispersion in capital returns – remains zero. Suppose instead that there is a benevolent planner with a limited amount of funds. These funds suffice to finance the investment project in some but not all plants. So the planner optimally finances the investment project in as many plants as the funds permit. Productivity dispersion would then rise, reflecting investment in some plants and no investment in others. A naive interpretation of this higher productivity dispersion would assume that the latter allocation is less efficient than the former where no investment takes place at all. Yet, in reality, the higher dispersion in the latter case is the outcome of a benevolent planner doing his best to overcome a financing constraint and generating more investment, capital and output in the aggregate.

This example demonstrates that productivity dispersion – commonly taken as an indicator for insufficient factor reallocation – may rise or fall with efficiency and thus aggregate output, while reallocation intensity and aggregate output are monotonically linked. It is therefore unclear how much of the empirically observed productivity dispersion reflects inefficient distortions and how much reflects (constrained) efficient behavior. This ambiguity warrants an empirical investigation into the origins of productivity dispersion. While research in the above mentioned literatures are concerned with the impact of misallocation and uncertainty on the aggregate economy, there is little work on why and where in the economy such significant dispersion originates in the first place. Yet, identifying the agents and frictions that are mainly responsible for the substantial productivity dispersion is crucial to assessing how much of that dispersion reflects constrained optimal behavior versus true misallocation. This knowledge in turn is relevant for policymakers interested in fostering growth and mitigating recessions. In the present paper, we fill this gap in the literature by jointly studying the dispersion of capital returns between firms *and* between plants within firms. As a starting point, we document two novel empirical findings based on the plant-level data from the U.S. Annual Survey of Manufactures: First, more than half of the overall dispersion originates across plants within the same firm rather than between firms. Second, when overall dispersion rises in recessions, we show that it rises primarily within firms rather than between firms. Our findings are quantitatively significant: a typical firm may increase its output by about 28% if it shifted an optimal amount of resources towards high-return plants. In strong downturns such as the 1980/82 recession, this potential output gain increases to 35%.

This focus on both the plant and the firm is relevant for a number of reasons: First, it allows us to highlight the fact that the firm and the market are two fundamentally different allocation mechanisms. In the micro-founded macroeconomic literature, plants are typically treated as independent decision makers. Yet, the allocation of resources across plants could conceivably be affected if they are part of an integrated firm or are required to share scarce firm inputs such as logistics and external credit. Consequently, understanding how firms work differently than markets may shed new light on the causes of resource (mis)allocation or how the economy copes with aggregate disturbances such as credit or uncertainty shocks.

Second, our findings potentially inform us on whether firms or markets are the most appropriate determinants of allocative efficiency and economic growth: if the dispersion of capital returns across plants reflects resource misallocation, then our empirical findings may indicate that firms are in fact inferior to markets at allocating inputs. In other words, distorsions such as principal-agent problems, transaction costs and asymmetric information may not be mitigated significantly by firms.

Third, the economy-wide effects of aggregate shocks are often largely shaped by frictions that hamper resource reallocation at the micro level. While a large variety of micro-level frictions has been suggested and studied in the literature, little effort has been made to understand at what granular level these frictions matter quantitatively. Since we are studying the dispersion of capital returns, we give some examples of frictions that interfere with capital reallocation and investment in general. For instance, some studies have highlighted the lumpy nature of investment at the *plant* level (see Doms and Dunne (1998) and Gourio and Kashyap (2007)), while others have found that capital expenditures at the level of the *firm* is fairly smooth (Eberly et al. (2012)). In the literature on uncertainty shocks, Bloom et al. (2012) combine such shocks with non-convex adjustment costs of capital that are most pertinent at the plant level. Christiano et al. (2014), in contrast, focus instead on financial constraints, a class of frictions that are of relevance at the level of the firm. These two separate ways of modeling the effects of uncertainty shocks require empirical productivity dispersion to fluctuate at different aggregation levels. Hence, documenting at which level of aggregation the dispersion of capital returns is most prevalent and volatile could eludicate the type of reallocation frictions that are most relevant.

All of the above reasons are strong arguments to investigate further the extent and dynamics of productivity dispersion. In this paper, our focus is on questions related to the link between within-firm dispersion in capital returns and investment activity. For instance, why do firms tolerate such large dispersion of capital returns across their plants? As an allocation mechanism of scarce resources, are firms less efficient than the market? Or is this dispersion simply optimal given the constraints, both internal and external, faced by the firm? How much do firms discriminate in allocating capital across their plants? At first sight, they appear to do so to a large extent: we find that dispersion in investment rates across plants is also large and more important within than between firms. In line with other studies highlighting the lumpy nature of investment (see Doms and Dunne (1998) and Gourio and Kashyap (2007)), we find that this dispersion is largely driven by investment spikes in a small number of plants. Yet, the correlation between a plant's capital return and a plant's investment within a firm is not very strong even when the firm invests as a whole. Moreover, we find that this correlation is even lower in periods of tight credit, exactly when the within-firm dispersion of capital returns is largest. Eisfeldt and Rampini (2006) document a similar mis-timing of capital returns dispersion and capital reallocation between firms. Though we confirm their evidence, we find the larger portion of dispersion and mis-timed reallocation to occur between plants within firms.

This last finding hints at a potential role for *external* financial frictions in shaping the degree of capital misallocation *within* the firm. To investigate this channel, we build a novel dynamic investment model of a multi-plant firm that features technological, organizational and financial interdependencies and constraints. The purpose of the model is threefold. First, we wish to establish how much of the empirically observed dispersion in capital returns and investment rates can be explained by our model. Second, we investigate the relative importance of the various frictions and their interactions as determinants of this apparent misallocation. Third, we study how much of the countercyclical dispersion in capital returns could result from optimal firm behavior.

Because of the focus in the literature on single-plant models, we are required to build a new model of a firm operating multiple distinct plants. Yet, in shaping our framework, we are mindful of the many frictions and imperfections that have been suggested in models of capital accumulation and reallocation (see Caballero (1999) for an overview). The investment literature, for example, has argued that "technological" frictions such as a fixed cost of investing are crucial to replicate the lumpy aspect of investment documented empirically (see among others Abel and Eberly (1994), Caballero et al. (1995), Cooper and Haltiwanger (2006), Gourio and Kashyap (2007)). Others have instead focused on the central role played by financing frictions, arguing that they offer a natural explanation for the documented role of cash flow in investment regressions (see among others Fazzari et al. (1988), Gilchrist and Himmelberg (1995)). Most of these papers have considered frictions in isolation.[1] We are the first to nest real and financial frictions in a model of a multi-plant firm. While some, mostly technological, frictions matter at the level of the plant, others such as financial or organizational imperfections affect the firm as a whole. We argue that modeling a multi-plant structure is crucial to understanding the effects of various frictions and their interactions. For instance, focusing solely on firm-level frictions would counterfactually predict the absence of any dispersion in capital returns across plants within firms. On the other hand, with only plant-level frictions, one would expect to see no difference between plants in single-unit versus multi-unit firms. our understanding about which micro-level frictions matter for aggregate outcomes may be affected by the multi-plant structure of firms in the economy.[2] Though its first aim is to characterize the investment policy of multi-plant firms, our model will also enable us to address questions about the aggregate effects of plant- and firm-level frictions.

In our model, firms operate plants that face both fixed and convex adjustment costs while the firm organizes internal and costly external financing of investment. When firms are credit constrained, they leverage internal capital markets and focus investment on only a few plants *even if most of their plants are equally productive.* This credit constrained investment policy leads to a rise in the dispersion of capital returns within the firm. In other words, had the firm unlimited access to external funds, it would invest in a way to minimize further this dispersion, to the extent allowed by the technological constraints it faces. This is consistent with Asker et al. (2014) and Bartelsmann et al. (2013), who have shown that non-convexities such as fixed investment adjustment costs or overhead labor can lead to endogenous dispersion of capital returns in equilibrium. Our model contains both features, but we show that their impacts can be magnified by the presence of costly external finance. In the face of such financing constraints, we document that the firm's optimal policy is to allocate internal funds to only a handful of its plants in a given period, postponing investment projects in its other plants. This staggered allocation of scarce resources implies that an externally credit constrained firm will display heightened capital returns and investment dispersion across its plants for several periods. The serial correlation of large investment projects, however, will increase in credit-constrained firms compared to their unconstrained counterparts. Using micro data from the U.S. Annual Survey of Manufactures, we find empirical support for these predictions at the micro and macro levels.

[1]Gomes (2001) and Eisfeldt and Muir (2013) who have combined real and financial frictions in a unified model of a firm operating one plant are notable exceptions.

[2]We study this question in a companion paper Kehrig and Vincent (2015): "Financial Frictions and Investment Dynamics in Multi-Plant Firms."

As described above, our model can explain dispersion of capital returns and investment rates within firms qualitatively. But dispersion may result from various other reasons beyond the scope of our model mechanism. We therefore analyze the quantitative predictions of our model. Calibrating it to micro-level moments in our data, we find that introducing financial firm-level constraints increases capital returns dispersion within the firm by 43% compared to dispersion already caused by technological non-convexities at the plant level. This suggests that our mechanism is quite powerful in generating substantial and countercyclical within-firm heterogeneity. This also means that dispersion of capital returns within financially-constrained firms is not necessarily an indicator of misallocation or inefficient investment policies, but rather the by-product of the firm's effort to mitigate external financial frictions. On the contrary, if the firm did not leverage its internal capital market, dispersion might remain at a lower level, but overall investment, growth and output would be lower as well. Internal capital markets thus help an economy cope more easily with periods of tight financial funding.

In a counterfactual exercise, we find that the existence of internal capital markets in multi-plant firms can increase within-firm dispersion by up to 20%, yet also raise aggregate investment and output by up to 8% and 5%, respectively, relative to an economy in which the pooling of resources within the firm is shut down. We view this finding as a cautionary tale to interpreting higher productivity dispersion as a sign of resource misallocation.

We see our project as a first step into modeling 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. Some theoretical research has been done on the efficiency of internal versus external capital markets: Stein (1997), Malenko (2012) study mostly principal-agent problems between a firm's owner and manager in a single-plant setup. Whether internal capital markets are more or less efficient is theoretically ambiguous: Gertner et al. (1994) show that division managers may exploit imperfect monitoring by firm headquarters to build up "inefficient empires." On the other hand, Scharfstein and Stein (2000) say that firms are likely to monitor investment spending better than outside financiers, thus improving capital allocation. 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), empirical research on within-firm dynamics is scarce. That said, our paper currently falls short of thoroughly studying the general equilibrium effects on multi-plant firms in the spirit of Thomas (2002) and Khan and Thomas (2008). On the empirical front, we believe that taking seriously this firm-plant dichotomy and the joint investment dynamics of plants within a firm offers a new dimension for the researcher to identify various types of frictions.

Our paper is organized as follows. In Section 2, we describe the data and show evidence on the importance of the within-firm dimension for the dispersion of capital returns and investment.

Section 3 describes our multi-plant model of the firm and analyses its predictions when an external financing constraint is introduced. Section 4 conducts various quantitative exercises geared towards understand the nature of productivity dispersion and carry out counterfactual exercises. In Section 5, we investigate whether the model predictions are borne out in the micro-level data. Section 6 concludes.

2 Returns dispersion between and within firms

2.1 Data sources and variables of interest

We use annual Census data on manufacturing plants described in detail in Appendix A.1. Our main object of interest are the dispersions of real returns of capital and the real investment rates across plants in the economy. They reflect the differences in capital returns and capital reallocation, respectively. In a standard frictionless economy, agents should choose investment rates in a way to equalize expected returns on capital. Hsieh and Klenow (2009), Asker et al. (2014) among others assess misallocation by studying the dispersion of revenue total factor productivity, denoted TFPR. While this is appropriate in their context, it is not in ours where we want to proceed to study investment. Investment should not necessarily flow towards the units with the highest TFPR if they already operate a large capital stock. In fact, investment should flow to units with high 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. We define industry output as the numéraire, and the price of next period's capital k' in terms of this period's numéraire is denoted by $P_t^{k'}$. Let δ denote the industry-wide depreciation rate, $MRPK_{nt}$ the marginal revenue product of capital in plant n in year t, y_{nt} and k_{nt} plant n's real value added and capital stock, respectively. Then the expected gross return, $\mathbb{E}\tilde{\mathcal{R}}_{nt+1}$, in a given year and industry is

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

We assume all units in an industry face the same price of capital, $P_t^{k'}$, and depreciation rate δ. 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, $\mathbb{E}\frac{y_{t+1}}{k_{t+1}}$. Equalizing expected capital returns is hence closely related to equalizing the marginal revenue product of capital. Since we do not measure expected returns, we approximate them by realized capital returns. 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 make the realized and the expected return different. All these assumptions are plausible and widely used in the macroeconomic and investment literature. From now on we study the logarithm of realised returns, labeled $\mathcal{R}_{nt+1}$, and refer to its dispersion as "(net) capital returns dispersion"

$$Var_t\left(\mathcal{R}_{nt+1}\right) = Var_t\left(\mathbb{E}\log\frac{y_{t+1}}{k_{t+1}}\right). \tag{1}$$

In order to study the dispersion of capital returns in equation (1) as well as that of capital reallocation as measured by investment rates requires us to compute real value added, the real capital stock and real investment, denoted by i_{nt}. To obtain real value added, we first compute nominal value added as sales less intermediate and energy inputs, corrected for inventory changes and resales, 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 for both structure and equipment capital are individually computed with the perpetual inventory method using investment expenditures and depreciation rates after initializing the capital stock at a transformed book value when observed for the first time. We transform these nominal book values into nominal market values and finally deflate this measure using BEA's price deflators for capital goods at the 3-digit NAICS industry level. Like capital, we compute real investment 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 BEA. Our capital measure denotes the beginning-of-year stock values while our investment and value added measures refer to flow values during the year.

2.2 Dispersion of returns and reallocation within firms

The goal of our empirical work is to study and document the dispersion of capital returns and capital reallocation. In addition to informing researchers about the empirical patterns necessary to calibrate micro-founded macroeconomic models, we place a novel emphasis on assessing the relative importance of dispersion across firms and within firms. To get an idea of the relevance of multi-plant firms, Table 1 displays aggregate economic statistics by firm type. While single-plant firms dominate in numbers, it is the multi-plant firms that operate the majority of the capital stock, produce most output and generate most investment. Moreover, about half the capital stock is operated by firms which consist of five or more plants. This suggests that firms have to consider a complex hierarchy of production units when allocating capital inside their firm. We now proceed to what extent capital returns and investment rates differ across plants within multi-plant firms and how that compares to the differences between firms.

We denote the cross-sectional variance of capital returns displayed in equation (1) by σ_t. Since

Table 1: Economic activity by firm type in U.S. manufacturing

	Share of...			
	plants	value added	capital stock	investment
Single-plant firms	0.760	0.298	0.260	0.293
Multi-plant firms	0.240	0.702	0.740	0.707
Firms with at least...				
... 5 plants	0.122	0.438	0.499	0.462
... 10 plants	0.073	0.244	0.293	0.267
... 15 plants	0.051	0.157	0.184	0.173

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.

capital returns are expressed in units of industry good per unit of capital, comparing dispersion of capital returns across industries is not very meaningful. Any arbitrary definition of what constitutes a typical good in an industry will automatically impact the dispersion of measured capital returns in the economy. We therefore focus on dispersion of capital returns within a typical industry and report results that are averaged across industries (details on the aggregation provided in Appendix A.3.1). We define an industry at the 4-digit NAICS level which makes goods within such industries sufficiently homogeneous to compare their capital returns while leaving enough observations in an industry-year cell to reliably study heterogeneity. In a given 4-digit NAICS industry, capital returns dispersion can be decomposed into the dispersion between firms, denoted by σ_t^B, and the average dispersion between plants within firms, denoted by σ_t^W:

$$Var_t\left(\mathcal{R}_{nt}\right) \equiv \sigma_t = \underbrace{\sum_j \omega_{jt}\left(\mathcal{R}_{jt} - \overline{\mathcal{R}}_t\right)^2}_{\sigma_t^B \text{ average between-firm}} + \underbrace{\sum_j \omega_{jt} \underbrace{\sum_{n \in j}^{N_j} \omega_{nt}^j \left(\mathcal{R}_{njt} - \mathcal{R}_{jt}\right)^2}_{\sigma_{jt}^W \text{ within firm } j}}_{\sigma_t^W \text{ average within-firm}} \tag{2}$$

where n indicates the plant, j the firm and t the year. $\mathcal{R}_{njt}$ denotes the logarithm of the net capital return of plant n belonging to firm j in year t, $\mathcal{R}_{jt}$ the average return in firm j in an industry, and $\overline{\mathcal{R}}_t$ the average level of returns in a given industry. ω_{njt} is the weight of plant n at time t, ω_{jt} that of firm j and ω_{nt}^j that of plant n just inside firm j. While unweighted dispersion is our benchmark, we also consider capital weighted dispersion to get a measure of "economic relevance." In the former case $\omega_{njt} = 1/N_t$ (where N_t is the number of observations), in the latter $\omega_{njt} = k_{njt}/k_t$ and accordingly for ω_{jt} and ω_{nt}^j. The basic results of that decomposition are displayed in Panel A.

of Table 2.

Table 2: Cross sectional moments of capital and investment

	Capital returns $\log(y/k)$	Investment rates i/k
A. Dispersion across and within firms		
Share of overall variance originating ...		
... between firms	0.399	0.321
... between plants within firms	0.601	0.679
B. General moments across plants		
Standard deviation	0.915	0.306
	(0.003)	(0.009)
Inter-decile range	2.162	0.164
	(0.012)	(0.036)
Coefficient of skewness	0.554	6.221
	(0.005)	(0.005)
Kelley skewness	0.118	0.491
Excess kurtosis	1.699	60.213
	(0.011)	(0.011)

Note: 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 all Census years 1982-2007 are computed for each industry and years first before being aggregated by industry and the averaged across years. For details see Appendix A.2.

Two results from our simple accounting exercise stand out: First, overall dispersion in capital returns is large. In the average industry and year, the standard deviation of log capital returns is 0.915. This means that a plant one standard deviation above the mean produces $e^{0.915} \approx 2.5$ times the value added as the average plant with the same capital stock; the difference between the plant at the 90th percentile compared to that at the 10th percentile even implies an $e^{2.162} \approx 8.7$-fold value added difference. Reallocating capital to high-return plants in the same industry should hence result in a considerable boost in aggregate output. If one interprets capital returns dispersion as misallocation, this means that the economy foregoes a lot of income.

The second conclusion of our accounting exercise is that the majority of dispersion occurs inside firms rather than between firms: over 60% of capital returns dispersion in a given industry is explained by variation of capital returns within firms. The aforementioned benefits from reallocation would thus firstly stem from reallocation across the production plants within a given firm. Analogously to the above exercise, we document that a plant one standard deviation (0.709) above the firm's average produces twice the value added with the same capital stock as the firm on average.

Interestingly, the cross-sectional distribution of capital returns is positively skewed which is consistent with the presence of credit constraints: productive firms may have too little capital which thus drives up their return and skews the returns distribution to the right. Following Kelley (1947), p. 250, we define the Kelley skewness as $\gamma_{Kelley} = \frac{\mathcal{R}^{90}+\mathcal{R}^{10}-2\mathcal{R}^{50}}{\mathcal{R}^{90}-\mathcal{R}^{10}}$. This quantiles based measure of skewness is 0.118 on average which means that the top half of the distribution, $\mathcal{R}^{90} - \mathcal{R}^{50}$ is about $1.118/0.882 = 27\%$ more spread out as the bottom half of the distribution, $\mathcal{R}^{50} - \mathcal{R}^{10}$.

Investment rates also differ substantially across production plants which means that capital is allocated very differently across these units. Since investment rates are measured in percentage points, they are not subject to the same industry-specific measurement problems present with capital returns. The cross-sectional standard deviation of 30.9% is large given that the average plant in the economy has an investment rate of 10.2%. Like with capital returns dispersion, the majority of investment rate dispersion originates within firms rather than between firms: within a given industry the within-firm share of investment rate variance amounts to almost 70%.

2.3 Robustness

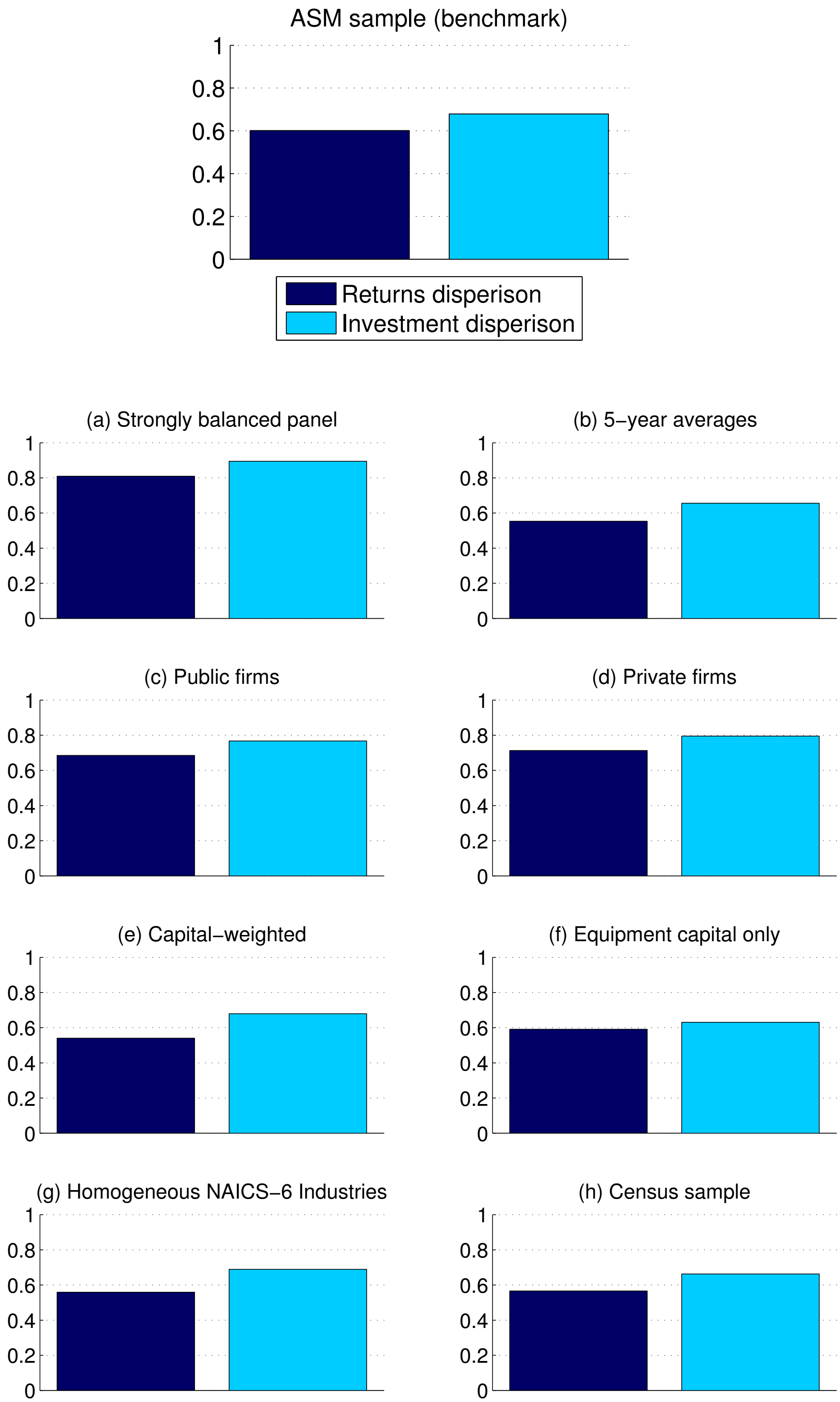
Our main empirical result is that most dispersion in capital returns and capital reallocation originates within firms rather than between firms. In this section, we briefly consider if this result is driven by a particular subset in our benchmark sample or could just reflect measurement error. Details for these robustness exercises can be found in Appendices A.3.2 and A.3.3. At this point, we give a concise overview of the main robustness checks we consider and whose results are displayed in Figure 1.

Note the baseline result of within-firm versus between-firm dispersion of both capital returns (navy blue on the left) and investment rates (light blue on the right) in the top panel of Figure 1. Our first robustness check rules out our result being driven by entry, exit or life cycle dynamics. We therefore redo the decomposition on a strongly balanced panel, but as seen in Panel (a) the share of within-firm dispersion is even larger at 81% and 90%, respectively.

Second, we want to rule out that temporary plant-level noise artificially increases the within-firm share of dispersion. In firms with many plants this plant-level noise would disappear at the firm level. The between-firm variance would then not be biased a lot while the within-firm variance would be biased upward. It is therefore conceivable that between-firm dispersion is more precisely measured as within-firm dispersion which might reflect a considerable portion of high-frequency noise. To rule out that effect, we construct rolling 5-year windows of average capital returns and investment rates for each plant. This time aggregation should equally filter out high-frequency noise at the plant level. Redoing the between-firm/within-firm decomposition of this sample of five-year window shows that the importance of the within-firm share of dispersion persists. As shown in Panel (b) of Figure 1, the within-firm share of overall dispersion in an industry are 55% and 66%, respectively. This suggests that plant-level noise does play some role in inflating the within-firm share, but it does not make up more than a small five percentage point difference.

In Panels (c) and (d) of Figure 1 we carry out the decomposition separately for publicly traded

Figure 1: Within-firm share of overall dispersion $\frac{\sigma^W}{\sigma^W+\sigma^B}$



and privately held firms. While within-firm dispersion dominates in both subsamples, private firms appear to be characterized by a larger share of within-firm dispersion of 71% and 80%, respectively. This points to differential access to finance as a potential source of higher dispersion.

Our next robustness checks revolve around economic relevance. Instead of using the unweighted decomposition, we consider capital weights for the ω's in equation (12) and redo the decomposition. Panel (e) shows the within-firm share of dispersion which declines slightly, but remains still dominant at 54% and 68%, respectively. In Panel (f), we display the results from focusing only on equipment capital which can more easily reallocated across production units than structure capital. Again, the results of the unweighted between-firm/within-firm decomposition are almost unchanged at 59% and 63%, respectively.

Even though our benchmark definition of an industry is fairly fine at the level of the 4-digit NAICS industry, heterogeneous products could potentially lead to spurious differences of capital returns that merely reflect changing product composition within industries. Instead of just defining industries finer at the 6-digit NAICS level, we go one step further and limit attention to those 6-digit NAICS industries considered in Foster et al. (2008) which produce an almost perfectly homogeneous good such as cement, sugar, coffee beans etc.[3] Naturally, we expect the within-firm share of dispersion to be smaller because many firms have operations in several 6-digit industries within the same 4-digit industry. But even in these homogeneous and fine industries, the within-firm share of dispersion in capital returns and investment rates displayed in Panel (g) amounts to 56% and 69% respectively.

Lastly, we only consider the Census years which allow us to focus on the full sample of manufacturing plants in the economy. This means we can consider many small firms with a low within-firm dispersion which are not sampled every year in the ASM. Panel (h) shows that the within-firm share or dispersion is slightly lower but still remains dominant at 57% and 66%, respectively.

Our robustness checks show that the dominance of the within-firm share of dispersion in both capital returns and investment rates does not appear to be an artifact of life-cycle dynamics, high-frequency measurement error at the plant level, the predominance of multi-plant firms with little capital, the type of capital heterogeneous products or the sampling of the ASM. In Appendix A.3.3 we rule out our results driven by further variants of measurement error.

3 A model of the multi-plant firm

In this section, we describe, solve, simulate and analyze a simple model of a firm comprised of several plants. We study how various plant- and firm-level frictions interact with the optimal allocation of capital by the firm across its plants. The interdependency of a plant's capital allocation varies. At

[3] These industries comprise Sugar (31131), Bakery products (311812), Coffee (31192), Block ice (312113), Plywood (321211), Corrugated Boxes (322211), Gasoline (32411), Carbon Black (32518), Cement (32731), Concrete (32732).

one extreme, the firm is a collection of disconnected plants: decisions are made on a plant-by-plant basis, without any interactions between them. We show that in the presence of frictions at the firm level, the firm alters the size and timing of plant-level investment plans which may result in dispersed capital returns across its plants. At the other extreme, firm-level frictions may induce the firm to treat several perfectly identical plants differently. After describing and solving the problem of the multi-plant firm, we will examine how much dispersion in capital returns a calibrated version of our model can generate and how much capital returns dispersion remains unexplained.

3.1 The problem of the firm

We focus on the basic problem of a firm that operates two plants n, $n = A, B$. We limit our model to only two plants in an effort to keep the numerical analysis of our model, which features non-convex policies, computationally feasible. A larger number of plants would increase the state vector of a firm which contains the capital stock and the technology level of each plant; in addition to that, any interactions between plants within the same firm would increase exponentially in the number of plants thus complicating the analysis without adding insight in the underlying fundamental economic mechanisms. Our choice to limit attention to two-plant firms allows us to consider the interactions of several plants within the organizational unit of the same firm while minimizing the computational burden. We first consider the technology and constraints at the level of the individual plant before extending the analysis to the level of the firm. In the following, lower case letters refer to plant variables, upper case letters to firm variables and bold upper case letters refer to vectors of a firm's plant variables. For better readability, we omit time subscripts and denote variables in the next period by a prime.

3.1.1 Technology and frictions at the plant level

Each plant n is characterized by total factor productivity which may include a firm-specific component that is common to both plants. The plant operates a Cobb-Douglas production function which 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. This means we can substitute out any static first-order condition for variable inputs and write plant revenues net of variable factor costs as

$$y_{nt} = z_{nt} k_{nt}^{\alpha}. \tag{3}$$

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

$$z_{njt} = \rho^p z_{njt-1} + \eta_{njt}$$
$$z_{jt} = \rho^f z_{jt-1} + \eta_{jt}.$$

with $\mathbb{E}\eta_{njt} = \mathbb{E}\eta_{jt} = 0$, $V(\eta_{njt}) = (\sigma^f)^2$ and $V(\eta_{jt}) = (\sigma^f)^2$.

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

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

As documented in a number of studies (see Cooper and Haltiwanger (1993), Cooper et al. (1999), Doms and Dunne (1998), 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."[4] 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 foregone 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 in Eberly, Rebelo and Vincent (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.[5] The parameter γ below captures the importance of this margin.

To summarize, frictions at the plant level will be expressed as:

$$\theta\left(i_{nt}, k_{nt}\right) = \left[\psi \mathbb{I}\left\{\frac{i_{nt}}{k_{nt}} > \vartheta\right\} + \gamma\left(\frac{i_{nt}}{k_{nt}}\right)^2\right] k_{nt} \tag{4}$$

where $\mathbb{I}$ is an indicator function equal to 1 if the plant investment rate is above ϑ; ψ is a parameter regulating the foregone 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 (3) and (4) above, plant cash flow is

$$\pi_{nt} = z_{nt}k_{nt}^{\alpha} - \theta(i_{nt}, k_{nt}). \tag{5}$$

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

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

3.1.2 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? This section discusses the allocation of funds within the firm.

A firm collects the cash flow from all plants and decides on how to allocate funds across its plants for investment projects. This means firm cash flow is

$$\Pi_t = \pi_{At} + \pi_{Bt} - \Phi. \tag{6}$$

Managing firm-wide operations requires overhead inputs such as firm management which every firm needs to pay in order to be functional. This overhead fixed cost, denoted by Φ, reflects items such as management costs, overall firm infrastructure, expenditures for R&D, marketing etc. Hence, compared to single-unit firms, multi-unit firms benefit from operating two plants with the same corporate overhead.

Second, while all production and investment activities take place at the level of the individual plant, only the firm is capable of organizing external finance. This assumption is realistic and sensible: while large and complex firms like General Electric operate hundreds of plants, only the firm issues bonds, borrows from banks or raises equity. Typically, the firms then allocate these funds in an internal capital market to individual plants. Consistent with this empirical pattern, we assume that it is the firm that co-ordinates investment plans across all its plants, organizes financing of investment through either internal cash flow or external borrowing and allocates funds to plants where investment is put in place. Only if desired firm-wide investment exceeds firm cash flow, the firm attempts to borrow amount B_{t+1} at net interest rate R_t so that all investment gets financed:

$$i_{At} + i_{Bt} = I_t \leq \Pi_t + B_{t+1}. \tag{7}$$

Organizing external financing, however, is an imperfect process. Following the literature on financial frictions, we assume there are two types of frictions: First, there is a financial participation cost ζK_t if the firm wants to borrow at all. This cost reflects the effort to establish a relationship with a lender, expenditures for various information and disclosure requirements and other administrative expenses independent of the loan amount. Like the real investment adjustment cost above, this cost is scaled by the capital stock. Second, we assume that the firm can divert a fraction $1/\lambda$ of the loan amount B_{t+1} to its private benefit. This diversion of resources cannot be observed or prevented by the lender. As a consequence, the lender will require collateral which it can seize in case it discovers ex post that the firm did divert funds. We assume that firms pledge a fraction ξ of the value of their capital stock, $K_t = (k_{At} + k_{Bt})$, as collateral. Given the posted collateral, lenders will limit the loan such that the diverted loan amount never exceeds the collateral in order to avoid ex post moral hazard unprofitable for the borrower:

$$\frac{B_{t+1}}{\lambda} \leq \xi K_t. \tag{8}$$

Note that one can think of $\lambda\xi$ as the maximum leverage the lender is willing to tolerate. We summarize the cost of external finance in the following function

$$\Theta(B_{t+1}, K_t) = \begin{cases} 0 & \text{if } B_{t+1} = 0 \\ \zeta K_t & \text{if } 0 < B_{t+1} \leq \xi\lambda K_t \\ \infty & \text{if } B_{t+1} > \xi\lambda K_t. \end{cases} \tag{9}$$

Note that $\Theta(\cdot)$ captures the net present value of the borrowing cost associated with B_{t+1}. The interest rate payments next periods have a net present value of $\beta R_t B_{t+1}$. Since there is no risk in our model, the firm's borrowing rate R_t is equal to the risk-free interest rate which in turn equals the inverse of the discount rate, so the net present value of the linear part of borrowing costs exactly equals B_{t+1}. Of course, this would change once we introduce any risk of bankruptcy along the lines of Townsend (1979) which would drive up R_t above the risk-free rate and make the middle portion of the borrowing cost in equation (9) monotonically increasing in the borrowing amount B_{t+1}.

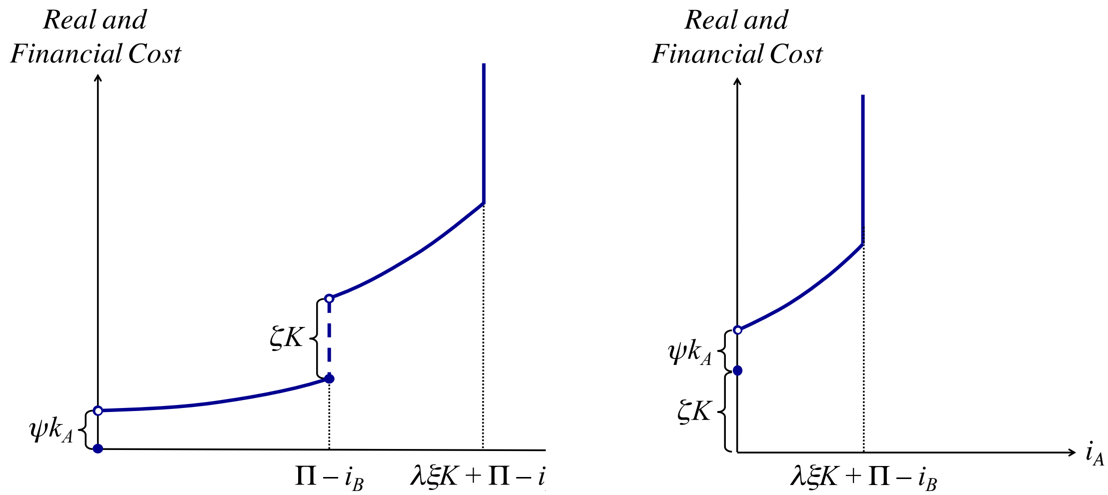
Total cost of investment in a given plant depends on the investment amount in that plant, the combined investment in the rest of the firm, whether or not the firm needs to borrow and if it possibly runs into the collateral constraint. The total cost of investment in plant A then consists of fixed and quadratic adjustment costs (real costs $\theta(i_{At}/k_{At})$ in equation (4)) as well as fixed borrowing costs (financial costs $\Theta(B_{t+1}, K_t)$ in equation (9)). The latter part depends on how much the other plant in the firm, plant B, invests as it dictates how fast and how much the firm needs to borrow. Thus, investment in one plant imposes an externality on investment in the rest of the firm because it depletes internal funds and imposes a borrowing cost that is shared by the entire firm.

We plot the total cost of investment in Figure 2 to illustrate the multiple non-convexities and how the interaction of investment across plants shape 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 $(\psi + \zeta)k_{nt}$.

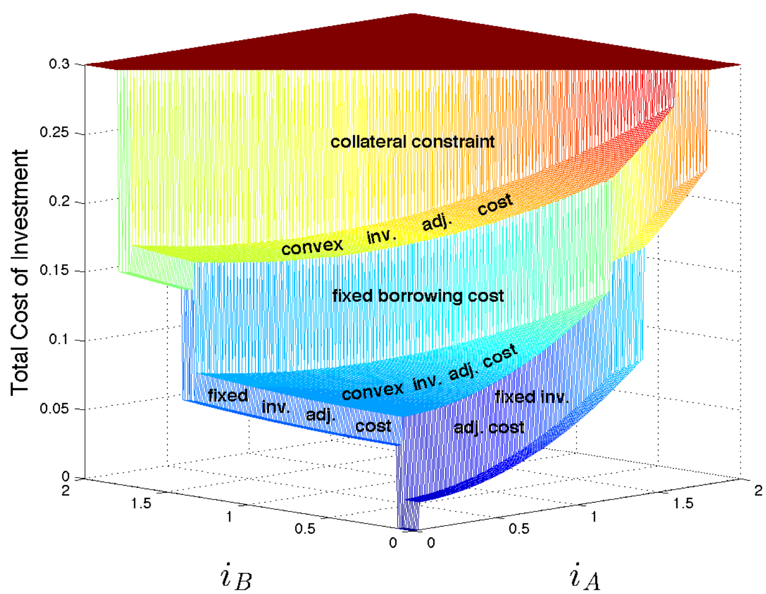
Figure 2: Total cost of investment

(a) Total cost of investing in plant A when i_{Bt} can be fully financed internally

(b) Total cost of investing in plant A when i_{Bt} cannot be financed internally



(c) Total investment costs from the firm's perspective



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 to the fixed borrowing cost 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 collateral constraint: $B_{t+1} \leq \xi\lambda K \Leftrightarrow i_{At} \leq \xi\lambda K + \Pi_t - i_{Bt}$. Panel (c) shows the the total cost jointly for $k_{At} = 1$, $k_{Bt} = 3$, $\zeta = \psi = 0.02$, $\gamma = 0.04$, $\vartheta = 0.03$, $\xi = 0.05$, $\lambda = 2$ and $\Pi_t = 1.4$.

3.1.3 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 will consist 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 and B in order to maximize firm value which consists of the net present value of discounted future gross profits net of investment and borrowing costs. When choosing the investment levels in either plant, the firm takes into account the various adjustment costs and whether or not borrowing is required to finance the desired level of investment. The firm's problem can be written in recursive form as:

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

This value function embeds three non-convexities: the fixed investment adjustment cost, ψ, the fixed borrowing cost, ζ, and the collateral constraint $\xi\lambda$.[6] All three non-convexities give rise to "inaction regions" in the firm's state space where the firm may choose to not change investment and/or borrowing even though the underlying productivity shocks z do. In those regions, the realized capital return will also differ from that of a frictionless investment model. The three non-convex investment/borrowing costs make the solution of the firm problem fairly complex and there are consequentially at most seven different[7] cases of investment-borrowing decisions each of which reflect investment and borrowing (in)activity in the various parts of the firm.

We illustrate the investment policy for plant A and the firm's borrowing policy in Figure 3 for a given amount of investment in plant B assuming that internal funds Π_t are sufficient to finance small amounts of investment. In principle, investment in either plant i_{At} is monotone in its productivity z_{At}. The fixed investment adjustment costs prevent small amounts of investment that would not justify paying the fixed cost ψk_{At}. So investment will not change unless z exceeds a threshold $z^0(k_{At})$. At $z^0(k_{At})$ the expected benefit from investing, $\mathbb{E}\frac{\partial V'}{\partial k'_{At}} i_{At}$, equals the fixed costs, ψk_{At}. Because the latter are modeled proportional to k_{At} and because of decreasing returns to scale, the first threshold will depend on k_{At}. Note that z^0 does not depend on firm variables. Investment

[6] We refer to these frictions as non-convexities even though the last one is strictly speaking only a non-differentiability; but it has similar results for the dispersion of capital returns within the firm as the non-convexities.

[7] With three binary decisions, there are eight cases, but when neither plant invests the firm would never borrow thus making one case redundant. We restrict the investment fixed adjustment cost ψ to be small. If it were extremely high, even the minimum investment which justifies paying the fixed cost would require firm borrowing and there would only be four cases. Then, there is no role for an internal capital market to finance investment because it never suffices even for the minimum investment in one of the two plants.

jumps and then increases monotonically in productivity. The quadratic investment adjustment cost regulated by γ increase the marginal cost of investing but do not break the monotonicity. Plant investment increases until the firm's internal funds are exhausted at $z^1_{At}(k_{At}; z_{Bt}, k_{Bt})$. Notice that this second threshold of productivity for plant A depends on investment in the rest of the firm, i.e. i_{Bt}, and vice versa. This is because investment expenditures in the other plants entail a negative externality for plant A as it depletes internal funds of the firm.

As with the fixed adjustment cost at the plant level, the fixed borrowing costs, ζK_t, prevent the firm from borrowing small external funds. This means there is another region of productivities $[z^1_{At}(\cdot), z^2_{At}(\cdot)]$ where investment does not respond to productivity. Only at productivities above $z^2_{At}(z_{At}, k_{At}; z_{Bt}, k_{Bt})$ is it worthwhile for the firm to pay the fixed borrowing cost. This third threshold depends on investment needs in the entire firm and hence depends on the levels and the distribution of capital stocks and the productivity shocks. Investment again increases monotonically in productivity above $z^2_{At}(\cdot)$ until productivity threshold $z^3_{At}(z_{At}, k_{At}; z_{Bt}, k_{Bt})$ at which point the firm's borrowing demands are limited by the collateral constraint.

3.2 The effects of financial constraints on capital allocation and capital returns

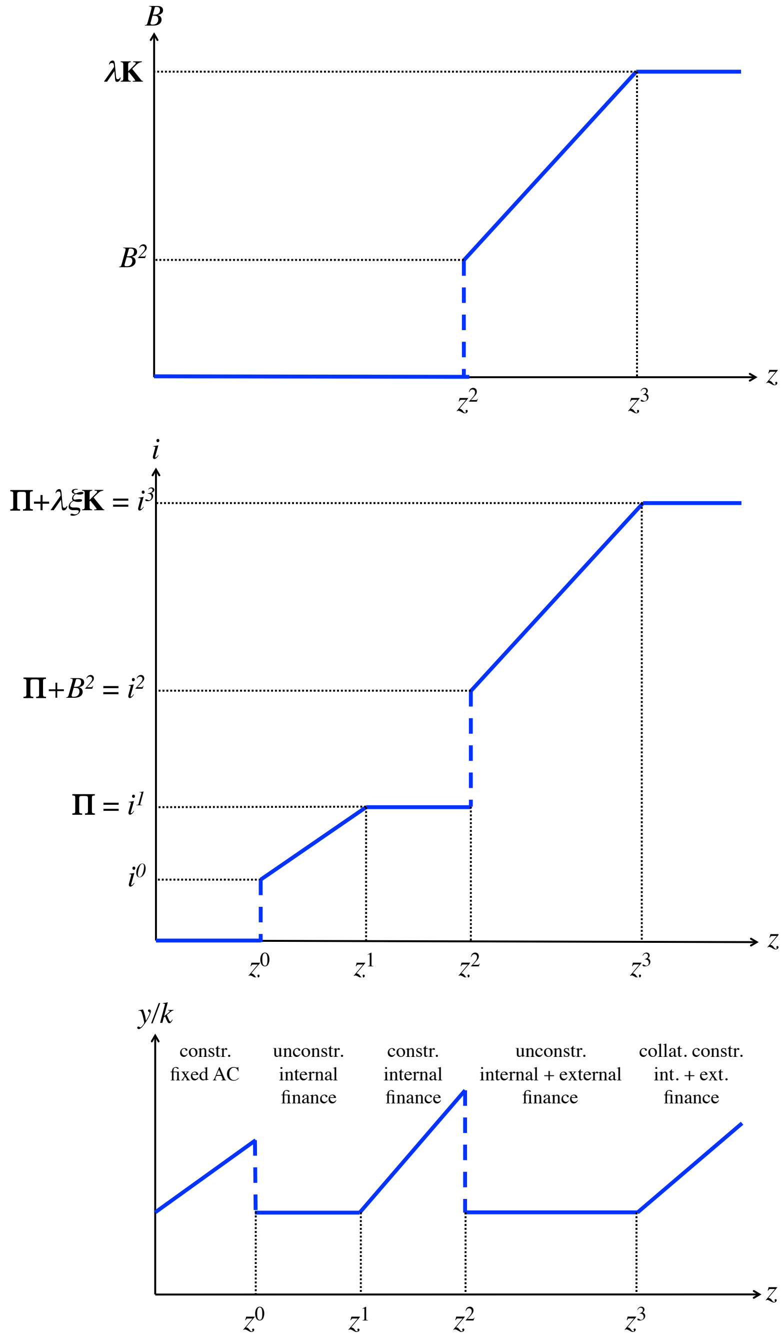
In the previous section, we described how investment at the individual plant is affected by firm-level frictions. Understanding these micro-level dynamics is helpful to gain intuition about the *joint dynamics* of capital returns and capital reallocation within the firm, which is central to the model. We will now describe the effect of each firm-level financial constraint on the dispersion of returns and investment rates and further moments that can be empirically checked in the data. We will explain how interactions between the plants within the firm shape the investment and borrowing policy qualitatively and then assess these patterns quantitatively.

For the quantitative analysis, we solve the model using a value function iteration procedure which is described in detail in Appendix B. Once we obtain the optimal investment and borrowing policy, we simulate a panel of 1,000 two-plant firms for $1,000$ periods and study the resulting within-firm and between-firm dynamics of investment and capital returns. We repeat this exercise several times changing the calibration of the various investment and financing frictions to assess their quantitative effects.

3.2.1 Benchmark model

The model with only quadratic investment adjustment costs is our benchmark. We choose this because comparing the effect of frictions to a completely frictionless model would make many moments such as capital returns dispersion zero thus rendering quantitative comparisons useless. We exclude fixed investment costs from the benchmark because this non-convexity will itself be a significant and known source of returns dispersion which we want to compare to the new firm-level

Figure 3: Firm borrowing and plant investment



Note: Schematic illustration of firm borrowing, plant investment policy functions and realized capital returns. This illustration assumes that the other plant does not invest at all and that physical adjustment costs ψ are smaller than fixed borrowing costs ζ, so that firms start investing without having to borrow; otherwise z^0 and z^2 would coincide.

frictions.

3.2.2 Effects of fixed borrowing costs: Less frequent borrowing, lumpier investment

The fixed borrowing costs imply that the firm will only borrow amounts large enough to justify paying the fixed costs. The firm tries to smooth these costs over time by "saving", which in our model happens through capital accumulation. More capital makes it less likely future borrowing will be needed as often, which in turn reduces the incidence of the fixed borrowing costs. This results in prolonged periods during which the firm will not borrow and finance all – if any – investment internally. Realized capital returns then fluctuate with productivity shocks and will start to diverge as the firm refrains from borrowing and investing in order to align capital returns. As a consequence, within-firm dispersion of capital returns may reflect the fixed borrowing costs faced by the firm rather than misallocation. Below, we will study the quantitative impact of this firm friction on within-firm dispersion.

Our model makes further predictions that we can utilize to study the effects of fixed borrowing costs. First, fixed borrowing costs lead to investment spikes that are synchronized across plants within the firm. Since the firm only borrows rarely, it will then find it optimal to borrow large amounts and pay the fixed investment adjustment costs in both plants. In those periods of firm-wide investment, the correlation between investment and capital returns is high, while the opposite is true during the long periods of internally-financed small investment projects. The latter tend to dominate, leading to lower overall correlation between investment and capital returns. Finally, because fixed borrowing costs make investment at the plant level even lumpier, the autocorrelation becomes even more negative and so does investment at the firm level.

3.2.3 Effects of collateral constraints: Staggered investment across plants

Collateral constraints also limit firm access to external finance. Unlike borrowing costs, however, they prevent borrowing large sums and do not discourage borrowing small amounts: in that sense, they have an effect opposite to that of fixed borrowing costs. The firm will only run into that constraint if it wants to borrow large sums, which happens when both plants receive large positive technology shocks. The firm will try to smooth out this type of borrowing constraint by staying away from it which means to borrow more often and smaller amounts. This constraint is particularly binding when both plants receive a technology shock. Because it faces a borrowing limit, the firm utilizes its internal capital market to finance capital projects in one plant at a time. In other words, it rotates investment plans across its plants. Seen over time, investment across the individual plants within a firm looks staggered.

Because the collateral constraint makes the firm borrow more often and smaller amounts, the correlation between investment and capital returns increases. Perhaps most interestingly, the correlation of investment within the firm drops sharply to negative reflecting the staggering pattern of investment across a firm's plants. Consistent with that, the firm avoids large investment spikes in its plants at the same time.

4 Quantitative analysis

4.1 Calibration

Table 3 summarizes the parameter values used 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. Of note, the fixed investment adjustment costs are lower than those estimated in Cooper and Haltiwanger (2006), Asker et al. (2014). This is because the fixed borrowing costs ζ reinforce the effects of fixed investment adjustment costs ψ. As a result, estimating a model that does not take into account firm structure will confound the two and bias the inferred value for ψ upwards. In addition, when we use the value from Cooper and Haltiwanger (2006), we find that the minimum investment projects needed to justify the fixed adjustment costs are so large that they usually require borrowing. While this case may be theoretically possible, it seems not very plausible that even minimal investment cannot be financed out of the combined firm's internal funds.

Given the lack of hard evidence to support the calibration of the borrowing fixed cost ζ, we remain conservative and pick a small value equal to only 0.01% of the firm capital stock. Also, we start with no firm fixed production cost, which mostly acts as a scale factor anyway.

Table 3: Model Calibration

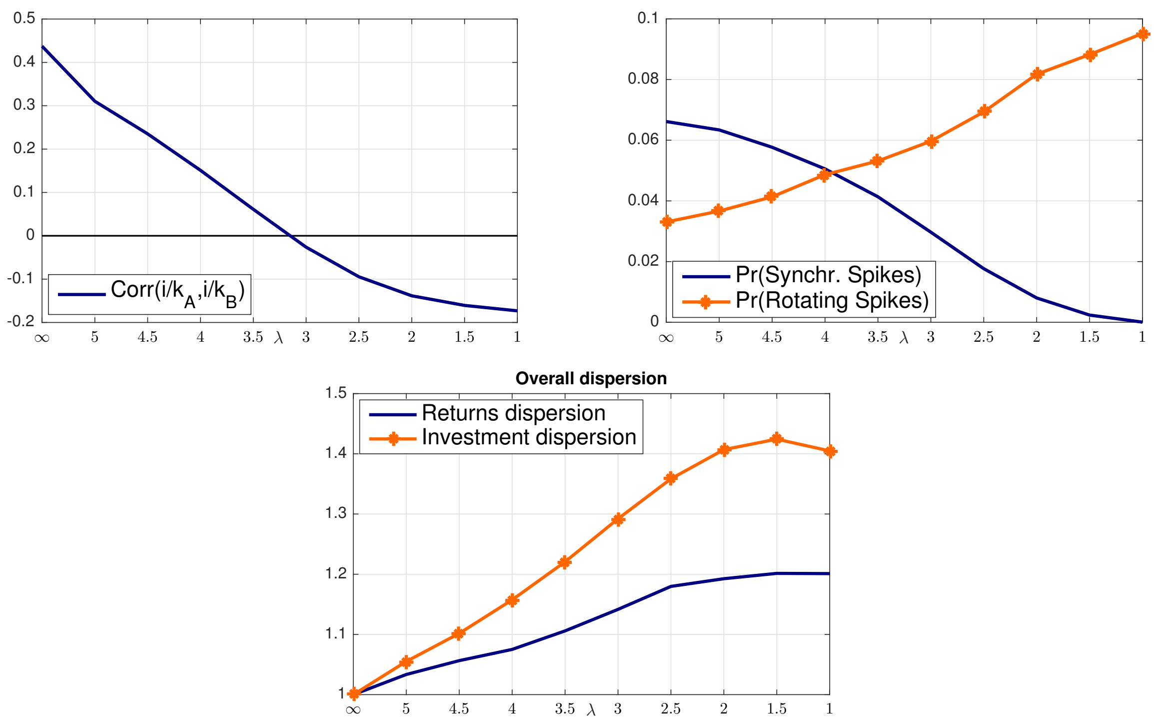
Parameter	Meaning	Value	Target/Source
β	Discount rate	0.95	Long-run real interest rate
α	Production elasticity	0.636	ASM data
δ	Depreciation rate	0.089	Mean from BLS Mfg capital tables
ψ	Fixed inv. adj. cost	0.039	Cooper and Haltiwanger (2006)
γ	Quadratic inv. adj. cost	0.049	Cooper and Haltiwanger (2006)
ζ	Fixed borrowing cost	10^{-4}	See text
ξ	Size of collateral	0.1	10% of capital can be posted as collateral
λ	Collateral constraint	2.5	Borrowing firm can divert 40% of loan
Φ	Firm fixed cost	0	See text
ρ^p	TFP persistence plant	0.6	serial correlation of $\log(y/k)_p$: 0.25
ρ^f	TFP persistence firm	0.85	serial correlation of $\log(y/k)_f$: 0.31
σ^p	TFP shock plant	0.25	volatility of $\log(y/k)_p$: 0.33
σ^f	TFP shock firm	0.24	volatility of $\log(y/k)_f$: 0.26

4.2 Financial constraints, investment activity and dispersion

In order to better understand the behavior of the firm, we now turn our attention to the model simulations. We focus on the impact of external financial constraints on various moments to get a better sense of the behavior of the multi-plant firm.

From the first panel of Figure 4, we can see where the increase in dispersion comes from: as the collateral constraint limits the ability of the firm to borrow, the correlation in investment activity across plants within the firm drops dramatically from its benchmark value of 0.42 in the unconstrained case. In fact, for low-enough values of λ, the cross-plant investment correlation turns *negative*, reaching a trough of -0.19 with our parameterization.

Figure 4: Effect of external financial constraints on the multi-plant firm economy



The second panel illustrates the mechanics behind the drop in correlation. Here, we plot two additional moments. The first one represents the probability of observing synchronized investment spikes, i.e. investment rates above 15% in both plants at the same time. As the collateral constraint is tightened and access to internal finance is limited, the firm cannot allocate large amounts of capital to both plants at the same time, even following a positive firm-level shock. Conceivably, an option for the firm would be to invest smaller amounts in both firms to restrain how much it needs to borrow on capital markets. But because of the presence of investment fixed costs, the firm would rather instead opt for desynchronization of investment activity across its plants. That is, in the event of large positive firm-level shock, the firm allocates capital first to the plant with the highest (expected) return on capital, then does the same for the second plant in the following period. This is why the frequency of rotating spikes, i.e. an investment spike in one plant followed by a spike in the other plant, rises as external financing becomes more limited.

Third, we plot in the the top panel of Figure 4 the relationship between the tightness of the collateral constraint, determined by the parameter λ, and the degree of dispersion of logged expected capital returns ($\log(y_{nt}/k_{nt})$) and investment (i_{nt}/k_{nt}) within the firm. Both moments are normalized to 1 in the scenario where the collateral constraint is not binding ($\lambda = \infty$). As the constraint is tightened (λ gets smaller), dispersion within the typical firm rises sharply, increasing by more than 40% for investment and 20% for capital returns once $\lambda = 2.5$ (firm cannot borrow more than 25% its capital stock). The increase in dispersion eventually slows down, and dispersion can even fall as the constraint becomes so binding that the firm rarely invests at all.

4.3 Aggregate implications: the social value of multi-plant firms

In the previous section, we studied how a multi-unit firm optimally allocates capital across its plants when it is faced with limited access to external financing in addition to the traditional frictions found in the literature. We found that the firm tends to stagger investment activity across its plants to limit borrowing activity. As a result, dispersion of investment but also capital returns within the firm rises.

In what follows, we ask ourselves a different but related question: could the rise in dispersion observed within the firm be socially optimal, i.e. associated with higher aggregate output? To investigate this issue, we simulate two almost identical economies, but for the presence of internal capital markets. The first economy is similar to the exercise we have done so far: we simulate a panel of 1,000 two-plant firms for 1,000 periods, where each firm can pool capital across its plants to post as collateral. The second economy also consists of 1,000 two-plant firms subjected to the same shocks. This time, however, each plant within the firm is considered to be a standalone unit: it must borrow on its own by posting its specific capital as collateral. Besides that, it faces exactly the same constraints and frictions as the plants in multi-unit firms. In other words, we create a "wall" between the plants that bars them from pooling resources when comes the time to access external financial markets. This exercise allows us to specifically isolate the role of the firm in creating internal capital markets and allocating new investment across its plants.[8]

Table 4 compares a number of moments across the two economies under the presence of collateral constraints ($\lambda = 2.5$), as well as the case in which such constraints are absent ($\lambda = \infty$). In this latter scenario, the single-plant and multi-plant firm economies are by definition identical. In line with our previous findings, we find that the presence of financial frictions raises dispersion in a

[8]We leave to future work the simulation of a full general equilibrium model with an endogenous interest rate.

world of multi-unit firms. With standalone plants, however, dispersion tends to fall relative to the unconstrained benchmark. Similarly, the correlation of investment across plants turns negative within multi-unit firms, yet increases somewhat once the plants cannot pool resources and need to borrow on their own.

The bottom panel of the table pertains to the aggregates across our economies: we sum up output, investment and capital across all firms for each period, and take the time-series average. One can notice a striking result: the multi-plant firm is capable of undoing most if not all the effets of financial frictions, with aggregate variables remaining almost unchanged relative to the unconstrained benchmark. This is true even if the presence of collateral constraints *raises* dispersion of capital returns. On the other hand, the same frictions imposed on an economy populated by firms comprised of standalone plants exhibits lower aggregate output *and* lower dispersion.

Table 4: Comparing multi-plant (MUF) and single-plant (SUF) firms

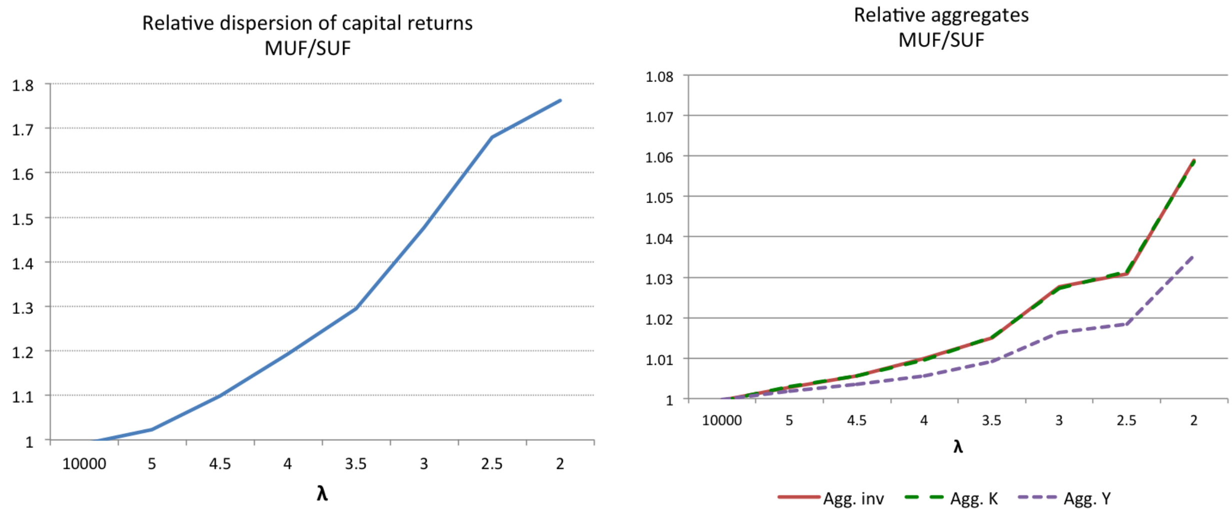
Moment	No frictions	$\lambda = 2.5$	
	(MUF = SUF)	MUF	SUF
1. Plant-level moments			
Share of within-firm dispersion of $E[\log(y/k)]$	0.150	0.193	0.123
Share of within-firm dispersion of i/k	0.298	0.615	0.353
$Corr(i/k_{At}, i/k_{Bt})$	0.437	-0.095	0.320
2. Aggregate Moments			
Aggr I	0.474	0.471	0.457
Aggr K	0.534	0.530	0.514
Aggr Y	0.130	0.130	0.127
Aggr V	314.2	312.7	307.7

We investigate this finding further by simulating the impact of a tightening of the collateral constraint in our two economies. The left panel of 5 shows the relative dispersion of capital returns as λ decreases, i.e. borrowing becomes more limited. In line with our previous findings, we can see that as the constraint is tightened, the multi-unit firm optimally decides to generate *more* dispersion, even if it could always decide to standalone-plants case by shutting down internal capital markets. This optimal strategy at the firm level translates into higher levels of aggregate investment, capital and output relative to an identical economy but for the fact that plants cannot pool resources to access external finance.

To summarize, the insight from our model is that higher capital returns dispersion can be an outcome of constrained efficient behavior by multi-plant firms. This higher dispersion is not a sign of resource misallocation in the economy, but instead of an optimal decision by the firm to stagger investment activity in order to circumvent financial constraints.

In other words, our findings cast doubt on the notion that dispersion in capital returns neces-

Figure 5: Multi-plant vs. single-plant firms, dispersion and aggregates



Note: Each value corresponds to the ratio of a specific moment in the multi-plant-firms economy (MUF) relative to the same moment for the economy composed of firms that cannot pool capital across plants (SUF). Ratios of average within-firm capital returns dispersion (left panel) and aggregate investment, capital and output (right panel) are plotted for various values of the collateral constraint parameter λ.

sarily indicates inefficiencies.

4.4 Rethinking the gains from eliminating misallocation

- Previous sections showed that measured dispersion reflects both inefficient distortions and constrained efficient behavior of multi-plant firms.
- Developing countries have almost no multi-plant firms while in the U.S. multi-plant firms account for he majority of economic activity.
- Thus, dispersion in developing economies reflects solely distortions while in the U.S. almost a quarter of dispersion can be attributed to constrained-efficient firm behavior (see quantitative analysis in Section 4.3 above).
- Fostering output in developing countries can be done in two ways:
  First, give them multi-plant firms, so these firms can reallocate internally; this would increase dispersion, output and welfare. Difference in returns dispersion now reflects distortions only, not differences in firm complexity; this difference is larger than previously thought.
  Second, reduce inefficient distortions. Because they were previously underestimated, bringing down distortions to U.S. levels comes with larger output and welfare gains.

- Conventional welfare gains exercises only consider the second avenue and underestimated the true dimension of distortions.

- So, there are two additional output and welfare gains here:
  First, gain from giving developing economies a more complex firm structure which facilitates within-firm reallocation and increases output (previously not considered).
  Second, gain which results from bringing down distortions to U.S. levels (previously underestimated).

- We intend to quantify both gains.

5 Collateral constraints in multi-plant firms: empirical evidence

5.1 Firm-level evidence

The previous sections documented a set of new facts and proposed a new theoretical model which challenged the notion that high dispersion in capital returns necessarily reflects misallocation. While our model is consistent with many data facts, we now want to demonstrate its empirical relevance more directly. To do that, we verify the crucial features of our model mechanism in firms which operate exactly two plants as in our model. The crux of the model mechanism was a collateral constraints which limited external finance. The tighter this financing constraint, the more relevant a firm's internal capital market becomes which the firm uses to finance investment in at least a subset of its plants. This meant that the more financially constrained, the more dispersed capital returns, capital reallocation and the more prevalent "rotating investment spikes" as opposed to "synchronized investment spikes." We check for these patterns in two-plant firms in the data. As an empirical proxy for the tightness of the collateral constraint, we choose the amount of the pledgeable collateral which we identify as proportional to the plant's capital stock.

5.2 Evidence for collateral constraints I: Staggering investment

We first check if unconstrained two-plant firms exhibit more synchronized investment spikes in the wake of a firm-level technology shock. As the pledgeable collateral of a firm shrinks, our model would predict that synchronized investment spikes become less common and firms transition towards "rotating investment spikes."

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

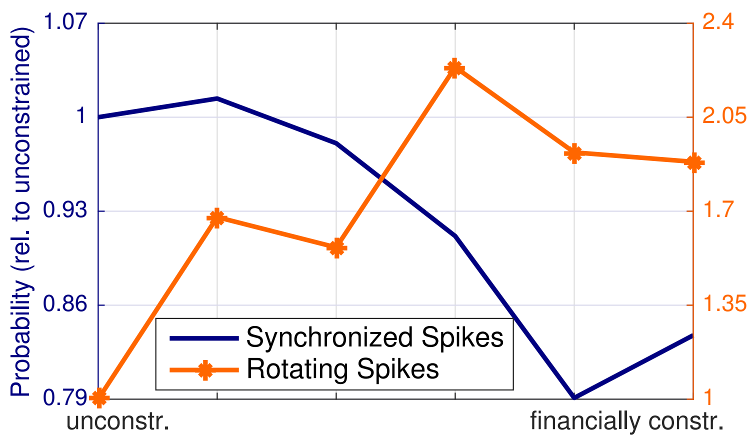
$$Pr(X_{jt}^{sync} = 1|X_{jt}) = \Phi\left(X_{jt}\beta\right) \tag{10}$$

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

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

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 (10) separately for ten deciles of two-plant firms along the value of the pledgeable collateral and evaluate the marginal likelihood of a synchronized investment spike in the wake of a firm-wide technology shock, i.e. a shock experienced by both plants. For disclosure requirements we have to smooth the results of these deciles using a cross-sectional rolling-window average of 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 Figure 6 (solid dark blue line). It shows that the financially most 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 6: Synchronized and rotating investment spikes after firm TFP shocks



In a similar vein, we estimate probit models where we regress a dummy variable if the firm experiences a "rotating investment spike," that is, both plants experience an investment spike in the wake of a firm productivity shock, but just in subsequent periods. We estimate

$$Pr(X_{jt}^{rotate} = 1 | X_{jt}) = \Phi\left(X_{jt}\beta\right) \tag{11}$$

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

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

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 as the starred orange line in Figure 6. In line with the model, the financially most constrained firms are more likely to respond to firm shocks with a rotating investment spike. That is, the firms contemporaneously invests in only one plant *and* invests in the other plant in the subsequent period. The empirical difference of the likelihood of a rotating investment spike increases a lot with our proxy measure of financial constraint: The most constrained firms are twice as likely to revert to rotating investment spikes than financially unconstrained ones. All in all, we view this evidence as strong support for the presence of our model mechanism.

5.3 Evidence for collateral constraints II: Within-firm dispersion

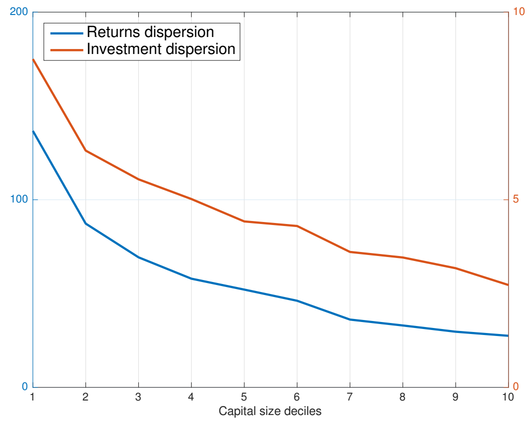
Lastly, we consider the within-firm dispersion in capital returns and capital reallocation. According to our model, firms which are more credit constrained should experience more dispersion while firms with a lot of pledgeable collateral should have an easier time managing to equate returns. We compute the two dispersion measures by decile of pledgeable collateral 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 as in the previous subsection and plot the results in Figure 7.

Indeed, computing the within-firm dispersion measures shows that investment dispersion monotonically declines in our proxy for credit constraints: the most constrained firms have investment dispersion within firms that is about three times as large as that for the least constrained firms. The difference in capital returns is even stronger: the most constrained firms are about six times as dispersed as the least constrained firms.

6 Conclusion

This paper showed that dispersion in capital returns need not indicate distortions. Motivated by the evidence that dispersion mostly occurred within firms rather than across firms, we built a model of a firm operating several plants. Such firms dispose over an internal capital market that helps easing external financial constraints and supports aggregate investment, capital and output. Most importantly, economies with multi-plant firms may well exhibit more dispersion in capital returns

Figure 7: Productivity and investment dispersion (CV) within 2-plant firms



than economies with single-plant firms, but still produce more aggregate output with the same technologies.

References

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

Ufuk Akcigit, Harun Alp, and Michael Peters. Lack of selection and limits to delegation: Firm dynamics in developing countries. *Working Paper*, 2016.

John Asker, Allan Collard-Wexler, and Jan De Loecker. Dynamic inputs and resource (mis)allocation. *Journal of Political Economy*, 122(5), October 2014.

Abhijit V. Banerjee and Benjamin Moll. Why does misallocation persist? *American Economic Journal: Macroeconomics*, 2(1):189–206, January 2010.

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.

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

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

Nicholas Bloom, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen Terry. Really uncertain business cycles. *NBER Working Paper No. 18245*, July 2012.

Richard Blundell and Stephen Bond. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143, August 1998.

Francisco J. Buera and Benjamin Moll. Aggregate implications of a credit crunch: The importance of heterogeneity. *American Economic Journal: Macroeconomics*, 7(3):1–42, July 2015.

Francisco J. Buera and Yongseok Shin. Financial frictions and the persistence of history: A quantitative exploration. *Journal of Political Economy*, 121(2):221–272, April 2013.

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.

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.

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.

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.

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

Gianluca Clementi, Julia K. Thomas, Aubhik Khan, and Berardino Palazzo. Credit shocks and aggregate fluctuations in an economy with production heterogeneity. *Working Paper*, 2011.

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

Russell Cooper and John C. Haltiwanger. On the nature of capital adjustment costs. *Review of Economic Studies*, 73(3):611–633, July 2006.

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.

Russell W. Cooper and Immo Schott. Capital reallocation and aggregate productivity. *NBER Working Paper No. 19715*, December 2013.

Steven J. Davis, John C. Haltiwanger, Lawrence F. Katz, and Robert Topel. Wage dispersion between and within U.S. manufacturing plants, 1963-86. *Brookings Papers on Economic Activity*, 1991:115–200, 1991. URL http://www.jstor.org/stable/2534792.

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

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

Andrea L. Eisfeldt and Tyler Muir. Aggregate issuance and savings waves. *Working Paper*, 2013.

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

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

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.

John G. Fernald and Brent Neiman. Growth accounting with misallocation: Or, doing less with more in Singapore. *American Economic Journal: Macroeconomics*, 3(2):29–74, April 2011.

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.

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

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.

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

Xavier Giroud and Holger M. Müller. Capital and labor reallocation inside firms. *Working Paper*, 2012.

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

João F. Gomes, Amir Yaron, and Lu Zhang. Asset pricing implications of firms' financing constraints. *Review of Financial Studies*, 19(4):1321–1356, Winter 2006.

Gita Gopinath, Şebnem Kalemli-Özcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. Capital allocation and productivity in South Europe. *Working Paper*, 2015.

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

Jan Grobovšek. Managerial delegation, law enforcement, and aggregate productivity. *Working Paper*, 2014.

Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante. Insurance and opportunities: A welfare analysis of labor market risk. *Journal of Monetary Economics*, 55(3):501–525, 2008.

Christopher A. Hennessy and Toni M. Whited. Debt dynamics. *Journal of Finance*, 60(3):1129–1165, July 2005.

Chang-Tai Hsieh and Peter J. Klenow. Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics*, 124(4), November 2009.

Matthias Kehrig. The cyclical nature of the productivity distribution. *Working Paper*, 2015.

Matthias Kehrig and Nicolas Vincent. Financial frictions and investment dynamics in multi-plant firms. *Working Paper*, 2015.

Truman L. Kelley. *Fundamentals of Statistics.* Harvard University Press, Cambridge, Massachusetts, 1947.

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

Aubhik Khan and Julia K. Thomas. Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica*, 76(2):395–436, March 2008.

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

Andrey Malenko. Optimal design of internal capital markets. *Working Paper*, February 2012.

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

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

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

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

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

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.

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

Jae Song, David J. Price, Fatih Guvenen, and Nicholas Bloom. Firming up inequality. *Working Paper*, 2015.

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

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

Robert Townsend. Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, 21(2):265–293, October 1979.

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

A Additional empirical evidence

A.1 Data

We mainly use confidential data on manufacturing establishments collected by the U.S. Census Bureau which 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, Census defines an "establishment" as a business location where the principal activity is production; we hence think of an "establishment" as a production plant. The Census data also contain information about the ownership of each plant (denoted by the variable `FIRMID`) which allows us to construct the ierarchical 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 Census of Manufactures (CMF) and the Annual Survey of Manufactures (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 dynamic implications of our model of investment in multi-plant firms the highest possible frequency. Many aspects of out 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 capital returns and reallocation. Every year, we observe about 55k plants which total up 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 the manufacturing data are merely needed to estimate productivity and the replacement value of capital at current market conditions.[9] 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`[10] in a given year and 4-digit NAICS industry. If the same firm is active in several industries, we define each industry operations as separate firms. Our within-firm dispersion measures are hence an understatement because we ignore the between-industry component if within-firm dispersion.

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

[10]Song et al. (2015) 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.

A.2 Empirics of cross-sectional moments

- Provide details how Table 2, Panel B. was constructed.
- First, we compute cross-plant moments $\mathcal{M}_{it}$ and their standard errors in a given industry i and year t. 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).
- Then, we aggregate across industries using that industry's share in the capital stock:

$$\mathcal{M}_t = \sum_i \omega_{it} \mathcal{M}_{it}$$

  Standard errors are computed according to this aggregation:

$$SE_{\mathcal{M}_t} = \sqrt{\sum_i \left(\omega_{it} SE_{\mathcal{M}_{it}} \right)^2}$$

- Resulting "aggregate" time series of cross-sectional moments reveals:
  - cross-plant standard deviation increases about 10 log points per decade; though the within-firm between-firm split remains unchanged
  - 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 a bottom tail).
- put some time series graphs here a la Gopinath et al. (2015)

A.3 Empirics of between-firm and within-firm moments

In this section, we detail how we compute the within-firm and between-firm dispersion in capital returns and capital reallocation which underly Table 2, Panel A. and the robustness exercises in Section 2.3.

A.3.1 Returns dispersion between and within firms

First, we decompose the overall variance in capital returns 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.

$$
\begin{aligned}
\sigma_t &= \sum_n \omega_{njit} \left(\mathcal{R}_{njit} - \overline{\mathcal{R}}_t\right)^2 \\
&= \underbrace{\sum_i \omega_{it} \left(\mathcal{R}_{it} - \overline{\mathcal{R}}_t\right)^2}_{\sigma_t^{Ind} \text{ between-industry}} + \underbrace{\sum_i \omega_{it} \underbrace{\sum_{j \in i} \omega_{jt}^i \left(\mathcal{R}_{jit} - \mathcal{R}_{it}\right)^2}_{\sigma_{it}^B \text{ between firms within ind. } i}}_{\sigma_t^B \text{ average between-firm}} + \underbrace{\sum_i \omega_{it} \sum_j \omega_{jt}^i \underbrace{\sum_{n \in j}^{N_j} \omega_{nt}^{ji} \left(\mathcal{R}_{njit} - \mathcal{R}_{jit}\right)^2}_{\sigma_{jit}^W \text{ within firm } j \text{ and industry } i}}_{\sigma_t^W \text{ average within-firm}}
\end{aligned} \tag{12}
$$

where n indicates the plant, j the firm, i the 4-digit NAICS industry and t the year. $\mathcal{R}_{njit}$ denotes the capital return of plant n belonging to firm j and industry i in year t, $\mathcal{R}_{jit}$ the average return in firm j in industry i, $\mathcal{R}_{it}$ the average return in industry i, and $\overline{\mathcal{R}}_t$ the average level of returns in the economy.

An industry's level of capital return 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 σ^{Ind} – an object which we ignore for its lack of economic meaning. In our empirical analysis in Section 2, we focus on σ_i^B and σ_i^W only as it is meaningful to compare them and how much of the dispersion in capital returns within an industry originates within firms as opposed to between firms in that same industry: $\mathcal{W}_i \equiv \frac{\sigma_i^W}{\sigma^W i + \sigma_i^B}$. 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 if we are looking at unweighted or capital weighted dispersion.

Although investment rates do not suffer from the industry-specific measurement issues like capital returns, we proceed in a similar way to assess between-firm and within-firm investment rate dispersion.

A.3.2 Robustness

Table 5 provides the underlying details for each sample cut presented in Section 2.3. Equation (12) is applied to each of these subsamples individually to check how much $\frac{\sigma^W}{\sigma^W + \sigma^B}$ changes.

"Full panel" comprises all plants sampled in the ASM 1972-2009 as described in Appendix A.1. "Mid-age" limits attention to plants which are at least three years old and three years away from death. "Balanced panel" refers to plants perpetually alive and perpetually sampled from 1972-2009. The two previous subsamples are aimed at filtering out the dispersion effects of life-cycle dynamics such as strong investment in early stages and divestment/depreciation in the later stages of a plant's life. "5-year averages" computes $1/T_{nt} \sum_{\tau=-2}^{2} \mathcal{R}_{nt+\tau}$ where T_{nt} is the number of periods plant n is observed in the five-year window around period t.

"Homogeneous industries" refers to those 6-digit NAICS industries considered by Foster et al. (2008) which produce almost homogeneous goods. While the shift to 6-digit NAICS industries likely reduces the importance of the within-firm share because firms are now defined to consist of plants at only a 6-digit NAICS industry, this exercise allows us to assess how much our results could be driven by product heterogeneity within 4-digit NAICS industries.

Splitting the sample into firms that are publicly traded and privately held examines the effect

Table 5: Capital returns dispersion within and between firms

Sample	Covered share of		$Var(\log(y/k))$		$Var(i/k)$	
	value added	capital stock	b/w plants within firms	b/w firms within ind.'s	b/w plants within firms	b/w firms within ind.'s
Full panel	0.643	0.724	0.601	0.399	0.679	0.321
Mid-age plants	0.518	0.586	0.595	0.405	0.685	0.315
Balanced panel	0.167	0.219	0.809	0.191	0.895	0.105
5y-averages	0.405	0.472	0.554	0.446	0.656	0.344
Public firms	0.043	0.086	0.685	0.315	0.767	0.233
Private firms	0.511	0.419	0.713	0.287	0.795	0.205
K-weighted	0.126	0.143	0.540	0.460	0.680	0.320
Equipment	0.643	0.724	0.591	0.409	0.631	0.369
Homog. Ind.'s	0.641	0.723	0.559	0.441	0.689	0.311
Census	0.708	0.778	0.566	0.434	0.663	0.337

of access to financial markets and their control matters for the dispersion of capital returns.

"K-weighed" refers to the decomposition of (12) where the ω weights are the capital stock of each plant, i.e. $\omega_{njt} = \frac{k_{njt}}{K_t}$ where K_t is the aggregate capital stock n year t. "Equipment" limits attention to value added per equipment capital only. "Census" finally comprises the cross section of all plants in Census years which includes many more smaller plants not sampled in the ASM but limits attention to the quinquennial frequency.

A.3.3 Measurement error

- Tackled already measurement error at plant level by looking at moving averages over 5-year windows
- Also consider measurement error which is more fixed over time; to tackle this, we consider capital returns which are computed using separate measures of capital and values added.
  - K – use appropriately deflated values of TAB instead of PIM
  - Y – use administrative data on sales from IRS instead TVS from CMF/ASM
  - Y – use collected data on actual production from PCU instead of TVS
- all these should be correlated with the K and Y measures in the ASM (since they measure the same underlying object), but still be different at time due to different coverage or handling by the statistical agency
- we recompute capital returns using the three alternative measures and redo the cross-sectional within-firm between-firm decomposition on these alternative measures.
- If dominance of within-firm share is true, then this should show up in all of these measures.

- Since using these alternative measures limits our sample at times, we also recompute the within-firm between-firm decomposition so we are comparing the moments for the same underlying sample where we have both our benchmark measure as well as the alternative.
- Turns out, 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 the benchmark differs from the alternative which yields an even higher within-firm share. Error bands constructed from averaging across 86 NAICS-4 industries.
- This means 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.

Table 6: Accounting for measurement error

	Alt. Measure	$Corr\left(\log\left(\frac{y}{k}\right)^{bench}, \log\left(\frac{y}{k}\right)^{alt}\right)$	$\left(\frac{\sigma^W}{\sigma^W+\sigma^B}\right)^{bench}$	$\left(\frac{\sigma^W}{\sigma^W+\sigma^B}\right)^{alt}$
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: 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 from the perpetual inventory method). Correlation of the computed capital returns measures are positive, some are high and the within-firm share of overall capital returns dispersion is not statistically different at the 95% 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.

A.3.4 Returns skewness between and within firms

Like the cross-sectional variance, the cross sectional skewness can be decomposed into a skewness component across plants within firms and one between firms. Let g denote the skewness of capital returns:

$$
\begin{aligned}
g &= \frac{\sum_n \omega_{njt}(R_{njt} - \overline{R}_t)^3}{V_t^{3/2}} \\
&= \frac{\sum_j \sum_{n\in j} \omega_{njt}(R_{njt} - R_{jt} + R_{jt} - \overline{R}_t)^3}{V_t^{3/2}} \\
&= \frac{\sum_j \omega_{jt} \sum_{n\in j} \omega^j_{nt}(R_{njt} - R_{jt})^3 + \sum_j \omega_{jt}(R_{jt} - \overline{R}_t) \sum_{n\in j} \omega^j_{nt}(R_{njt} - R_{jt})^2 + \sum_j \omega_{jt}(R_{jt} - \overline{R}_t)^3}{V_t^{3/2}} \\
&= \sum_j \omega_{jt} \left(\frac{V^W_{jt}}{V_t}\right)^{3/2} \frac{\sum_{n\in j} \omega^j_{nt}(R_{njt} - R_{jt})^3}{(V^W_{jt})^{3/2}} + \frac{\sum_j \omega_{jt}(R_{jt} - \overline{R}_t) \sum_{n\in j} \omega^j_{nt}(R_{njt} - R_{jt})^2}{V_t^{3/2}} + \frac{\sum_j \omega_{jt}(R_{jt} - \overline{R}_t)^3}{V_t^{3/2}} \\
&= \underbrace{\sum_j \omega_{jt} \left(\frac{V^W_{jt}}{V_t}\right)^{3/2} g_{jt}}_{\text{average within-firm skewness}} + \underbrace{\sum_j \omega_{jt} \left(\frac{V^W_{jt}}{V_t}\right) \frac{R_{jt} - \overline{R}_t}{V_t^{1/2}}} + \underbrace{\left(\frac{V^B_t}{V_t}\right)^{3/2} g^B}_{\text{between-firm skewness}}
\end{aligned}
$$

where $\omega_{jt} = \sum_{n\in j} \omega_{njt}$ and $\omega^j_{nt} = \omega_{njt}/\omega_{jt}$.

B Model solution

— to be done —

C Estimating technology for multi-plant firms

C.1 Setup

Our quantitative analysis rests on the calibrated values for the production function and technology shocks. Identifying technology shocks is hence a crucial determinant of our quantitative analysis. Firm-level and plant-level shocks regulate how much capital returns dispersion emerge from the volatility of shock processes rather than constraints the firm faces when choosing its internal resource allocation. In this section we describe how we identify parameters α, ρ^f, ρ^p, σ^f and σ^p in the model. We estimate these consistent with the assumptions of the economic model in Section 3. In doing that, we extend the estimation of technology shocks in Cooper and Haltiwanger (2006) to a multi-plant setting. In addition to identifying technology shocks in a setting for multi-plant firms, we also refine their previous empirical work by defining a 3-digit NAICS industry as the common environment for firms. This allows production functions, long-run technology growth and TFP shocks to differ across those industries.

Plants face an inverse demand curve $p = y^{-\nu}$, so sales are $y^{1-\nu}$. Output y is produced using a Cobb-Douglas technology with capital k and other variable inputs which are chosen free of adjustment costs in markets where the firm is price taker. Assuming that the production elasticities of all variable inputs sum to ξ, gross profits of plant n which belongs to firm j at time t can be written (in logs) as

$$\begin{aligned}\pi_{njt} &= \log(1-\xi) + (1-\nu)y_{njt} \\ &= \log(1-\xi) + (1-\nu)\left[\frac{\tilde{\alpha}}{1-\xi}k_{njt} + tfp_{njt}\right].\end{aligned} \tag{13}$$

In contrast to the body of the paper, all variables are logged for ease of reading. We assume that log technology follows a stochastic process with trend growth

$$tfp_{njt} = \tilde{A} + \tilde{g}t + \tilde{A}_t + \tilde{\alpha}_{nj} + \tilde{z}_{njt} + \tilde{z}_{jt}. \tag{14}$$

Trend growth $\tilde{g}$ is common across all units as is the scaling parameter $\tilde{A}$ and the stochastic component $\tilde{A}_t$ which is stationary and fluctuates around zero. Every plant has a fixed technology level $\tilde{\alpha}_{nj}$ which embeds long-run plant profitability. In addition to the aggregate effect $\tilde{A}_t$, the stochastic portion of technology contains a firm component, $\tilde{z}_{jt}$, and an idiosyncratic plant component, $\tilde{z}_{njt}$. The firm component is defined as the firm average, so that z_{njt} denotes a plant's deviation from the firm average technology, that is: $\sum_{n\in j} z_{njt} = 0\ \forall t$. Substituting equation (14) into equation (13) yields

$$\begin{aligned}\pi_{njt} &= \log(1-\xi) + (1-\nu)\left[\frac{\tilde{\alpha}}{1-\xi}k_{njt} + \tilde{A} + \tilde{g}t + \tilde{A}_t + \tilde{\alpha}_{nj} + \tilde{z}_{njt} + \tilde{z}_{jt}\right] \\ &= \alpha k_{njt} + \overline{A} + gt + A_t + \alpha_{nj} + z_{njt} + z_{jt}\end{aligned} \tag{15}$$

where $\alpha = \frac{(1-\nu)\tilde{\alpha}}{1-\xi}$, $\overline{A} = \log(1-\xi) + (1-\nu)\tilde{A}$, $g = (1-\nu)\tilde{g}$ and analogous definitions for A_t, α_{nj}, z_{njt} and z_{jt}. If one abstracts from common and long-run productivity factors, equation (15) corresponds to the logged version of equation (3) in the model section. We can also map this equation into the data: we measure π_{njt} as the log value of sales less all variable input costs (salaries, wages, fringe benefits, materials and energy) deflated by the 5-digit shipment price deflator of the NBER-CES manufacturing database and k_{njt} as the log value of the real stock of structure and equipment capital.

We assume that both plant- and firm-specific components follow AR(1) stochastic processes:

$$\begin{aligned}z_{njt} &= \rho^p z_{njt-1} + \eta_{njt} \\ z_{jt} &= \rho^f z_{jt-1} + \eta_{jt}.\end{aligned}$$

with $\mathbb{E}\eta_{njt} = \mathbb{E}\eta_{jt} = 0$, $V(\eta_{njt}) = (\sigma^f)^2$ and $V(\eta_{jt}) = (\sigma^f)^2$. This persistence in technology processes is quite realistic, but prevents a direct empirical implementation of equation (15). There are several ways to overcome the problem which we detail below.

C.2 Identifying technology: Preferred benchmark

C.2.1 Estimating the production function and firm-level technology shocks

We can achieve identification of all parameters in equation (15) if we assume that the plant fixed productivity effects sum to zero inside firms. This is not a strong a assumption for two reasons: First, all α_{nj} have to sum up to zero – any common long-run component is contained in $\overline{A}$ – which is not very restrictive because "common" means 3-digit NAICS industry here. Second, the median firm in U.S. manufacturing operates 36 plants which samples a considerable number of plant fixed effects to sum to zero. We allow the stochastic processes to be different for firms and plants as highlighted above. To get overcome the problem of autocorrelated error terms in equation (15), we aggregate this equation to the firm level and then take quasi differences:

$$\begin{aligned}
\pi_{jt} &= \alpha k_{jt} + \overline{A} + gt + A_t + z_{jt} \qquad (16)\\
&= \alpha k_{jt} + \overline{A} + gt + A_t + \rho^f z_{jt-1} + \eta_{jt}\\
\pi_{jt} - \rho^f \pi_{jt-1} &= \alpha k_{jt} - \rho^f \alpha k_{jt-1} + (1-\rho^f)\overline{A} + \rho^f g + (1-\rho^f)gt + A_t - \rho^f A_{t-1} + \eta_{nj}. \qquad (17)
\end{aligned}$$

where $\pi_{jt} = \frac{1}{N_j}\sum_{n\in j}\pi_{jt}$ is the average log profit across the N_j plants of firm j (analogously for capital). We also used the fact that $\frac{1}{N_j}\sum_{n\in j} z_{njt} = \frac{1}{N_j}\sum_{n\in j}\alpha_{nj} = 0$. Since η_{jt} are iid over time and across firms, we can estimate equation (17). Since it's over identified, we use GMM with the following instruments: lagged and twice lagged values of π_{jt}, current, lagged and twice lagged values of k_{jt} and a complete set of time dummies to capture $(1-\rho^f)gt + A_t - \rho^f A_t$. This way, we do not have to specify a stochastic process for the aggregate technology shock.

C.2.2 Estimating plant-level components of TFP

We identify plant-level technology shocks from equation (15) as follows. Define the following two objects at the plant and firm level respectively:

$$\begin{aligned}
X_{njt} &\equiv \pi_{njt} - \alpha k_{njt} = \overline{A} + gt + A_t + \alpha_{nj} + z_{njt} + z_{jt}\\
X_{jt} &\equiv \pi_{jt} - \alpha k_{jt} = \overline{A} + gt + A_t + z_{jt}.
\end{aligned}$$

We can compute the difference between those two objects and use it to estimate plant-level technology shocks in a fixed effects panel regression:

$$\begin{aligned}
Y_{njt} \equiv X_{njt} - X_{jt} &= \alpha_{nj} + z_{njt}\\
&= \alpha_{nj} + \rho^p z_{njt-1} + \eta_{njt}\\
&= \alpha_{nj} + \rho^p\left[Y_{njt-1} - \alpha_{nj}\right] + \eta_{njt}\\
&= (1-\rho^p)\alpha_{nj} + \rho^p Y_{njt-1} + \eta_{njt} \qquad (18)
\end{aligned}$$

This gives us estimates for the outstanding parameters ρ^p and σ^p which we report alongside the other estimates for the median industry in Table 7.

C.3 Identifying technology: General case

We also present a more general identification than above which proceeds along the lines of Blundell and Bond (1998). In particular, we relax the assumption $\sum_{n\in j}\alpha_{nj} = 0$. This means firms within an industry are allowed to have long-run productivity differences. This added flexibility comes at the cost of potentially weaker identification. For filtering out firm-level fixed effects requires differencing equation (16) and then taking again the same quasi differences as above to overcome

Table 7: Estimates

	Aggregate	Firm level		Plant level	
	α	ρ^f	σ^f	ρ^p	σ^p
(Ia)	0.6268	0.7360	0.8533	0.2806	0.4985
	(0.0028)	(0.0016)	(???)	(0.0121)	(???)
(Ib)	0.6268	0.7136	0.8628	0.3414	0.5939
	(0.0118)	(0.0073)	(???)	(0.0037)	(???)
(Ic)	0.5924	0.6980	0.8407	0.3423	0.5955
	(0.0206)	(0.0131)	(???)	(0.0071)	(???)
(II)	0.6268	0.1627	0.6153	0.2806	0.4985
	(0.0028)	(0.0502)	(???)	(0.0121)	(???)

Note: Results from the GMM estimation of equation (17) and the fixed effects panel regression in equation (18), standard errors in parentheses. Row Ia reports GMM estimates for α and the firm process (ρ_f and σ^f) and FE panel estimates for the plant process (ρ_f and σ^f); row (II) reports GMM estimate of α and FE panel estimates for both the firm and the plant processes.
(Ib) is like (Ia) except that all parameters are estimated at the NAICS-3 industry level and then averaged using value added industry weights. (Ic) does the whole thing at the NAICS-4 level.
Estimating everything at the 3-digit NAICS industry level and then reporting median industry values is fairly similar.

the persistence problem.

$$\begin{aligned}
\Delta\pi_{jt} &= \alpha\Delta k_{jt} + g + \Delta A_t + \Delta z_{jt} \\
&= \alpha\Delta k_{jt} + g + \Delta A_t + \rho^f \Delta z_{jt-1} + \Delta\eta_{jt} \\
&= \alpha\Delta k_{jt} + g + \Delta A_t + \rho^f \left[\Delta\pi_{jt-1} - \alpha\Delta k_{jt-1} - g - \Delta A_{t-1}\right] + \Delta\eta_{jt} \\
\Delta\pi_{jt} - \rho^f \Delta\pi_{jt-1} &= \alpha\Delta k_{jt} - \rho^f \alpha\Delta k_{jt-1} + (1-\rho^f)g + \Delta A_t - \rho^f \Delta A_{t-1} + \Delta\eta_{jt}. \qquad (19)
\end{aligned}$$

As in our benchmark case, we estimate equation (19) with GMM. As instruments, we choose lagged values of gross profits and current and lagged values of capital as above. But in this setup, where we need to instrument for growth rates rather than levels, these instruments are notoriously weak. This weakness manifests itself in estimates which are quite different from the accepted standard in the literature (see Cooper and Haltiwanger (2006) or Kehrig (2015)), so we are not confident they make a solid calibration target for your quantitative work. We get $\hat{\alpha} = 0.17$, $\hat{\rho}^f = -0.15$. In principle, one could follow the subsequent procedures outlined above to obtain estimates for the aggregate and plant-level components of technology.

C.4 Identifying technology: Previous standard

Previous work that estimated technology shocks in the context of investment models ignored the firm dimension. This implicitly assumed the same persistence of plant and firm technology and also assumed no plant-level fixed effects. If we were to follow that procedure, the same persistence allows us to write $\zeta_{njt} = z_{njt} + z_{jt} = \rho\zeta_{njt-1} + \eta^{\zeta}_{njt}$. Then, we taking quasi differences as above:

$$\begin{aligned}
\pi_{njt} &= \alpha k_{njt} + \overline{A} + gt + A_t + \zeta_{njt} \\
&= \alpha k_{njt} + \overline{A} + gt + A_t + \rho^{\zeta}\zeta_{njt-1} + \eta^{\zeta}_{nj} \\
\pi_{njt} - \rho^{\zeta}\pi_{njt-1} &= \alpha k_{njt} - \rho^{\zeta}\alpha k_{njt-1} + (1-\rho^{\zeta})\overline{A} + \rho^{\zeta}g + (1-\rho^{\zeta})gt + A_t - \rho^{\zeta}A_{t-1} + \eta^{\zeta}_{nj}.
\end{aligned}$$

We estimate this equation by GMM using as current and up to twice lagged values of capital and lagged values of gross profits. In principle, one could follow the subsequent procedures outlined above to obtain estimates for the aggregate and plant-level components of technology.

D How financial constraints shape investment empirically

D.1 Credit Constraints

D.1.1 Introducing time-varying financial frictions

In order to determine whether the predictions of the model with time-varying financial frictions are in line with what we observe in the data, we need some measure of financial conditions that covers a long enough time period. The National Financial Conditions Index (NFCI) from the Federal Reserve Bank of Chicago seems well suited to our purposes. The NFCI is a weighted average of a large number variables of financial activity, relative to their means. The index is therefore centered around zero by construction. We will be using the Adjusted National Financial Conditions Index (ANFCI), a version that isolates the component of financial conditions that is orthogonal to current economic conditions, allowing us to focus solely on the impact of fluctuations in credit tightness without worrying about endogeneity. Figure 8 shows the evolution of both the raw and adjusted NFCI.

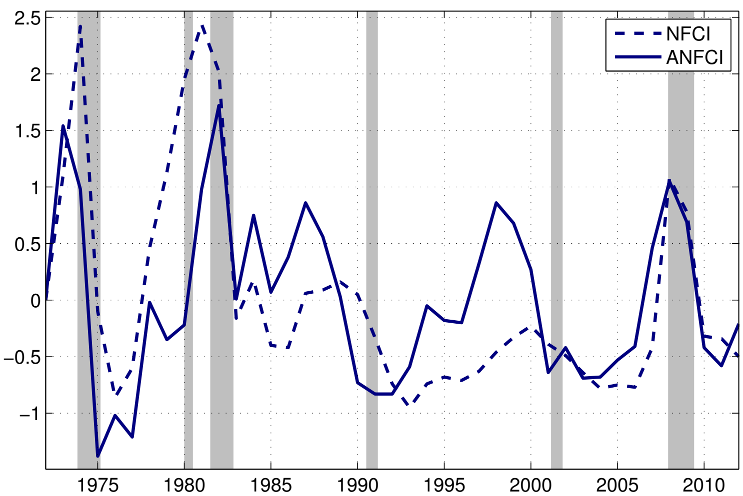
As one can see, the adjusted indicator lines nicely with NBER recessions in the 1970's, 1980's and in 2008/09 while it does not increase before or during the 1991 and the 2001 recession – that financial indicators such as credit spreads did not catch those recessions is a well-known fact and we follow the literature to interpret these as non-financial recessions. In the subsequent analysis, we will use the ANFCI time series as an exogenous variable that constrains a firm's ability to borrow funds for investment purposes.

D.1.2 Empirical joint dynamics in 2-plant firms

In this section we move away from comparative statics by incorporating a two-state process for the financial friction: the firm will alternate between states with low or high degrees of financial frictions. This exercise allows us to run on the simulated data regressions that are similar to the empirical specifications discussed in the next section where we use a time-series index of financial conditions to determine how investment dynamics are affected by changes in financial frictions.

For this exercise, we continue to set the fixed cost of borrowing, ζ, to 1% of the capital stock. This parameter will be time-invariant and can be interpretred for example as the management costs related to preparing a loan application and interacting with the financial intermediary The other borrowing friction parameter, η, is time-varying: it can take the values 0 (low borrowing cost) or 5

Figure 8: Credit Tightness in the U.S. Economy



Note: Annualised time series of the National Financial Conditions Index in the raw version (dashed line) and the version that is adjusted for endogenous responses of financial indicators to non-financial shocks (solid line). Shaded areas are NBER recessions.

(high borrowing cost) depending on the degree of credit tightness, and the probability of switching between the two states is equal to 0.1. All the other parameters of the model are unchanged.

We simulate the model and run regressions on the simulated data. The move to regression analysis is important if we want to link the predictions of the model with the data: while our model allows us to perfectly isolate the role of borrowing constraints, in the real world investment dynamics may be affected by multiple sources of heterogeneity unrelated to financial frictions. The use of controls is therefore crucial.

It should be noted that we are in no way trying to calibrate the size and relative importance of the firm-level financial frictions at this point. Therefore, what we are interested in determining whether the model predictions are in line with what we observe in the data from a qualitative, not quantitative, standpoint.

Table 8 shows results for plant-level investment regressions. Each regression uses the investment-to-capital ratio for plant A at time t as the dependent variable (the two plants are perfectly symmetric in the model), $(i/k)_{At}$. In order to determine the impact of firm-level financial frictions on investment dynamics, we define a dummy variable, ς_t, equal to one if credit is tight in period t (i.e. $\eta = 5$) and zero otherwise.

The first two regressions of Table 8 focus on the role of output/cash flow variables in explaining movements in investment. Not surprisingly, plant-level investment is on average lower in periods of high borrowing costs ($\varsigma_t = 1$). In the top panel we can see that a 10% increase in the output-to-capital ratio of the plant raises its i/k ratio by about 1.3%. Interestingly, plant A's investment is also affected by plant B shocks, to a lesser degree. There are two possible reasons for this result. First, simultaneous increases in output at both plants A and B are potentially indicative of a firm-wide shock. Since firm shocks are more persistent than plant-level shocks, the optimal decision is to invest more.

Second, the internal finance channel is also at play: in a context where funds are scarce and borrowing costly (recall that $\zeta = 0.01K$ in all states), a good shock in plant B generates precious cash flow that can be used to finance investment in plant A. This is also evident in the results for the second regression of the same table. There, plant B's output is replaced by cash flow at the level of the firm, net of adjustment and fixed operating costs (notice that we do not use logs as cash flow is sometimes negative) Not only is plant A investment higher when firm cash flows are higher (conditional on the plant-specific shock), but we can see that this dependence on the firm's financial resources is particularly strong when credit is tight: the coefficient on cf_t/k_t more than doubles when $\varsigma_t = 1$. In other words, the existence of internal capital markets is particularly relevant when external financial constraints are more binding.

The last regression of Table 8 looks at the correlation in investment activity across plants within the firm. In periods where credit is cheap, $(i/k)_{At}$ is a positive function of investment activity in plant B, though the relationship is somewhat weak: a 1 percentage point increase in $(i/k)_{Bt}$ raises $(i/k)_{At}$ by less than 0.1 percentage point. The relationship, however, changes dramatically when credit is tight: in periods where $\eta = 5$, the same 1 p.p. change in investment at plant B leads instead to a *fall* of almost 0.5 percentage point in $(i/k)_{At}$ as investment activity in one plant crowds out investment elsewhere. With scarce funds, the firm selects the most profitable projects, postponing others as their marginal benefit is outweighed by the marginal cost of external funds.

The firm-level regression in Table 8 highlights another impact of financial frictions in multi-unit firms: the serial correlation of firm-level investment increases. As shown earlier in Table 4, our baseline calibration implies that the autocorrelation of i_t/k_t is slightly negative, which should not be too surprising as these firms are very small. However, a financial constraint shock makes firm-level investment significantly smoother. On potential explanation for this result is related to our earlier findings: when credit is tight, it makes the firm less likely to invest in both plants at the same time in order to avoid costly borrowing. To understand why this may lead to higher autocorrelation of investment for the firm, consider the example of a positive firm-level shock. Since both z_{At} and z_{Bt} are now higher, both plants would now like to invest to reach their new optimal level of capital. But given limited cash flows, this implies that the firm would need to obtain costly funds on capital markets. Instead, it will sometimes find it optimal to stagger investment: plant A invests today, while plant B waits until tomorrow. By construction, this makes firm-level investment smoother. We show additional evidence for this kind of behaviour below.

In Table 10 we instead focus on investment spikes, having showed earlier that they were the main contributor to aggregate investment empirically. All regressions are linear probability models. In the first panel, the dependent variable is a dummy equal to 1 if the investment-to-capital ratio in plant A is greater than 15%, our threshold for a spike. Both output levels for plants A and B have a positive impact on the probability of a spike, in line with what we found earlier. Our focus, however, is on the spike indicator for plant B: very clearly, the spike activity in one plant does matter for the probability of a spike in the other. In periods of tight credit ($\varsigma_t = 1$), the probability of observing an investment spike in plant A is 0.5 percentage point lower if plant B is already spiking.

Another way to confirm this finding is to see whether occurrences of double spikes (i.e. spikes in both plants) is more or less likely when external funds are more costly. Conditional on firm cash flow and capital stock (coefficients not reported), the second panel of the same table shows that the probability of observing both plants spiking in the same period basically falls to zero when credit is tight.

Table 8: Plant-level investment regressions

Constant		0.419***
Credit tightness	ς_t	−0.075***
Output plant A	$\log(y/k)_{At}$	0.129***
	$\varsigma_t \cdot \log(y/k)_{At}$	−0.014***
Output plant B	$\log(y/k)_{Bt}$	0.063***
	$\varsigma_t \cdot \log(y/k)_{Bt}$	−0.011**
R^2		0.33
Constant		1.647***
Credit tightness	ς_t	−0.074***
Output plant A	$(y/k)_{At}$	0.584***
Firm cash flow	$(cf/k)_t$	0.246***
	$\varsigma_t \cdot (cf/k)_t$	0.356***
R^2		0.34
Constant		1.882***
Credit tightness	ς_t	0.017***
i/k plant B	$(i/k)_{Bt}$	0.085***
	$\varsigma_t \cdot (i/k)_{Bt}$	−0.545***
Cash flow firm	$(cf/k)_t$	−0.007
	$\varsigma_t \cdot (cf/k)_t$	0.882***
R^2		0.40

Note: Dependent variable is i/k of plant A. $\varsigma_t = 1$ in periods of high borrowing cost, 0 otherwise. Controls such as capital stock or cash flow are included but not always reported. *, ** and *** indicate significance at the 10, 5 and 1 percent level respectively.

Table 9: Firm-level investment regressions

$$(i/k)_t = \beta_0 + \beta_1 \varsigma_t + \beta_2 \log(i/k)_{t-1} + \beta_3 \varsigma_t \cdot \log(i/k)_{t-1} + \varepsilon_t$$

Variable		Estimate
Constant	β_0	0.419^{***}
Credit tightness	β_1	-0.038^{***}
Lagged firm investment	β_2	-0.034^{***}
... interacted with credit tightness	β_3	0.125^{***}
R^2		0.44

Note: Dependent variable is i/k of the firm. $\varsigma_t = 1$ in periods of high borrowing cost, 0 otherwise. Controls such as capital stock and firm cash flow are included but not reported. *, ** and *** indicate significance at the 10, 5 and 1 percent level respectively.

Finally, the last regression in Table 10 revisits the question of staggered investment across plants. We regress the spike indicator for plant A on its own lag, the spike indicator for plant B at time t as well as its lagged value, both by itself and multiplied by the credit dummy, ς_t. Our focus is on this last variable: if the firm is more likely to stagger investment activity when borrowing is costly, the coefficient on the interaction term should be positive. This is what we obtain, with a value of 0.036, though the effect seems to be relatively small.

All these results seem to indicate that the firm-level financial frictions, whether in the form of fixed or quadratic costs, alter significantly plant-level investment dynamics. In summary, the firm optimises the timing of investment spikes, making sure that joint spikes are avoided in order to minimise the need to borrow in a given period. Instead, following a positive firm-level shock affecting both plants, one plant spikes immediately while the other waits one period to invest. This, in turn, makes investment aggregated at the firm level more serially correlated in periods of tight credit. In the next section, we investigate whether these predictions are borne out empirically using a proxy for financial conditions.

D.2 Investment Dynamics of Plants, of Firms and Within Firms

We first focus on the time series properties of investment rates at the plant level, the firm level and the joint investment dynamics of plants within firms. Table 4 is our benchmark.

D.2.1 Autocorrelation of plant-level investment

We start by examine the autocorrelation of investment at the plant and the firm level. Table 4 shows how plant-level investment becomes less and firm-level investment becomes more autocorrelated in the model. The latter is a response of firms where costly external credit induces the firm to smooth borrowing and thus investment. Predictions about the autocorrelation of plant-level investment are not as sharp and that is reflected in the data: Table 11 displays the estimates of regressing plant-level investment on the credit constraints indicator (denoted by ς_t). The estimates from the panel regression (our preferred specification) are not significant. A simple pooled OLS regression indicates that in tight credit times investment becomes more autocorrelated.

Table 10: Spike ($i/k > 0.15$) regressions

	Dependent variable	$\mathbf{I}\{(i/k)_{At} > 0.15\}$
Constant		6.261^{***}
Credit tightness	ς_t	-0.039^{***}
Output plant A	$\log(y/k)_{At}$	0.504^{***}
Output plant B	$\log(y/k)_{Bt}$	0.330^{***}
Spike plant B	$\mathbf{I}\{(i/k)_{Bt} > 0.15\}$	-0.254^{***}
	$\varsigma_t \cdot \mathbf{I}\{(i/k)_{Bt} > 0.15\}$	-0.261^{***}
R^2		0.44
	Dependent variable	$\mathbf{I}\{(i/k)_{At}, (i/k)_{Bt} > 0.15\}$
Constant		0.069^{***}
Credit tightness	ς_t	-0.075^{***}
R^2		0.13
	Dependent variable	$\mathbf{I}\{(i/k)_{At} > 0.15\}$
Constant		6.651^{***}
Credit tightness	ς_t	-0.108^{***}
Spike plant B	$\mathbf{I}\{(i/k)_{Bt} > 0.15\}$	-0.478^{***}
Lag spike plant A	$\mathbf{I}\left\{(i/k)_{At-1} > 0.15\right\}$	-0.048^{***}
Lag spike plant B	$\mathbf{I}\{(i/k)_{Bt} > 0.15\}$	0.061^{**}
	$\varsigma_t \cdot \mathbf{I}\{(i/k)_{Bt} > 0.15\}$	0.036^{**}
R^2		0.48

Note: Dependent variable is spike dummy. $\varsigma_t = 1$ in periods of high borrowing cost, 0 otherwise. *, ** and *** indicate significance at the 10, 5 and 1 percent level respectively.

Table 11: Credit Constraints and Plant- and Firm-Level Investment

left panel (plant): $(i/k)_{nt} = \beta_0 + \beta_1 (i/k)_{nt-1} + \beta_2 \varsigma_t \cdot (i/k)_{nt-1} + \beta_3 \varsigma_t + \varepsilon_{nt}$
right panel (firm): $(i/k)_{jt} = \beta_0 + \beta_1 (i/k)_{jt-1} + \beta_2 \varsigma_t \cdot (i/k)_{jt-1} + \beta_3 \varsigma_t + \varepsilon_{jt}$

Coefficient	OLS	Panel	Coefficient	OLS	Panel
	(0.0471)	(0.0638)		(0.0565)	(0.0614)
β_2	0.0145**	0.0109	β_2	0.0264***	0.0288***
	(0.0069)	(0.0095)		(0.0074)	(0.0075)
β_3	0.0004	0.0017	β_3	0.0003	0.0003
	(0.0008)	(0.0011)		(0.0007)	(0.0007)
Controls	Yes	Yes	Controls	Yes	Yes
N	49k	17k	N	24k	9k

The predictions about firm-level investment are much sharper. Table 11 reveals that investment at the firm level significantly becomes more autocorrelated: A standard deviation to credit tightness almost doubles the autocorrelation. The effect at the firm level is much stronger than any change at the plant level where the autocorrelation increases only by 10%. Interestingly, tight credit itself does not significantly lower investment levels as the estimate of β_3 is not significantly negative as one may expect – and neither it is with the plant regression in Table 11.

D.2.2 Investment Correlation Within Firms

A particularly sharp implication of credit constraints is that investment across the two plants within the firm falls; Table 4 shows in fact that it may even become negative. This obviously reflects the fact that in times of tight credit the firm needs to scale back investment in general. So if one plants invests a lot, the the other one probably suffers when credit is tight.

Table 12 confirms this prediction of the model. The estimates of β_2 are significantly negative across both panel and OLS regressions. The estimate imply that a doubling of investment in the other plant lowers the investment rate in the other plant by two percentage points.

Table 12: Credit Constraints and Within-Firm Investment

$$(i/k)_{At} = \beta_0 + \beta_1 (i/k)_{Bt} + \beta_2 \varsigma_t \cdot (i/k)_{Bt} + \varepsilon_{nt}$$

Coefficient	OLS	Panel
	(0.0017)	(0.0016)
β_2	–0.0035**	–0.0031**
	(0.0014)	(0.0013)
Controls	Yes	Yes
N	66k	66k

D.2.3 Probability of investment spikes and spike size

We look at the likelihood of investment spikes in episodes of tight credit. It's not obvious whether there will be more or less investment spikes when credit is tight. On the one hand, spikes will happen less often because tight credit limits overall investment resources. Then, tight credit results in a "lumpiness" effect because investment looks more lumpy. On the other hand, tight credit makes the firm smooth its borrowing so that investment spikes become more frequent. If that latter outcome prevails, one would expect investment spikes to become smaller; we label this latter effect the "smoothing effect." We test both of these possible predictions and display the results of this regression in Table 13. The results are overall weak and borderline significant. But if at all, one sees that the probability of investment spikes increases and the level of investment spikes decreases significantly.

Table 13: Credit Constraints and Investment Spikes

$$\mathbf{I}\left\{(i/k)_{nt} > 0.15\right\} = \beta_0 + \beta_1 \varsigma_t + \beta_2 \varsigma_t \cdot \mathbf{I}\left\{\text{Public}\right\} + \varepsilon_{nt}$$
$$(i/k)_{nt} = \tilde{\beta}_0 + \tilde{\beta}_1 \varsigma_t + \tilde{\beta}_2 \varsigma_t \cdot \mathbf{I}\left\{\text{Public}\right\} + \varepsilon_{nt} \quad \forall n, \text{ s.t. } (i/k)_{nt} > 0.15$$

Coefficient	OLS	Panel
β_0	0.0411***	0.0519***
	(0.0.131)	(0.0092)
β_1	0.0002	0.0003
	(0.0012)	(0.0011)
β_2	–0.0055	–0.0004
	(0.0067)	(0.0066)
Controls	Yes	Yes
N	66k	66k

Coefficient	OLS	Panel
$\tilde{\beta}_0$		.
$\tilde{\beta}_1$		negative (significant)
$\tilde{\beta}_2$		positive (significant) $\Rightarrow$ publicly traded firms unaffected
Controls	Yes	Yes
N	66k	66k

D.2.4 Joint distribution of investment spikes within firms

While predictions about single investment spikes are not conclusive, the model has fairly strong predictions about the *joint* distribution of investment spikes within firms. When credit is tight, a firm cannot allow both of its plants to undergo an investment spike in the same period – at least not if it's financially constrained. As a consequence, the likelihood to see both plants undergoing an investment spike should drop significantly when credit is dear. This is clearly borne out in the data as Table 4 shows. An it does show up in the data as well: When credit gets tight, the likelihood to undergo an investment spike when the other plant undergoes one is 2% lower than when credit is loose. In that latter scenario, we can expect that one plant spiking raises the likelihood of the other one spiking by 18% – probably reflecting a positive firm-specific productivity shock.

Note that this logic of reduced simultaneous spiker plants only applies to financially constrained firms. If a firm wasn't financially constrained, we would expect the coefficient β_2 to be zero. To test for that hypothesis, we include an interaction term of credit tightness and the other plant spiking with a dummy variable that indicates whether or not the firm is publicly traded or not. The idea is that publicly traded firm probably are not affected by ς_t. As we can see, this is borne out in the data, albeit it's borderline significant.

D.2.5 Serial correlation of single investment spikes

So what can firms that are credit constrained do if both of its plants are so productive that ideally they should btw undergo investment projects? If credit is tight and thus external finance particularly costly, then it may see no other possibility than to focus its funds on investing in one plant and postponing investment in the other plant. We call this spacing out of investment spikes "adjacent investment spikes" and test for them by regressing a dummy variable that indicates such "adjacent spike"-firms on credit conditions. The unconditional probability (without any especially

Table 14: Credit Constraints and Firm-level Investment

$$\mathbf{I}\left\{(i/k)_{At}>0.15\right\}=\beta_0+\beta_1\mathbf{I}\left\{(i/k)_{Bt}>0.15\right\}+\beta_2\varsigma_t\mathbf{I}\left\{(i/k)_{Bt}>0.15\right\}+\beta_3\varsigma_t\mathbf{I}\left\{(i/k)_{Bt}>0.15\right\}\mathbf{I}\left\{\text{Public}\right\}+\varepsilon_{nt}$$

Coefficient	OLS	Panel
β_0	0.0378***	0.0463***
	(0.0130)	(0.0089)
β_1	0.1846***	0.1812***
	(0.0056)	(0.0056)
β_2	–0.0211***	–0.0205***
	(0.0051)	(0.0051)
β_3	0.0237*	0.0201
	(0.0132)	(0.0135)
Controls	Yes	Yes
N	66k	40k

tight credit) is about 10% (see Table **??**) and the regression results in Table 15 tell us that this probability drops by 0.7%.[11]

[11] In this analysis, we consider firm that undergo exactly one spike today and one or two spikes last year. Instead, we should have restricted the sample to only those firms that have *at least* 2 spikes within a two year window and then see if they are more likely to stretch the at least 2 spikes out or if they choose to do it simultaneously and also consider interaction terms with the publicly traded dummy. Since this sample is smaller than the one we consider here, we see our estimates as a lower bound.

Table 15: Credit Constraints and Serial Correlation of Investment Spikes

$$\mathbf{I}\left\{(i/k)_{At},(i/k)_{At-1} > 0.15\right\} = \beta_0 + \beta_1 \varsigma_t + \beta_2 \varsigma_t \cdot (i/k)_{Bt} + \varepsilon_{nt}$$

Coefficient	OLS	Panel
β_0	0.2148***	0.1836***
	(0.0354)	(0.0274)
β_1	0.0069**	0.0069**
	(0.0030)	(0.0032)
β_2	–0.0141	–0.0141
	(0.0130)	(0.0133)
Controls	Yes	Yes
N	17k	9k