Investment, Innovation, and Financial Frictions*

Pablo Ottonello
University of Michigan and NBER
pottonel@umich.edu

Thomas Winberry
Wharton and NBER
twinb@wharton.upenn.edu

July 27, 2023

Abstract

We study the role of financial frictions in determining long-run economic growth. Empirically, firms do more capital investment when they are small but do more innovation as they grow and become large. We interpret these findings using an endogenous growth model with heterogeneous firms and financial frictions. In our model, small firms are investment-intensive because they face a higher return to capital. Financial frictions lower aggregate growth by slowing the rate at which firms exhaust the returns to capital and start innovating. If ideas are non rival, policy should not only raise innovation but also lower investment expenditures among constrained firms.

*We thank Andy Abel, Fernando Alvarez, Cristina Arellano, David Argente, Marco Bassetto, Gideon Bornstein, Jeff Buckley, Satyajit Chatterjee, Nicolas Crouzet, Maarten De Ridder, Nik Engbom, Vincent Glode, Jeremy Greenwood, Hugo Hopenhayn, Pete Klenow, Pablo Kurlat, Simon Mongey, Guillermo Ordonez, Victor Rios-Rull, Richard Rogerson, Pierre-Daniel Sarte, and seminar participants at the Jackson Hole Macro Conference 2023, the Penn-Wharton Brownbag, the Minneapolis Fed, the University of Wisconsin, the Richmond Fed, the conference in honor of Richard Rogerson's 65th birthday, the London School of Economics, the Philadelphia Fed, the Cowles Macro Conference 2023, and the NBER Summer Institute 2023. We thank Samuel Stern and Erin Gibson for excellent research assistance.

1 Introduction

In the long run, economic growth is driven by new ideas which push out the technological frontier. While many of these new ideas are manifested in new firms entering the economy, a substantial share of ideas come from existing firms already in the economy. Because they already have ideas in place, existing firms must decide not only how much to innovate — which creates new ideas — but also how much to invest — which scale up production using their existing ideas. To the extent that a firm is financially constrained, these two activities will compete for the same funds within the firm. How do financial frictions distort the mix of investment and innovation within firms at the micro level? Do these distortions quantitatively matter for economic growth at the macro level?

We address these questions using new cross-sectional evidence and an endogenous growth model with heterogeneous firms and financial frictions. Empirically, we find a pecking order of firm growth; firms do more capital investment when they are small but do more innovation as they grow and become large. Our model matches this finding because firms have a higher return to capital when they are small. Financial frictions lower the aggregate growth rate because they delay the rate at which firms accumulate capital, lower its return, and start innovating. If ideas are non rival, this allocation is suboptimal because firms do not internalize how their innovations raise other firms' productivity. Of course, this externality motivates higher innovation, but supporting higher innovation also entails lowering investment expenditures among constrained firms. We show how these changes can be achieved by subsidizing innovation or, to a lesser extent, by lowering taxes on investment.

Our empirical work focuses on how the composition of investment and innovation changes within existing firms as they grow over time. We measure investment in physical capital and proxy for innovation using R&D expenditures and patenting activity. Our baseline sample is drawn from Compustat, a panel of publicly-listed U.S. firms. The long panel dimension of Compustat allows us to account for fixed differences across firms, which turns out to be crucial for our results. However, Compustat is a selected subsample of larger and older firms in the economy. We address this selection in our economic model by calibrating it to a broader set of firms in the economy and explicitly modeling the selection into Compustat.

We also find that our results hold in Orbis data, which contains some large private firms.

We find that firms do roughly twice as much investment but half as much innovation when they are small compared to when they grow large. These patterns hold for various measures of size, such as capital, employment, or sales. These patterns also hold by age — firms do more investment when they are young and but do more innovation when they are old — as well as net worth — firms do more investment when they have few assets but do more innovation when they have many assets. We interpret these findings to indicate that the relative return to innovation is decreasing in the shadow price of external finance, as proxied by size, age, or asset position.[1]

Motivated by this evidence, we develop a heterogeneous-firm endogenous growth model in which firms face financial constraints. In order to focus on the decisions of incumbent firms, we assume that new entrants simply draw an idea, embodied in their productivity, from the existing stock of ideas in the economy. Firms must then decide how much to spend on investment, which increases the capital stock used in production, and how much to spend on innovation, which increases the probability of receiving a new idea and raising the firm's productivity. The firm's mix of investment and innovation is determined by the relative return on these two activities. The return to capital is determined by its marginal product and value as collateral in external borrowing, while the return to innovation is determined by the probability of generating a new idea and the present value of that idea to the firm.

Our model generates the same pecking order of firm growth we found in the data; firms are investment intensive when they are small but become innovation intensive when they grow large. The return to capital is high when a firm is small because the marginal product of capital has not yet diminished and its use as collateral is more highly valued. As the firm accumulates capital and grows, the marginal product and collateral value of capital

[1] Most existing empirical studies find that innovation rates are higher among small firms rather than large firms, in contrast to our results (see, for instance, Akcigit and Kerr (2018)). There are two possible reasons for this difference. First, most of the literature focuses on innovation by new entrants into the economy, while we focus on innovation among incumbent firms already established in the economy. These two sets of firms may face fundamentally different tradeoffs, which we discuss in Section 4. Second, within our Compustat sample of incumbent firms, we find that our use of *within-firm* variation is key to uncovering our empirical results. If we instead look *across* firms, we find that small firms do more innovation, in line with the existing literature. While these across-firm comparisons are interesting, they are driven by differences in fixed characteristics, such as technological factor intensities, that our outside of our model.

fall, inducing the firm to shift into innovation. Financial frictions are central to this process because they control how quickly the firm can accumulate capital and drive down its return. In fact, without any frictions, firms would immediately lever up to their optimal scale and begin innovating, leaving investment and innovation counterfactually independent of size, age, or net worth. In this sense, financial frictions create the pecking order of firm growth in our model.

Quantitatively, the strength of the model's pecking order is governed not only by the strength of financial frictions but also by the innovation technology, which determines the return to innovation. Inferring the properties of this technology is difficult because its output, new ideas, are difficult to directly measure in the data. We infer the realization of new ideas using what firms reveal to us through their forward-looking investment decisions. In particular, our model predicts that new ideas should generate investment spikes — short-lived bursts of investment — because ideas are complementary with capital in production. Consistent with this prediction, R&D expenditures are strongly associated with investment spikes in our Compustat data; having one standard deviation higher R&D increases the probability of a spike by approximately 40% in the following year. We use these joint dynamics of R&D and investment spikes to discipline the innovation technology in our model.

We validate the quantitative strength of financial frictions in our model using new empirical evidence on the response of innovation to changes in the after-tax price of capital. Specifically, we study the Bonus Depreciation Allowance, a countercyclical investment stimulus that increased tax deductions for investment during the 2001 and 2008 recessions. Following Zwick and Mahon (2017), we exploit heterogeneity in the policy treatment across firms to estimate the effect of lower investment taxes on R&D expenditures using a difference-in-difference empirical design. We find that lower taxes on investment significantly raise R&D expenditures in our Compustat data. We then feed in a similar-sized shock to the relative price of investment to our model and replicate these regressions on model-simulated data. We find that the model's implied regression coefficient is similar to the data.

We use our calibrated model to quantify the effects of financial frictions long-run economic growth. To do so, we compare the balanced growth path (BGP) in our model to a frictionless version of the model in which firms face no financial constraints. The frictionless model has no pecking order of firm growth because firms immediately accumulate their optimal scale of capital and begin innovating. Financial frictions generate two distortions relative to this benchmark. First, given the existing distribution of productivity, financial frictions misallocate capital across firms, lowering the *level* of aggregate TFP as in Hsieh and Klenow (2009). Second, and novel to our framework, financial frictions delay the point at which individual firms begin innovating, lowering the *growth rate* of aggregate TFP.

We find that financial frictions lower aggregate economic growth by 45 basis points per year. Over long horizons, this lower growth rate leads to much larger efficiency costs than the costs from capital misallocation. In addition, because financial frictions primarily lower innovation by small firms, the majority of innovation activity in our model is concentrated among the largest, most unconstrained firms in the economy. This distribution of innovation implies that financial crises, modeled as an unexpected tightening of financial constraints, do not persistently reduce innovation or aggregate growth in the medium run.

To the extent that ideas are non rival, this equilibrium allocation of investment and innovation is socially inefficient. In order to understand this externality, we study the allocation chosen by a social planner who is subject to the same financial constraints as individual firms. Clearly, the planner wants higher innovation because it internalizes the positive externality from non-rival ideas. We instead focus on how the subtler question of how the planner reallocates investment in order to support this goal. The planner's choice is heterogeneous across firms: higher innovation requires lower investment expenditures for constrained firms (due to the flow-of-funds constraint) but incentivizes higher investment for unconstrained firms (due to the complementarity of productivity and capital). We study how the planner balances these forces along a transition path starting from the equilibrium BGP. In the early phase of this transition, the substitutability for constrained firms dominates in the sense that aggregate investment falls. Over time, however, the resulting growth builds up the distribution of net worth, and eventually the complementarity for unconstrained firms dominates in the sense that aggregate investment increases.

We use these insights to evaluate the effects of practical policies on the allocation of investment and innovation. Unfortunately, simple policies cannot exactly decentralize the planner's allocation because the planner's incentives vary across firms and over time. However, we find that an appropriately-chosen innovation subsidy that is constant across firms and time gets very close to the planner's allocation in terms of aggregate variables. We also find that an investment tax cut can raise innovation in the long run, even for constrained firms (since the lower tax rate ultimately reduces after-tax *expenditures* on investment). This result contrasts with the neoclassical growth model, in which investment tax cuts have no effect on long-run growth. However, the tax cut creates a temporary surge in investment which raises the real interest rate and reduces innovation in the short run. In this sense, an investment tax cut can only partially substitute for the nearly-optimal innovation subsidy.

Related Literature Our findings on the long-run costs of financial frictions contribute to the literature on input misallocation across firms. Restuccia and Rogerson (2008) and Hsieh and Klenow (2009) show how reduced-form wedges can lower aggregate TFP by generating dispersion of marginal products across firms. Buera, Kaboski and Shin (2011), Midrigan and Xu (2014), and Moll (2014), among others, endogenize these wedges using structural models of financial frictions. However, these papers take the distribution of productivity as given, so the costs of financial frictions only arise from distorting capital as a function of productivity. In our model, financial frictions also distort the distribution of productivity by affecting innovation decisions. These innovation distortions lower the growth rate of TFP, while capital misallocation only lowers the level of TFP. We find that the level effects from capital misallocation are smaller than the growth effects of lower innovation, at least in the context of the U.S. economy to which our model is calibrated.

Our model combines elements of the Hopenhayn (1992) framework, in which firm dynamics are determined given an exogenous process for productivity, and the endogenous growth framework, in which productivity is determined through innovation. A key feature of Hopenhayn (1992) is decreasing returns to scale, which implies that firms have an optimal scale given their level of productivity. The literature has studied how various frictions impede the ability of firms to reach this optimal scale, such as firing costs in Hopenhayn and Rogerson (1993), capital adjustment costs in Khan and Thomas (2008), selection upon entry in Clementi and Palazzo (2016), or — most closely related to our model — financial frictions in Khan and Thomas (2013). We contribute to this literature by incorporating innovation, which endogenizes the productivity process and therefore the distribution of optimal size.[2]

On the endogenous growth side, our focus on firm dynamics is related to the creative destruction literature pioneered by Aghion and Howitt (1992) and Grossman and Helpman (1991), and more recently used in quantitative analyses by, e.g., Klette and Kortum (2004), Akcigit and Kerr (2018), or Acemoglu et al. (2018). Most of these models abstract from frictions to factor accumulation and, therefore, have no source of sluggish input dynamics.[3] We contribute to this literature by incorporating capital accumulation and the sluggish dynamics induced by financial frictions. Our focus on differences in innovation intensity across firms is most closely related to Akcigit and Kerr (2018), who study how firms choose between two different types of innovation. We abstract from different types of innovation to instead study the choice between innovation and capital investment.[4]

Road Map The rest of our paper is organized as follows. Section 2 documents the pecking order of firm growth in the data. Motivated by this evidence, Section 3 develops the model and Section 4 describes how the model matches the pecking order. Section 5 calibrates the key parameters of the model and shows that it matches our new evidence on the response of innovation to changes in the tax rate on investment. Section 6 uses the calibrated model to quantify the aggregate effects of financial frictions. Section 7 shows how the non rivalry of ideas opens the door to policy intervention and evaluates the effects of innovation subsidies and investment tax cuts. Section 8 concludes.

2 Pecking Order of Firm Growth

Both our empirical and model analysis focus on a particular subset of firms in the economy. The vast majority of firms pursue little to no innovation and their scale is very small, perhaps for non-pecuniary reasons (see Hurst and Pugsley (2011)). We omit these firms from our analysis and instead focus on firms who eventually innovate and meaningfully contribute to

[2]Relatedly, Atkeson and Burstein (2010) embed innovation decisions in a Melitz (2003)-style model without capital in order to study the dynamic gains from trade.

[3]An important exception is Bilal et al. (2021), who study a version of the creative destruction model with search frictions in the labor market.

[4]In a broadly related vein, Crouzet et al. (2022) develop a model with two different types of capital that differ in their degree of non rivalry.

economic growth. We conceptualize the lifecycle of these firms in two phases. In the first *entry phase*, an entrepreneur uses their own time and skills to generate a new idea. If a new idea materializes, the entrepreneur creates a firm and implements the idea in practice. Once the production process using that idea is established, which may include the first couple years of the firm's life, the firm enters the *incumbent phase*. In this phase, the firm must decide how much to scale up its existing idea through investment and how much to attempt to generate a new idea through innovation.

We focus on the incumbent phase of a firm's life for three reasons. First, this phase contains the tradeoff between investment and innovation which motivates our paper. Second, the incumbent phase accounts for a significant share of aggregate R&D expenditures and patenting activity. Third, recent work by Akcigit and Kerr (2018) and Garcia-Macia, Hsieh and Klenow (2019) estimate that around three-quarters of aggregate growth comes from incumbent firms rather than new entrants. That said, our analysis largely abstracts from how financial frictions affect innovation among new entrants. Therefore, our aggregate results should be interpreted as a lower bound on the growth effects of financial frictions.

2.1 Data Description

Our main analysis uses annual firm-level data from Compustat, a panel of publicly listed U.S. firms, from 1975 – 2018. This dataset has two key advantages for our analysis. First, it contains joint information on firms' investment, R&D expenditures, and financial positions, allowing us to measure our key variables of interest. Second, it is a long panel, which allows us to absorb permanent differences across firms using fixed effects. We show in Section 2.3 that using within-firm variation is critical for our results.

Of course, Compustat is a selected subsample of the types of innovative incumbent firms in which we are interested. We address selection into Compustat in two ways. First, we show that our results also hold in Orbis, which contains some larger private firms.[5] Second,

[5] An alternative source of data on the investment and innovation decisions of privately-held firms is the Annual Survey of Manufacturers (ASM) from the U.S. Census. However, for small firms, these data are a rotating sample, which limits the within-firm variation we could use (see the discussion in Kehrig (2015)). In addition, these data only cover the manufacturing sector, while Compustat contains broader sectoral coverage of the economy.

we calibrate the key features of our economic model to be representative of a broader set of firms in the economy, and explicitly account for the selection of firms into our observed Compustat data within the model.

Our main variables of interest are firms' investment and innovation activity. We measure the investment rate as the ratio of capital expenditures to the lagged value of plant, property, and equipment. Innovation activity is more difficult to measure, so we proxy for it in two ways. First, we proxy for the inputs into the innovation process using the ratio of R&D expenditures to sales; since sales are noisy, we average them over the last five years. Second, we proxy for the outputs of the innovation process using approved patents collected from the United States Patent and Trademark Office by Kogan et al. (2017). Appendix A describes the details of how we clean the data and presents descriptive statistics of our final sample. For our baseline analysis, we exclude observations associated with acquisitions in order to focus on innovation and investment occurring within firms.

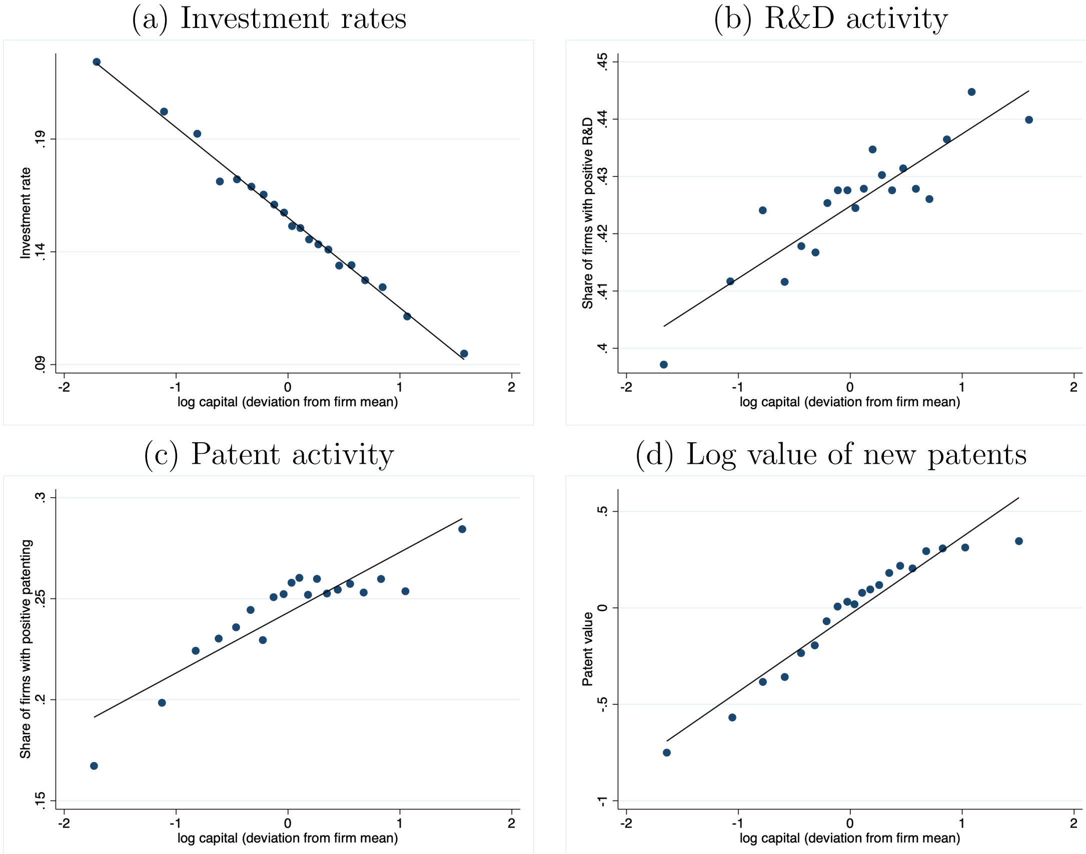
2.2 The Pecking Order of Firm Growth

We illustrate our pecking order of firm growth using simple binned scatterplots of investment and innovation by the size of the firm's capital stock. We isolate within-firm variation by de-meaning both size and investment/innovation at the firm level, which is equivalent to using a firm fixed effect in a regression context. We condition on firms with at least 20 years of observations in order to precisely estimate the firm-level mean, but Appendix A shows our results also hold for the whole sample of firms. In order to make the units of the outcome variables more interpretable, we add back in their mean values across all firms (which is a normalization that does not affect any results).

Figure 1 illustrates our two key empirical results. First, panel (a) shows that firms' investment activity *decreases* as firms grow; the smallest firms' investment rates exceed 0.2 but fall below 0.1 as they grow large. This pattern is consistent with the notion that firms face a higher relative return to capital when they are small than when they are large.

Second, panels (b) – (d) shows that firms' innovation activity instead *increases* as firms grow larger. In terms of innovation inputs, panel (b) shows that the share of firms with positive R&D expenditures increases by about 10% as firms grow. We report this extensive

Figure 1: The Pecking Order of Firm Growth



Notes: Binned scatter plots of investment rates, the share of firms with positive R&D, the share of firms with positive patenting, and the log of market value of new patents (measured by Kogan et al. (2017)) by firm size (measured by the log of real capital). All variables are demeaned at the firm level. In order to make the units of the outcome variable more interpretable, we add back in the unconditional mean across all firms of investment rates, share of firms with positive R&D, and share of firms with positive patenting. For variable definitions and sample selection, see Appendix A.

margin measure to highlight the fact that more than half of firms do not report positive R&D in a given year; this inaction in innovation activity will be an implication of our model as well.[6] Appendix A shows that the intensive margin, i.e. the amount of R&D expenditures among firms who report positive R&D, is also increasing in firm size.

In terms of innovation outputs, panels (c) and (d) show that patenting activity increases as firms grow larger. On the extensive margin, panel (c) shows that firms are about half as likely to obtain a successful patent when they are small compared to when they are large.

[6] A potential concern is that the inaction in the data merely reflects reporting error rather than true inaction. We address this concern in Appendix A by conditioning on observations after the firm reports its first positive R&D expenditure (and therefore has presumably set up the accounting infrastructure to report R&D expenditures going forward). We find similar results in this subsample.

Table 1
The Pecking Order by Various Measures of Size

	(1) Investment rate	(2) R&D activity	(3) R&D-to-sales	(4) Patent activity	(5) log value per patent
log capital					
$\hat{\gamma}$	-0.088***	0.026***	0.002**	0.067***	0.56***
	(0.003)	(0.008)	(0.0007)	(0.009)	(0.050)
N	50464	47145	46439	54419	16731
R^2	0.27	0.86	0.87	0.63	0.83
log employment					
$\hat{\gamma}$	-0.047***	0.035***	0.002***	0.11***	0.50***
	(0.005)	(0.011)	(0.0007)	(0.011)	(0.059)
N	46885	44563	43575	49591	16323
R^2	0.21	0.86	0.88	0.64	0.82
log sales					
$\hat{\gamma}$	-0.068***	0.025***	0.002***	0.075***	0.65***
	(0.004)	(0.008)	(0.0007)	(0.009)	(0.050)
N	50464	47145	46439	54419	16731
R^2	0.23	0.86	0.87	0.63	0.84
log net worth					
$\hat{\gamma}$	-0.069***	0.015**	0.002**	0.050***	0.43***
	(0.003)	(0.007)	(0.0006)	(0.007)	(0.041)
N	46368	43289	42666	49821	15614
R^2	0.26	0.86	0.88	0.64	0.82
Mean	0.13	0.44	0.02	0.34	0

Notes: Results from estimating the regression $o_{jt} = \alpha_j + \gamma \log s_{jt} + \varepsilon_{jt}$, where o_{jt} is the outcome of interest (investment rate, indicator for positive R&D, R&D-to-sales ratio, indicator for positive patenting, or log market value per patent computed following Kogan et al. (2017)), s_{jt} is the measure of size (capital, sales, employment, or net worth), and α_j is a firm fixed effect. We standardize the size measures $\log s_{jt}$ and the log value per patent over the entire sample. Standard errors are clustered at the firm level.

On the intensive margin, panel (d) shows that the market value of those new patents — which Kogan et al. (2017) compute as the change in firm equity value in a narrow window around the patent approval — increases by about one log point (approximately 170%) as firms grow. While individually none of these measures of R&D or patenting activity fully captures firms' innovation, they are collectively consistent with the notion that firms face a higher relative return to innovation as they grow.

Table 1 shows that the pecking order of firm growth holds for other measures firm size as well. We summarize the pecking order using the regression

$$o_{jt} = \alpha_j + \gamma \log s_{jt} + \varepsilon_{jt}, \tag{1}$$

where o_{jt} is the outcome of interest (investment rate, indicator for positive R&D, indicator for positive patenting, or log value of new patents) and s_{jt} is the measure of size (capital, sales, employment, or net worth). As discussed above, the firm fixed effects α_j ensure we use within-firm variation. The coefficient of interest is γ, which measures how the outcome of interest varies with the particular measure of firm size. We standardize each size variable $\log s_{jt}$ over the entire sample in order to make the units of the coefficient γ easier to interpret. We cluster standard errors at the firm level.

The first row of Table 1 quantifies the magnitudes and statistical significance of the binscatters using the regression (1). Column (1) shows that having one standard deviation more capital than the average firm lowers the firm's investment rate by nearly 9 percentage points relative to the unconditional mean of 13 percentage points — a more than 60% decline in investment as firms grow. In contrast, columns (2) – (5) show that having more capital systematically raises our various proxies of innovation activity. For example, having one standard-deviation higher capital increases the market value associated with that year's patenting activity by one-half a standard deviation.

The remaining rows of Table 1 show that these general patterns hold for other measures of size as well. The second and third rows proxy for size using employment or sales, which are common in the literature. The final row proxies for size using net worth, which is the relevant state variable in our economic model. The magnitudes of the pecking order are similar for all of these measures of firm size.

Robustness Appendix A contains additional robustness analysis. Most importantly, we show that our pecking order of firm growth also hold among Orbis data, which contains privately-held firms. We also show that our results are robust to alternative versions of our baseline specification. First, our results are robust to including all firms in the sample (rather than only those who have survived at least 20 years, as in our baseline analysis). Second, our results hold for other measures of innovation activity. Third, the pecking order also holds using within-sector, rather than within-firm, variation. Fourth, our results are robust to including time fixed effects to capture secular trends.

2.3 Comparison to Existing Literature

We now compare our results to the existing literature, which tends to find that small firms do more innovation than large firms (in contrast to our findings). We reconcile this difference in two ways. First, within our sample, we show that using purely within-firm variation is key to obtaining our results. Second, the exiting literature often focuses on new entrants that are outside our sample. These new entrants may behave in fundamentally different ways from the firms in our sample. Due to data limitations, the existing literature often does not focus on the purely within-firm variation among these new entrants (see Footnote 5). Nevertheless, in Section 4, we show that allowing new entrants to have a comparative advantage in innovation does not affect the key results in our paper.

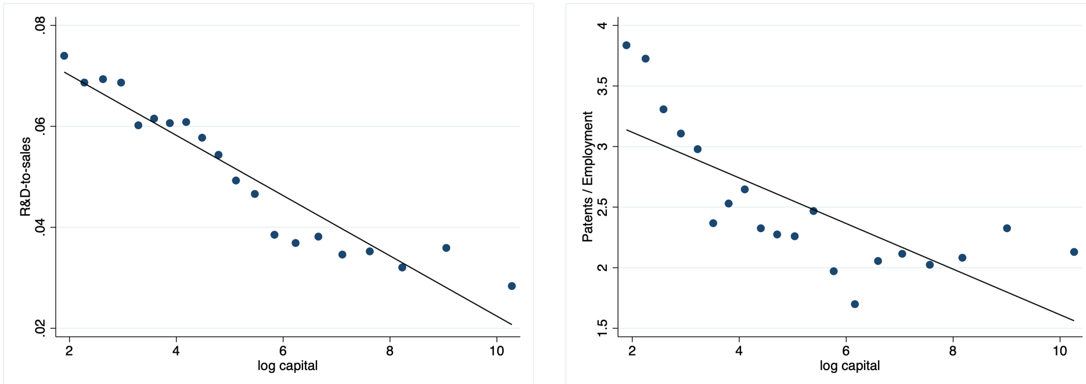
To explain the role of within-firm variation in our sample, we find it useful to directly compare our results with Akcigit and Kerr (2018), which studies how the R&D-to-sales and patents-to-employee ratios vary with firm size in the Longitudinal Business Database (LBD) from the U.S. Census Bureau.[7] We use Akcigit and Kerr (2018)'s sample selection of "continuously innovative firms," i.e., firms that have conducted positive R&D or patenting activity over the last five years. However, Appendix A shows that the comparisons from this subsection hold for the full sample of firms as well.

The top row of Figure 2 confirms that Akcigit and Kerr (2018)'s results also hold within our sample if we use all variation in the data; small firms are more innovation-intensive than large firms. However, the bottom row shows that the opposite result holds using only within-firm variation; in this case, firms are less innovation-intensive when they are small compared to when they are large. Appendix A shows that these comparisons also obtain

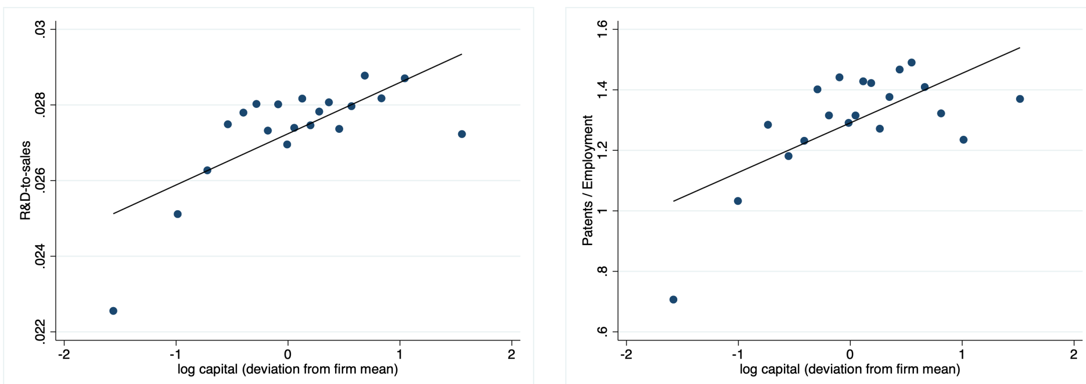
[7]While the LBD contain information on sales and employment for a large set of firms, it does not contain information on capital investment. To our knowledge, the Census data that also contains investment only covers the manufacturing sector, as discussed in Footnote 5.

Figure 2: Role of Within-Firm Variation

All variation



Within-firm variation



Notes: Binned scatter plots of R&D to sales and patents per employee by firm size (measured by the log of real capital). All variables are demeaned at the firm level. In order to make the units of the outcome variable more interpretable, we add back in the unconditional mean across all firms of investment rates, share of firms with positive R&D, and share of firms with positive patenting. "Continuously innovative firms" indicates that we restrict to the sample to firms that have conducted positive R&D or patenting activity over the last five years, as in Akcigit and Kerr (2018). For variable definitions and sample selection, see Appendix A.

using sales, employment, or net worth to measure size instead of capital.

While the across-firm variation driving the results in the top panel of Figure 2 is an interesting source of firm heterogeneity, we do not focus on this dimension of the data for the rest of our paper. Our economic model makes predictions about how financial frictions affect firms' decisions relative to what that same firm would have done absent those frictions. These predictions are most directly disciplined by variation in how the same firm has behaved at different points in time.[8]

[8] In contrast, the variation across firms also includes fixed differences across firms in things like technological factor intensities. For example, large manufacturing firms may have a permanently high scale but engage in little innovation in order to focus on production, while small technology firms may have a permanently low

3 Model

We now develop our model of investment and innovation that is consistent with the evidence presented above. The model is set in discrete time and there is no aggregate uncertainty.

3.1 Environment

We purposefully keep the entry phase of our model as simple as possible in order to focus on the tradeoff between investment and innovation in the incumbent phase.

Entry Phase There is a fixed flow π_d of new entrants each period that are endowed with zero debt and draw their initial levels of productivity and capital from some distribution $\Phi^0_t(z,k)$. In order to capture imitation by new entrants (as in, e.g., Luttmer (2007)), we assume this initial distribution of productivity is related to the distribution of incumbent firms in the economy; we will parameterize and calibrate this dependence in Section 5. Imitation is one sense in which ideas are non rival in our model and is necessary to ensure the model generates a positive growth rate along the balanced growth path (BGP).

While our model of the entry phase is extremely simple, the equilibrium BGP of our model is identical to a version of the model with free entry in which the entry cost grows along with the economy (which Klenow and Li (2022) argue is the case empirically). However, neither model captures how financial frictions constrain the ability of potential entrants to come up their first idea. Quantitatively modeling this margin would require introducing a separate set of model features — such as the innovation technology of new entrants, the ability of entrepreneurs to smooth consumption, and features of the venture capital market such as expert advise and the dilution/moral hazard of founders — that are largely unrelated to the tradeoff between investment and innovation in which we are interested. Therefore, we prefer to abstract from these issues, focus on how financial frictions affect incumbent firms, and interpret our analysis as a lower bound on the effect of financial frictions on long-run growth.

optimal scale but focus on innovation rather than production. Consistent with this example, Appendix A shows that our pecking order of firm growth also holds if we only use within-sector, rather than within-firm, variation. This result suggests that fixed differences across sectors largely drive the results in the top panel of Figure 2.

Incumbent Phase Firms in the incumbent phase produce output $y_{jt} = (A_t z_{jt})^{1-\alpha} k_{jt}^{\alpha}$ where j indexes a firm, A_t is aggregate productivity, z_{jt} is firm-specific productivity, and k_{jt} is the firm's capital stock.[9] Decreasing returns to capital $\alpha < 1$ ensure there is an optimal scale of the firm for each level of productivity, as in Hopenhayn (1992).[10] At the beginning of the period, a random subset of firms learn that they must exit the economy, in which case they produce, sell their undepreciated capital $(1-\delta)k_{jt}$, and pay back their existing debt. This exit shock occurs with probability π_d, the same as the flow of new entrants, which ensures the total mass of firms in production is constant over time.[11] Exit shocks are a common tool to ensure that firms do not outgrow their financial frictions in the long run.

Firms that will continue into the next period spend resources on investment and innovation. Investment expenditures x_{jt} yield capital in the next period following the standard accumulation equation $k_{jt+1} = (1-\delta)k_{jt} + x_{jt}$. Innovation i_{jt} yields a higher probability of a successful innovation $\eta(i_{jt})$, which permanently raises productivity by a factor Δ:

$$\log z_{jt+1} = \begin{cases} \log z_{jt} + \Delta + \varepsilon_{jt+1} \text{ with probability } \eta(i_{jt}) \\ \log z_{jt} + \varepsilon_{jt+1} \text{ with probability } 1-\eta(i_{jt}) \end{cases}, \tag{2}$$

where $\varepsilon_{jt+1} \sim N(0, \sigma_\varepsilon)$ are idiosyncratic shocks to productivity growth unrelated to innovation. These innovations can capture new technologies, process innovation, or the introduction of new products. We assume that the total amount of expenditures required to achieve this probability of success is $i_{jt} \times A_t z_{jt}$. This assumption captures the notion that successive ideas become harder to find because the easy ones are "fished out." We will show in Section 4 that

[9]Appendix D shows that this production function can be derived from a model in which labor is a variable input in production: $y_{jt} = (A_t z_{jt})^{1-\widetilde{\alpha}} k_{jt}^{\widetilde{\alpha}} \ell_{jt}^{\widetilde{\nu}}$ with $\widetilde{\alpha} + \widetilde{\nu} < 1$. The production function in the main text is equal to the reduced-form variable profit function $\max_{\ell_{jt}} (A_t z_{jt})^{1-\widetilde{\alpha}} k_{jt}^{\widetilde{\alpha}} \ell_{jt}^{\widetilde{\nu}} - w_t \ell_{jt}$. In this case, the production parameter α reflects the elasticity of revenue with respect this combination of inputs through $\alpha = \frac{\widetilde{\alpha}}{1-\widetilde{\nu}}$. We calibrate the model under this interpretation. Hence, incorporating variable labor would not affect our results, so we omit it from the main text for the sake of parsimony.

[10]The endogenous growth literature typically assumes constant returns to scale in objects (here capital) on the basis of the replication argument, which then implies increasing returns to scale in objects and ideas jointly (here capital and productivity). Our model is consistent with this view if we interpret decreasing returns as reflecting a downward-sloping demand curve or if we allow for free entry.

[11]The fixed mass of firms is the only specification which is consistent with an stable average firm size in the model with endogenous labor and fixed population discussed in Footnote 9. If the mass of firms were to grow or shrink over time, then the average firm size would asymptote to zero or infinity. Allowing for growth in the number of firms therefore requires incorporating population growth, which is outside the scope of this paper.

this specification of the innovation technology implies that all financially unconstrained firms have the same growth rate, i.e. Gibrat's law holds for unconstrained firms. This property provides useful benchmark both because Gibrat's law arguably holds among large firms in the data and because it is a common property of other endogenous growth models. In contrast, financially constrained firms will grow faster than unconstrained firms because constrained firms face a higher marginal product of capital.

We assume that firms cannot sell their existing ideas, i.e. innovation expenditures must be non-negative $i_{jt} \geq 0$. In principle, financially constrained firms may have an incentive to sell their ideas in order to finance investment. In practice, the "market for ideas" — licensing arrangements, patent sales, mergers and acquisitions, etc. — is rife with frictions. We view our assumption that $i_{jt} \geq 0$ as the limit in which those frictions are sufficiently large to prevent trade in the market for ideas altogether. These frictions allow the model to generate the inaction in innovation rates we documented empirically in Section 2, and we discuss how to relax our stark assumptions about these frictions in Section 8.

Firms have two sources of finance for their investment and innovation expenditures. First, firms can borrow externally, but this borrowing is subject to the collateral constraint $b_{jt+1} \leq \theta k_{jt+1}$. This constraint can be derived from an environment in which firms lack commitment to repay their debts, and lenders can seize a fraction $\frac{\theta}{1-\delta}$ of their capital if firms default. Second, firms can use their internal resources, but they cannot raise new equity. This assumption implies that dividend payments must be nonnegative:

$$d_{jt} = (A_t z_{jt})^{1-\alpha} k_{jt}^{\alpha} + (1-\delta)k_{jt} - b_{jt} - k_{jt+1} - A_t z_{jt} i_{jt} + \frac{b_{jt+1}}{1+r_t} \geq 0.$$

This no equity-issuance constraint may seem overly strong at face value in light of the role of venture capital play in financing young firms in the technology sector. However, equity issuance among established incumbent firms, which are the focus of our model, is relatively rate. In any event, we show in Section 4 that it is the combined effects of the equity issuance costs and borrowing constraints which matter for firms decisions. Any combination of debt- and equity-financing frictions which delivers the same overall costs will have similar effects in our model. We validate our particular calibration of these combined effects in Section 5.

Aggregate productivity A_t depends on the distribution of individual productivity z_{jt}

$$A_t = \left(\int z_{jt} dj \right)^a . \tag{3}$$

This assumption captures the notion that firms can observe some average level of accumulated ideas in the economy, but only a fraction $a \geq 0$ is relevant for their own production decisions and/or may be appropriated for use. This assumption is the second way in which ideas are non rival in our model, capturing spillovers onto incumbent firms. We choose this form of spillover to clearly illustrate how financial frictions amplify the positive externality from the non-rivalry of ideas in Section 7. Allowing for the externality to enter through the innovation process would yield similar results.

Key Differences Between Capital and Ideas There are three key differences between capital k and ideas z in our model:

(i) *Return wedge*: there is a wedge between the returns to capital and returns to ideas. Generally speaking, the pecking order of firm growth is determined by how this return wedge varies with firm size. In our model, the wedge is derived from the constraint that firms cannot sell ideas $i_{jt} \geq 0$, and is shaped by forces (ii) and (iii) below.

(ii) *Diminishing returns to scale*: capital is subject to decreasing returns to scale, so all else equal the marginal product of capital is higher for small firms. This force leads to a higher return wedge for these firms.

(iii) *Collateral value*: capital is more collateralizable than ideas in external borrowing, further increasing the return wedge for small firms.

While we have made specific parametric choices regarding these differences, their general spirit will drive the results to come.

Households To close the model, there is a representative household with preferences represented by the utility function $\sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\sigma}-1}{1-\sigma}$, where $1/\sigma$ is the elasticity of intertemporal substitution (EIS). Since there is no aggregate uncertainty, firms discount future profits using the implicit risk-free rate

$$\frac{1}{1+r_t} = \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\sigma}. \tag{4}$$

3.2 Equilibrium

In order to define the equilibrium, it is convenient to formulate firms' decisions recursively. The firm's individual state variables are its individual productivity z_{jt} and its net worth $n_{jt} = (A_t z_{jt})^{1-\alpha} k_{jt}^{\alpha} + (1-\delta)k_{jt} - b_{jt}$.[12] Exiting firms set $k_{jt+1} = b_{jt+1} = 0$, while continuing firms' decisions are characterized by the Bellman equation

$$v_t^{\text{cont}}(z,n) = \max_{k',i,b'} n - k' - A_t z i + \frac{b'}{1+r_t} + \frac{1}{1+r_t}\mathbb{E}_t\left[v_{t+1}(z',n')\right] \text{ s.t. } d \geq 0 \text{ and } b' \leq \theta k', \tag{5}$$

where $\mathbb{E}_t\left[v_{t+1}(z',n')\right]$ integrates over the next period's exit shock, innovation success, and idiosyncratic productivity shocks. The implied decision rules induce a law of motion for the measure of firms, $\Phi_{t+1}(z,n) = T(\Phi_t; k'(\cdot), i(\cdot), b'(\cdot))(z,n)$.

A *competitive equilibrium* is a sequence of value functions $v_t(z,n)$; policies $k'_t(z,n)$, $i_t(z,n)$, and $b'_t(z,n)$; distribution of firms $\Phi_t(z,n)$; real interest rate r_t; and aggregate productivity A_t such that (i) firms optimize and the associated policy functions solve the Bellman equation (5); (ii) the evolution of $\Phi_t(z,n)$ is consistent with firm decisions; (iii) the real interest rate r_t is given by (4) with $C_t = \int (y_{jt} - (k_{jt+1} - (1-\delta)k_{jt}) - A_t z_{jt} i_{jt})\, dj$; and (iv) aggregate productivity is given by the definition (3).

Balanced Growth Path Because productivity grows over time, the limiting behavior of the model exhibits a balanced growth path. The BGP is characterized by the growth rate of $Z_t = A_t \times (\int z_{jt} dj) = (\int z_{jt} dz)^{1+a}$. The growth rate $1 + g^* = \frac{Z_{t+1}}{Z_t}$ equals the growth rate of all macroeconomic aggregates and scales the individual decisions and measure of firms.[13] Appendix B provides details.

[12] In fact, given that productivity has a unit root, it may be possible to further reduce the firm's state variable to the ratio n_{jt}/z_{jt}. We do not pursue this option in the main text for expositional convenience.

[13] Our model is a fully endogenous growth model in the sense that the extended version with labor described in Footnote 9 exhibits what Jones (2005) terms "strong scale effects." However, we could eliminate these strong scale effects by assuming that new ideas increase productivity $\log z_{jt}$ by $\Delta \times Z_t^{\phi}$ with $\phi < 0$ (our baseline model corresponds to $\phi = 0$). As discussed by Jones (2005), the effects of various counterfactuals or policies at any point along a transition path are continuous in the parameter ϕ.

4 The Pecking Order of Firm Growth in the Model

We now show that our model generates a pecking order of firm growth consistent with the data. We also discuss the key economic forces governing this pecking order, motivating how we calibrate the model in Section 5.

Characterizing Decision Rules In order to characterize the firm's decision rules, we first note that the marginal cost of spending resources on either investment or innovation is given by the firm's *shadow value of funds*, $\frac{\partial v_t(z,n)}{\partial n}$. This object represents the marginal value of keeping resources inside the firm and is therefore the opportunity cost of instead spending those resources on investment or innovation. Appendix B shows two straightforward results about the shadow value of funds. First, it is equal to $\frac{\partial v_t(z,n)}{\partial n} = 1 + \lambda_t(z,n)$, where $\lambda_t(z,n)$ is the Lagrange multiplier on the non-negativity constraint on dividends; that is, the shadow value of funds is equal to the household's value, 1, plus the shadow price of issuing equity, $\lambda_t(z,n)$. The second result is that firms equate this shadow price $\lambda_t(z,n)$ to the shadow price of additional borrowing when constrained, i.e. the expected value of the multipliers on future collateral constraints $\mu_t(z,n)$ in all possible states of the world. Hence, the multiplier $\lambda_t(z,n)$ encodes how both financial frictions affect the marginal cost of firm growth through either investment or innovation.

Building on this discussion, Proposition 1 characterizes firms' optimal choices of investment and innovation. This proposition extends a similar result from Khan and Thomas (2013)'s model without innovation.

***Proposition* 1.** *Consider a firm in period t that will continue operations in $t+1$, has productivity z, and has net worth n. Then there exist two functions $\overline{n}_t(z)$ and $\underline{n}_t(z,n)$ that partition the individual state space:*

(i) ***Financially unconstrained****: If $n \geq \overline{n}_t(z)$, then the financial wedge $\lambda_t(z,n) = 0$. Being financially unconstrained is an absorbing state.*

(ii) ***Currently constrained****: If $n \leq \underline{n}_t(z,n)$, then both the collateral constraint binds $b' = \theta k'$ and the financial wedge is positive $\lambda_t(z,n) > 0$.*

(iii) ***Potentially constrained****: Otherwise, the collateral constraint is not currently binding $b' < \theta k'$ but the financial wedge is positive $\lambda_t(z,n) > 0$.*

In any of these cases, the optimal choices for investment $k'_t(z,n)$, innovation $i_t(z,n)$, and external financing $b'_t(z,n)$ solve the system

$$1+\lambda_t(z,n) = \frac{1}{1+r_t}\mathbb{E}_t\left[(MPK_{t+1}(z',k')+1-\delta)\times(1+\lambda_{t+1}(z',n'))\right]+\theta\mu_t(z,n) \tag{6}$$

$$1+\lambda_t(z,n) \geq \frac{\eta'(i)}{A_t z}\frac{1}{1+r_t}\left(\mathbb{E}_t[v_{t+1}(z',n'|\text{ success })]-\mathbb{E}_t[v_{t+1}(z',n'|\text{ failure })]\right), \quad = \quad \text{if } i>0 \tag{7}$$

$$k'+A_t z i = n+\frac{b'}{1+r_t} \text{ if } \lambda_t(z,n)>0; \text{ otherwise, } b'_t(z,n)=b^*_t(z), \tag{8}$$

where $MPK_{t+1}(z',k') = \alpha\left(\frac{A_{t+1}z'}{k'}\right)^{1-\alpha}$ is the marginal product of capital, $\lambda_t(z,n)$ is the Lagrange multiplier on the no equity issuance constraint $d \geq 0$, and $\mu_t(z,n)$ is the multiplier on the collateral constraint $b' \leq \theta k'$.

Proof. See Appendix B. ■

The first part of Proposition 1 describes three different regimes of financial constraints. *Financially unconstrained* firms have zero probability of facing a binding collateral constraint, which implies that their financial wedge is $\lambda_t(z,n) = 0$. These firms are able to follow the policy rules from the version of the model without financial frictions and are indifferent over any combination of external financing b' and internal financing d leaves them financially unconstrained. As in Khan and Thomas (2013), we resolve this indeterminacy by requiring that firms pursue the "minimum savings policy," i.e., the smallest level of $b' \equiv b^*(z)$ that leaves them unconstrained with probability one.

The remaining firms are affected by financial frictions in some way. *Currently constrained* firms' collateral constraint binds in the current period, which directly limits their ability to accumulate capital. *Potentially constrained* firms do not face a binding collateral constraint in the current period, but there is a positive probability of reaching a future state in which the constraint becomes binding. Financial frictions still affect these firms' decisions through precautionary motives.

The second part of Proposition 1 characterizes the investment and innovation decisions for any of these three types of firms.[14] Equations (6) and (7) are the first-order conditions for investment and innovation. As discussed above, the marginal cost on the left-hand side of these equations is given by the shadow value of funds $1+\lambda_t(z,n)$. The marginal benefit on the right-hand side of the capital first-order condition (6) is given by two terms: the discounted expected marginal product of capital in the next period and the marginal collateral benefit provided by additional capital. This first-order condition always holds with equality because firms can freely sell capital.

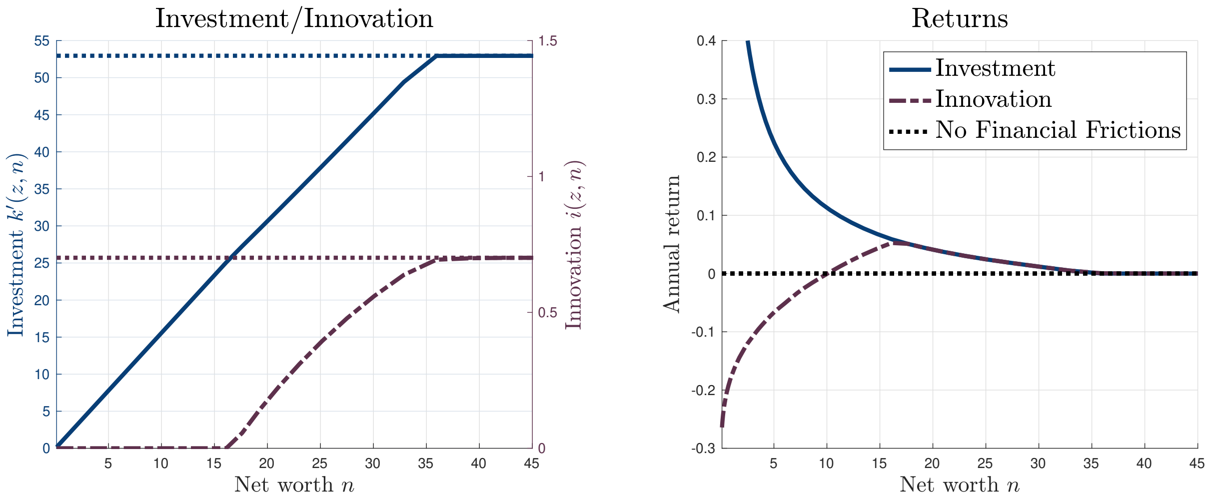
In contrast, the innovation first-order condition (7) may not hold with equality because firms face a non-negativity constraint on innovation $i_t(z,n) \geq 0$. The marginal benefit of innovation on the right-hand side of (7) is the marginal improvement in the probability of success per unit of innovation expenditure times the expected improvement in firm value from a successful innovation. Equation (8) is the nonnegativity constraint on dividends, which binds as long as the firm has a positive financial wedge $\lambda_t(z,n) > 0$. In this case, innovation and investment expenditures must be financed out of either internal resources or new borrowing.

Illustrating the Pecking Order We plot the decision rules which emerge from these decisions in order to illustrate the key economic tradeoffs in our model. We generate these plots using our calibrated parameter values from Section 5, but the qualitative properties that emerge hold for a wide range of the parameter space. The left panel of Figure 3 plots the investment and innovation policies as a function of net worth. The right panel plots the net returns associated with each activity, i.e. the right-hand side of the respective first-order condition minus one. We fix the level of productivity z to its average value in these figures to illustrate how the decision rules depend on net worth.

The model's pecking order can be summarized by three distinct regions in net worth space for a given level of productivity z. In the first region, the firm spends all of its available

[14]Our numerical algorithm solves the firm's problem by jointly iterating over the policy functions and the Lagrange multipliers $\lambda_t(z,n)$ in (the detrended version of) this system (6) - (8). This procedure is very fast because it avoids any numerical maximization or equation solving. In practice, we find computational runtimes comparable to using Carroll (2006)'s endogenous grid method, even though that method does not apply to this model. Our algorithm is applicable to other investment models in which the endogenous grid method does not apply. See Appendix C for details.

FIGURE 3: The Pecking Order of Firm Growth in the Model



Notes: the left panel plots capital expenditures $k_t(z,n)$ (left axis) and innovation expenditures $i_t(z,n)$ (right axis) in market equilibrium BGP of the calibrated model for fixed z. The right panel plots the return to these activities, defined as the RHS of Euler equations (6) and (7) minus 1. "No financial frictions" refers to the model in which all firms following the unconstrained policies $k^*(z)$ and $i^*(z)$ from Proposition 1, but using the same real interest rate from the market BGP.

resources on capital and sets innovation expenditures $i_t(z,n) = 0$. This choice is optimal because the return to capital lies strictly above the return to innovation. As net worth increases, the firm is able to accumulate more capital, which drives down its return due to the diminishing marginal product of capital and the lower shadow value of collateral. At the same time, higher capital also raises the return to innovation because TFP and capital are complements in production. However, in this region, the return to capital is always higher than the return to innovation, so the firm only grows its size by accumulating capital.

As the firm continues to grow, the returns to capital and innovation eventually intersect and the firm begins innovating. In this second region of the pecking order, the innovation first-order condition (7) holds with equality, so the returns to capital and innovation must be equalized. However, both (net) returns are still greater than zero because the financial wedge is positive $\lambda_t(z,n) > 0$, implying that firms do not pay dividends (see equation (8)). In this case, investment and innovation are substitutes because higher investment must be accompanied by lower innovation for a given level of net worth (or vice-versa). On the other hand, an increase in net worth will increase both investment and innovation, with the associated sensitivities determined by the slope of the expenditure curves in the left panel. Hence, in this region of the pecking order, the firm grows both through capital accumulation and through the potential realization of successful innovations.

Finally, for sufficiently high levels of net worth, the firm enters the last region of the pecking order in which it becomes financially unconstrained. Conditional on its value of productivity z, a firm in this region has reached its optimal scale given its current level of productivity, $k_t^*(z)$. At this point, the financial wedge $\lambda_t(z, n) = 0$, driving the net returns to investment and innovation to zero and implying that firm's policies become independent of net worth. In this case, the only way in which the firm will grow further is the realization a successful innovation. If this happens, the firm's productivity z will jump up, at which point both returns will also jump up and the firm may re-enter the dynamics described above.

This discussion illustrates how our model is consistent with empirical patterns of investment and innovation that we documented in Section 2. First, firms tend to enter the economy with a new idea but less capital than the implied optimal scale, $k < k_t^*(z)$. This initial condition places most new entrants in the first region of the pecking order in which they start growing only through investment. Second, established firms who receive a new idea from a successful innovation will similarly enter a situation in which their current capital stock is below their new, higher optimal scale $k < k_t^*(z)$. These firms will again enter an earlier region of the pecking order and prioritize investment before innovating again. Although we presented our baseline empirical patterns as a function of the firm's capital stock, Appendix A confirms the same patterns hold using net worth, which is the model's state variable in these figures.

This discussion also shows that our model matches two stylized facts about scale-dependent firm growth in the data. First, employment and sales of small firms grow faster than large firms (see, e.g., Akcigit and Kerr (2018)); this occurs in our model because small firms have a high marginal product of capital in the early stages of the pecking order. Second, the average growth rates of large firms are roughly constant, which is known as Gibrat's law; this occurs in our model because all unconstrained firms choose the same innovation intensity i_{jt} and therefore have the same probability of receiving a new idea (see the plots in Appendix C). The fact that Gibrat's law does not hold for all firms in our model complicates aggregation relative to the typical endogenous growth model. In particular, we need to keep track of the entire distribution of firms in order to compute the model; see Appendix C for details.

Role of Key Parameters While in principle all parameters shape the pecking order of firm growth, two sets of parameters are particularly important in practice: the degree of financial frictions and the efficiency of the innovation technology.

Figure 3 shows that, without financial frictions, the model would not have a pecking order at all; firms would immediately be able to leverage up to their optimal scale given current productivity. In this case, investment and innovation become independent of net worth, size, and age, inconsistent with the evidence presented in Section 2. Hence, financial frictions are the key model ingredient which allows us to be consistent with that evidence.[15]

While some form of financial constraints are necessary to generate the pecking order in our model, the precise form we've chosen here is not. In general, financial constraints play two roles in our model. First, they imply that small firms have a higher shadow value of funds $1 + \lambda_t(z, n)$ and, therefore, face a higher marginal cost spending resources. Any financial constraint with this feature will imply that small firms face a high marginal product of capital, generating a high return to capital. Second, the financial constraints determine the collateral value of either investment or innovation, which may affect the return on either activity. In our calibrated model, the collateral value component of the return to capital turns out to be fairly small relative to the marginal product itself. While alternative specifications of financial constraints may shift the returns to innovation or investment on the margin, they are unlikely to overturn the pecking order itself.[16]

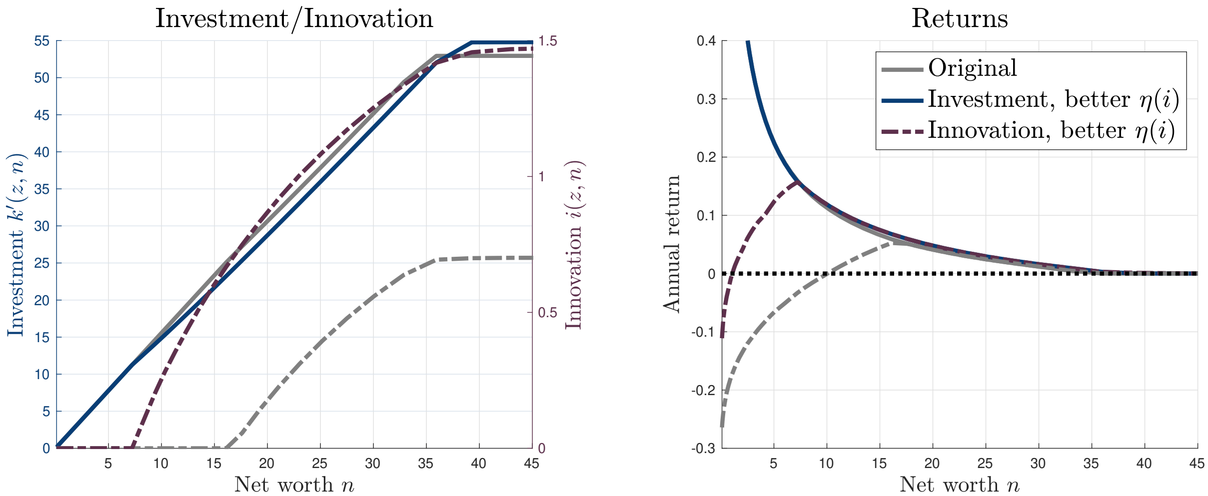
[15]Other sources of adjustment frictions, such as capital adjustment costs, customer base accumulation, or learning how to implement new ideas may also match pecking order of firm growth. For example, in unreported results, we find that capital adjustment costs match the pecking order because the frictions keep small firms' current capital k below their optimal scale $k^*(z)$. We view these alternative frictions are complementary explanations for the pecking order. We focus on financial frictions given the empirical importance of net worth documented in Section 2. Financial frictions uniquely imply that investment and innovation are substitutes for constrained firms, which will be important for our policy analysis.

[16]Some forms of financial constraints may imply that ideas are partially collateralizable, which would shift up the return to innovation. For example, consider an earnings-based constraint in the spirit of Lian and Ma (2021) and Greenwald et al. (2019):

$$b_{jt+1} \leq \widetilde{\theta}(A_{t+1}z_{jt+1})^{1-\alpha}k_{jt+1}^{\alpha}, \tag{9}$$

In this specification, ideas are partially collateralizable because they are reflected in cash flows. This property also amplifies the positive externality of innovation relative to our baseline model because new ideas directly

FIGURE 4: Role of Innovation Technology in the Pecking Order



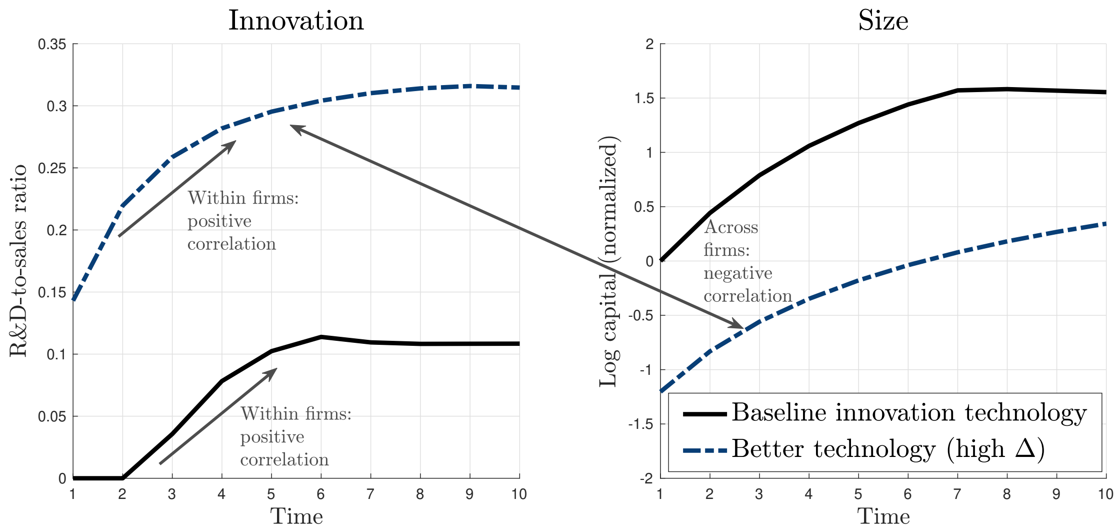
Notes: the left panel plots capital expenditures $k_t(z,n)$ (left axis) and innovation expenditures $i_t(z,n)$ (right axis) in market equilibrium BGP of the calibrated model for fixed z. The right panel plots the return to these activities, defined as the RHS of the Euler equations (6) and (7) minus 1. "Better $\eta(i)$" refers to the model with higher η_0 than in our baseline calibration (see Section 5), but using the same real interest rate from the market BGP.

Figure 4 illustrates the effects of a more efficient innovation technology, which raises the success probability $\eta(i)$ for any level of innovation i. The higher success probability shifts up the returns to innovation, which implies that it intersects the returns to capital at a lower level of net worth. Therefore, firms begin innovating and enter the second region of the pecking order for lower levels of net worth than in the baseline. The firm's innovation rate is also higher in this region, which forces it to slightly reduce its capital expenditures given the flow of funds constraint (8). However, once the firm becomes unconstrained, its level of capital accumulation is higher than in the baseline scenario because higher innovation increases the marginal product of capital.

Interpreting Existing Literature Through Our Model In Section 2, we discussed how existing empirical studies tend to find that small firms do more innovation, in contrast with our pecking order of firm growth. We also showed that, within our sample, the existing literature's results are driven by fixed differences across firms (which are not the focus our

shift out everyone's borrowing constraint (the right-hand side of (9) is increasing in A_{t+1}).

FIGURE 5: Allowing For Differences in Innovation Success Across Firms



Notes: figure plots time path of R&D-to-sales ratios (left panel) and detrended log capital stock (right panel) for two types of firms: one with the baseline innovation technology, and one with a better technology with higher step size Δ but starts with a smaller log capital stock. Log capital in right panel is normalized to zero in the first year for the firm with the baseline technology.

analysis). However, we cannot rule out the possibility that the newest firms in the economy are not more innovation-intensive before they enter our sample. For example, new firms may have have a higher technological return to innovation, leading them to do more innovation.

In this subsection, we show how our model can accommodate this possibility without affecting the key tradeoff between investment and innovation in our model. We illustrate this result by comparing the time paths of two different types of firms. The first firm is representative of a small incumbent firm in our Compustat sample; it has our calibrated innovation technology and starts with a level of capital similar to a smaller firm within Compustat (we discuss how we model the selection into Compustat in Section 5). The second firm is representative of a new entrant with a higher return to innovation than incumbents; its innovation technology has a larger increase in productivity upon success Δ, but the firm starts with a lower level of capital. We then simulate the time paths of the R&D-to-sales ratio and the capital stock for these two firms over a ten year period.

Figure 5 shows that mechanism implies that innovation is negatively correlated with size across firms, as in the existing literature, but positively correlated with size within firms, as in our pecking order of firm growth. The across-firm correlation is negative because the smaller firm has a higher technological return to innovation Δ, and therefore pursues more innovation and less investment for any level of net worth. However, the within-firm correlation is still positive because firms are more willing and able to finance higher innovation as they accumulate capital. Hence, our pecking order of firm growth still applies in this extended model with heterogeneous returns to innovation Δ. Given this robustness, we abstract from heterogeneity in the technological return to innovation Δ for the rest of our paper.[17]

5 Parameterization

We now calibrate the model to ensure it is in line with key features of the data. Section 5.1 describes the moments we use to discipline the key parameters governing the pecking order described above. Section 5.2 uses these moments, and others, to calibrate the model. Finally, Section 5.3 shows that the model matches various untargeted statistics, including the response of innovation to investment tax shocks.

5.1 Strategy for Disciplining Key Forces

Following much of the literature, we will choose the tightness of the collateral constraint θ to match the average leverage of firms in the data; we validate this choice using the response of innovation to investment tax shocks in Section 5.3. Therefore, the main challenge in our calibration is to pin down the properties of the innovation technology: (i) the probability of a successful innovation $\eta(i)$ and (ii) the size of successful innovations Δ.

While we can arguably measure innovation in the data using R&D expenditures, there is no direct measure of its output (successful innovations).[18] Given these difficulties, our ap-

[17] Akcigit and Kerr (2018) develop a model without capital but with two types of innovation activities that differ in the size of their success, which loosely maps into heterogeneity in our Δ parameter. In their model, innovation declines in firm size because firms endogenously choose to do the lower-Δ type of innovation as they grow. Akcigit and Kerr (2018) discipline the strength of this force using the fact that small firms' employment grow faster than large firms. Our model also implies that small firms grow faster than large firms, but because of the diminishing marginal product of capital rather than the innovation technology. We therefore view our mechanism and Akcigit and Kerr (2018)'s as complementary channels which may both be operative depending on the type of firm being studied. However, only our mechanism predicts that firm-level innovation rates increase in firm size, as we found in the Compustat and Orbis data.

[18] R&D expenditures may understate true innovation activity, especially for small firms with less devel-

proach is to infer successful innovations from what firms reveal through their forward-looking investment decisions. In our model, firms that receive a successful innovation experience an *investment spike*—a large but short-lived surge in their investment rate — in order to adapt their capital stock to the new, higher level of productivity. Therefore, the responsiveness of investment spikes to R&D expenditures is informative about the innovation technology through the lens of our model.

We study the relationship between investment spikes and R&D expenditures in our Compustat sample. Following Cooper and Haltiwanger (2006), we define investment spikes as years in which a firm's investment rate is above 20%. In our sample, the frequency of investment spikes is 20% and the average size of an investment spike is 31.5%, similar to Cooper and Haltiwanger (2006)'s Census sample. The average R&D-to-sales ratio is 2.9%, and its standard deviation 5.7%.

We estimate the linear probability model

$$\mathbb{1}\{\frac{x_{jt}}{k_{jt}} \geq 0.2\} = \alpha_j + \alpha_{st} + \sum_{h=1}^{H} \beta_h \left(\frac{\widehat{i}_{jt-h}}{\tilde{y}_{jt-h}} \right) + \Gamma' X_{jt} + \epsilon_{jt}, \tag{10}$$

where $\frac{x_{jt}}{k_{jt}}$ denotes the investment rate of firm j in period t; $\frac{\widehat{i}_{jt}}{\tilde{y}_{jt}}$ the R&D-to-sales ratio; α_j and α_{st} firm and time by sector fixed effects; X_{jt} is a vector of firm-level controls; and ϵ_{jt} is a random error term. Our coefficient of interest, β_1, measures how the probability of an investment spike is related to its previous R&D expenditure. The vector of controls X_{jt} includes the ratio of cash flows to lagged capital—to absorb the effect that changes in firms' cash-on-hand has on both investment spikes and R&D expenditure—and two variables that the lumpy investment literature identifies as predictors of investment spikes: the number of years since the previous spike and the capital-employment ratio. Appendix A.1 details the construction of these additional control variables. We set $H = 4$ for our baseline model and explore alternative lags in robustness analysis. We cluster standard errors two ways to account for correlation within firms and within years.

Table 2 shows that, consistent with our model, R&D expenditures are a strong predic-

oped accounting infrastructures. However, this concern is smaller in the context of our Compustat sample, which only includes large, publicly traded companies. In any event, we are not aware of a better-measured alternative to innovation inputs in the data.

Table 2

Investment spikes and Innovation

	(1)	(2)	(3)
$\frac{i_{jt-1}}{\tilde{y}_{jt-1}}$	1.29	1.10	1.12
	(0.16)	(0.15)	(0.15)
$\frac{\text{cf}_{jt}}{k_{jt}}$		0.12	0.11
		(0.02)	(0.02)
years since spike$_{t-1}$			0.003
			(0.0008)
$\frac{k_{jt}}{n_{jt-1}}$			-0.016
			(0.003)
Observations	53,577	53,577	53,577
Adj. R^2	0.261	0.281	0.282

Notes: Results from estimating alternative versions of $\mathbb{1}\{\frac{x_{jt}}{k_{jt}} \geq 0.2\} = \alpha_j + \alpha_{ts} + \sum_{h=1}^{4} \beta_h \left(\frac{\widehat{i}_{jt-h}}{\tilde{y}_{jt-h}}\right) + \Gamma' X_{jt} + \varepsilon_{jt}$, where $\frac{x_{jt}}{k_{jt}}$ denotes the investment rate of firm j in period t; $\frac{\widehat{i}_{jt}}{\tilde{y}_{jt}}$ the R&D-to-sales ratio; α_j and α_{ts} firm and time by sector fixed effects; X_{jt} is a vector of firm-level controls; and ε_{jt} is a random error term. Column (1) reports estimates for a specification without including-firm level time-varying controls; Column (2) those that include cash flows ($\frac{\text{cf}_{jt}}{k_{jt}}$) as a control; and Column (3) those that also include the lumpy-investment controls (years since the last investment spike, years since spike$_{t-1}$, and the standardized capital-output ratio, $\frac{k_{jt}}{n_{jt-1}}$). To estimate the models reported in Columns (1) and (2), we restrict the sample to that with available observations in Column (3). For variable definitions and descriptive statistics, see Appendix A.

tor of investment spikes. Column (1) reports the estimated coefficient β_1 from the linear probability model (10) without any additional controls X_{jt}. Quantitatively, the estimated coefficient implies that have an R&D-to-sales ratio one standard deviation above the mean increases the probability of an investment spike by 7 percentage points, i.e. a nearly 40% increase in the probability of a spike relative to its unconditional mean. Column (2) shows that estimate survives controlling for changes in cash flow which may independently affect both investment and R&D expenditures.

While we admittedly do not have exogenous variation in R&D expenditures to identify the causal effect of innovation on investment spikes, we view these results are suggesting a tight link between the two. Therefore, we will target the estimated coefficient β_1 in our model calibration by running the same regression on model-simulated data.

One may be concerned that investment spikes in the data are driven by fixed capital adjustment costs, not the arrival of new ideas as in our model. We address this concern in three ways. First, we target the passthrough of R&D expenditures to investment spikes, not the overall frequency of investment spikes. Second, fixed costs most naturally occur at the unit of the plant or even production line, and we note that Compustat firms aggregate over many such units. Finally, Column (3) in Table 2 shows that our regression coefficient is virtually unaffected by controlling for the years since the last spike and the capital to labor ratio, which are important summaries of the incentives to invest in fixed cost models.

Appendix A.4 presents robustness analysis and additional supportive evidence about the relationship between innovation and investment spikes. In particular, we show that the results presented in Table 10 are robust alternative model specifications: an alternative definition of investment spikes (that considers a sector-level threshold instead of an absolute threshold), alternative lags of R&D-to-sales ratios, and additional controls used in the investment literature (e.g., size, sales growth, and the share of current assets). We also present a complementary event study which shows that R&D-to-sales tend to increase before investment spikes.

5.2 Calibration

With this evidence in hand, we now calibrate the model. We proceed in two steps. First, we fix a subset of parameters to match standard aggregate targets. Second, we choose the remaining parameters to match moments in the data.

Table 3 contains the parameters that we exogenously fix. We set the EIS $1/\sigma = 1.5$, in line with estimates from the finance literature. We make this choice because the main role of the EIS in our analysis is to control movements in the real interest rate. Changes in the real interest rate are very powerful in our model because unconstrained firms face no other adjustment costs in our model. Given this value of the EIS, we set the household's discount factor $\beta = 0.97$ so that the real interest rate is 4% annually along the BGP. We set the elasticity of output with respect to inputs to be $\alpha = 0.55$, close to the 0.59 value Cooper and Haltiwanger (2006) estimate for manufacturing plants.[19] We set the depreciation rate to $\delta = 8\%$ annually to imply an aggregate investment-to-capital ratio of 10% along the BGP.

[19]In the model with labor discussed in Footnote 9, our choice of α is consistent with an underlying production function in which the labor share is 2/3 and the total returns to scale is 0.85.

Table 3
Fixed Parameters

Parameter	Description	Value
Household		
β	Discount factor	0.97
$1/\sigma$	EIS	1.50
Firms		
α	Output elasticity w.r.t inputs	0.55
δ	Depreciation rate	0.08
π_d	Exit rate	0.08

Notes: parameters chosen exogenously to match external targets.

Finally, we assume $\pi_d = 8\%$ of firms exit per year, broadly consistent with exit rates in both the Business Dynamics Statistics (BDS) and in our Compustat sample.

Table 4 contains the endogenously chosen parameters and the moments we target in the data. For targets drawn from Compustat data, we mirror the sample selection into Compustat by conditioning on firms that have survived at least seven years, as in Ottonello and Winberry (2020). The first three parameters govern the innovation technology. We assume that the probability of success is given by $\eta(i_{jt}) = 1 - e^{-\eta_0 i_{jt}}$, so that η_0 controls the efficiency of the success probability. The parameter Δ is the size of successful innovations, and a controls the non-rivalry of ideas in A_t. The corresponding first three moments in the right panel of the table contain the intuitively strongest source of identifying variation in the data (though of course all parameters are jointly chosen to match all targets). As discussed above, the regression coefficient of the probability of investment spikes on lagged R&D spending from Table 2 pins down the efficiency of the success probability $\eta(i)$. The average size of the resulting investment spikes then pins down the size of successful innovations Δ. Given this innovation technology, we then infer the degree of non-rivalry of ideas among incumbent firms, a, to match a long-run growth rate of 2% per year.

The most natural point of comparison for our estimated innovation technology is to the empirical literature on the response of patenting to R&D spending, which is often used to discipline creative destruction models (see, e.g., Acemoglu et al. (2018)). These studies typically find an average elasticity of successful innovation to R&D around 0.5; our estimates imply an average elasticity 0.49, which is very close to the literature.

Table 4
Fitted Parameters

Parameter	Description	Value	Target (all joint)	Data	Model
Innovation technology					
η_0	Efficiency	0.19	Regression coefficient	1.10	1.08
Δ	Size of innovations	0.22	$\mathbb{E}[x_{jt}/k_{jt}\vert\text{spike}]$	0.32	0.32
a	Non-rivalry of ideas	0.40	Growth rate	0.02	0.02
Financial frictions					
θ	Collateral constraint	0.39	$\mathbb{E}[b_{jt}/k_{jt}]$	0.13	0.15
Productivity shocks					
σ_ε	SD of shocks	0.01	$\sigma(x_{jt}/k_{jt})$	0.13	0.10

Notes: left panel contains the parameters chosen to match the moments in the right table. "Regression coefficient" is the regression coefficient on lagged R&D-to-sales in Table 2 column (1). $\mathbb{E}[x_{jt}/k_{jt}|\text{spike}]$ is the average size of investment spikes in our Compustat sample described in Section 5.1. "Growth rate" is the aggregate growth rate along the market equilibrium BGP. $\mathbb{E}[b_{jt}/k_{jt}]$ is the average (net) leverage of firms in our sample. Finally, $\sigma(x_{jt}/k_{jt})$ is the standard deviation of investment rates in our sample.

The remaining parameters are pinned down by standard moments. We choose the degree of financial frictions θ in order to match the average (net) leverage of firms in our Compustat sample. While this target may be problematic given that we do not model the tax advantage of debt, we show in the next section that it implies realistic heterogeneity in the response of innovation to investment tax shocks. We choose the dispersion of idiosyncratic shocks σ_ε to match the dispersion of investment rates in the data.

Finally, as discussed in Section 3, we choose the distribution of new entrants to capture the idea that new entrants imitate incumbent firms. We assume that productivity is drawn from log-normal distributions whose mean equals the mean of the distribution of incumbent firms, and set the dispersion in those draws to $\sigma_z = \Delta$. We assume that entrants' initial capital stock is roughly 3% of average incumbents' capital; this choice would imply that new entrants' employment is approximately 10% of the average firm, in line with the data (see the discussion in Khan and Thomas (2013)). We choose the dispersion of new entrants' capital to match the relative dispersion of new entrants to incumbents (roughly 20%).[20]

[20] While this entry process is representative for the typical firm, it may not capture "revolutionary entrants" who enter the with a substantially better idea than other firms in the economy. These revolutionary entrants can be thought of as coming from another distribution of initial state variables with very high productivity z. Financial frictions would be even more binding on these firms because their optimal scale of capital would be even further above their initial capital stock.

5.3 Validation

We now show that the calibrated model matches untargeted statistics in the data. Importantly, the model matches new evidence on the response of innovation to exogenous changes in the after-tax price of investment across the firm size distribution, validating the strength of financial frictions in driving innovation decisions in our model.

Sources of Firm Heterogeneity Appendix E analyzes the two sources of firm heterogeneity in our calibrated model: lifecycle dynamics and productivity differences (due to either successful innovations or productivity shocks). Following the pecking order of firm growth from Section 4, most young firms start investment-intensive but become more innovation-intensive as they age. Increases in productivity raise the marginal product of capital and shadow value of funds $1 + \lambda_t(z, n)$, which induces firms to invest and borrow more but innovate less. These dynamics imply positive investment- and innovation-cash flow sensitivities, as in the data. We also show that the model matches a number of untargeted moments from our Compustat sample.

Investment Tax Shocks To study the response of innovation to changes in the incentives to invest, we exploit variation in the after-tax price of investment induced by the Bonus Depreciation Allowance. The Bonus is a countercyclical investment stimulus policy used in the 2001 and 2008 recessions. The policy allowed firms to deduct a fraction $b_t \in [0, 1]$ of investment expenses from their tax bill immediately (and apply the standard depreciation schedule to the remaining $1 - b_t$ fraction of expenditures). By bringing forward future tax deductions into the present, the policy increases the present value of tax deductions by $\Delta\zeta_t = b_t(1 - \zeta)$ where $\zeta < 1$ is the present value of deductions under the baseline schedule.

Zwick and Mahon (2017) show that sectoral heterogeneity in the baseline tax depreciation schedule across sectors, ζ_s, provides exogenous variation that can be used to identify the effect of the Bonus, $\Delta\zeta_{st} = b_t(1 - \zeta_s)$, on investment. Table 5 Column (1) replicates Zwick and Mahon (2017)'s estimates of the effect of the Bonus on investment in our Compustat sample.

TABLE 5
BONUS DEPRECIATION ALLOWANCE IN THE DATA AND THE MODEL

	(1) $\frac{x_{jt}}{k_{jt}}$, data	(2) $\frac{x_{jt}}{k_{jt}}$, model	(3) $\frac{i_{jt}}{y_{jt}}$, data	(4) $\frac{i_{jt}}{y_{jt}}$, model	(5) $\frac{i_{jt}}{y_{jt}}$, small	(6) $\frac{i_{jt}}{y_{jt}}$, large
$\frac{1-\tau PV_{st}}{1-\tau}$	-1.41	-1.68	-0.24	-0.33	-0.53	-0.13
	(0.25)		(0.06)		(0.23)	(0.05)
R^2	0.41		0.89		0.85	0.91

Notes: estimates of $\widehat{\gamma}$ from the regression (11) in columns (1) and (2) or from the regression (12) in columns (3) - (6). Standard errors are two-way clustered by firms and years. "Model" columns (2) and (4) replicate the regressions on model-simulated data in response to a shock equivalent to a 50% bonus depreciation allowance which reverts back to its long-run average following an AR(1) process with annual persistence 0.75 (giving a half-life of roughly two years). "Small" firms in column (5) are those whose average sales are in the bottom 3 deciles of the sales distribution. "Large" firms in column (6) have average sales in the top 3 deciles of the sales distribution.

Specifically, we estimate the regression

$$\frac{x_{jt}}{k_{jt}} = \alpha_i + \alpha_t + \gamma \frac{1 - \tau\zeta_{st}}{1 - \tau} + \Gamma' X_{jt} + \epsilon_{jt}, \tag{11}$$

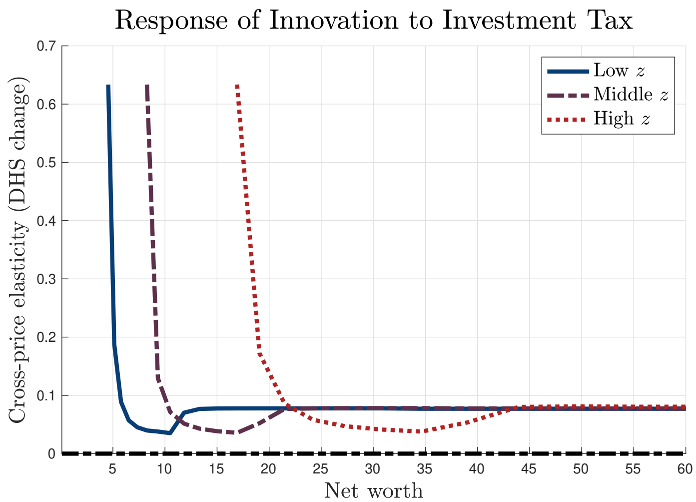
where τ is the corporate tax rate against which firms deduct investment expenditures, α_i is a firm fixed effect, α_t is a time fixed effect, X_{jt} controls for cash flows to lagged capital, and ϵ_{jt} are residuals.

In the data, we estimate the regression coefficient on investment $\widehat{\gamma} = -1.41$, which is close to Zwick and Mahon (2017)'s estimate of -1.53 using firm-level IRS microdata. A 50% bonus would increase increased the average value of $\frac{1-\tau\zeta_{st}}{1-\tau}$ by -0.03, implying its direct effect increased the average firm's investment rate by $-0.03 \times -1.41 = 0.04$, compared to its unconditional average of 0.14.

We replicate the Bonus Depreciation Allowance in our model by feeding in an exogenous shock to the relative price of investment of size $\tau\Delta\zeta_t = b_t(1-\zeta)$. Appendix D shows that the Bonus is isomorphic to a temporary shock to the relative price of capital in our model. We assume that the shock mean-reverts according to an AR(1) with an annual autocorrelation coefficient of 0.8. This coefficient is broadly in line with the data and, as we show below, implies a similar response of investment in our model.

In order to compare the model to the data, we simulate a panel of firms and estimate the regression equation (11). In this regression, we assume all firms face the same present value

Figure 6: Heterogeneous Responses to the Bonus Depreciation Allowance



Notes: cross-price elasticity $\frac{\partial \log i(z,n)}{\partial \log(1-\zeta_t)}$ in response to a shock equivalent to a 50% bonus depreciation allowance which reverts back to its long-run average following an AR(1) process with annual persistence 0.75 (giving a half-life of roughly two years). Elasticities computed in impact period of the shock.

of tax deductions ζ_t, i.e. there is no sectoral heterogeneity. Since the empirical specification (12) includes time fixed effects which absorb general equilibrium effects, we keep the real interest rate fixed at its initial value $r_t = r^*$ for this exercise. We do not include controls X_{jt} that are outside of our model. As in our calibration, we mirror the sample selection into Compustat using firm age. Column (2) in Table 5 shows that the model's implied regression coefficient is $\widehat{\gamma} = -1.51$, very close to the data.

Column (3) in Table 5 documents a new empirical finding: the Bonus also substantially raises innovation expenditures. We estimate the regression

$$\frac{\widehat{i}_{jt}}{\widetilde{y}_{jt}} = \alpha_i + \alpha_t + \gamma \frac{1 - \tau \widehat{\zeta}_{st}}{1 - \tau} + \Gamma' X_{jt} + \epsilon_{jt}, \tag{12}$$

which replaces the investment rate on the LHS of (11) with the RD-to-sales ratio $\widehat{i}_{jt}/\widetilde{y}_{jt}$. Note that the denominator $\widetilde{y}_{jt}$ is lagged sales in the past five years, so it is predetermined in the period of the shock. Quantitatively, this estimated coefficient implies that a 50% bonus directly raises the average firm's RD-to-sales ratio by about 0.8pp relative to its unconditional average of 2.9pp — a nearly 30% increase in innovation expenditures. Column (4) shows that our model matches the empirical response of innovation to the bonus well.

In order to understand the role of financial frictions in driving the model's success, Figure 6 shows that the model's cross-price elasticity of innovation with respect to investment $\frac{\partial \log i(z,n)}{\partial \log(1-\tau\zeta_t)}$ is positive but heterogeneous across firms. Unconstrained firms have a positive elasticity because higher investment also raises the return to innovation due to the complementarity between capital and productivity. On the other hand, constrained firms have a positive elasticity because the shock lowers their after-tax expenditures on investment, freeing up cash flows to finance innovation. Quantitatively, this cash flow channel is larger than the complementarity channel for most constrained firms.

Columns (5) and (6) in Table 5 confirm these size-dependent responses are consistent with the data, providing further validation of the role of financial frictions in linking innovation and investment. Following Zwick and Mahon (2017), we define small firms as those whose average sales are in the bottom three deciles of the distribution and large firms whose sales are in the top three deciles. Small firms' innovation expenditures are about five times as responsive to the bonus as are large firms, consistent with Figure 6 in our model.

6 The Aggregate Effects of Financial Frictions

We now use our calibrated model to assess the quantitative implications of financial frictions for the macroeconomy. In Section 4, we showed that financial frictions delay the point at which individual firms begin innovating. Aggregating across firms, this fact implies that there will be less innovation expenditures in the aggregate. Lower innovation expenditures then lowers the long-run growth rate g^* through:

$$g^* \approx (1+a)(e^{\Delta}-1)\int \eta(i_{jt})dj. \tag{13}$$

The aggregate growth rate is approximately equal to the number of new ideas, $\int \eta(i_{jt})dj$, times the productivity improvement from each new idea, $e^{\Delta}-1$, times the positive spillovers, $1+a$.

Our goal in this section is to quantify these negative growth effects of financial frictions.

TABLE 6

ROLE OF FINANCIAL FRICTIONS IN LONG-RUN GROWTH

		TFP losses after			
	Growth rate g	20 years	30 years	40 years	50 years
Calibrated model	2.00%				
Frictionless model	2.45%	9%	14%	19%	25%

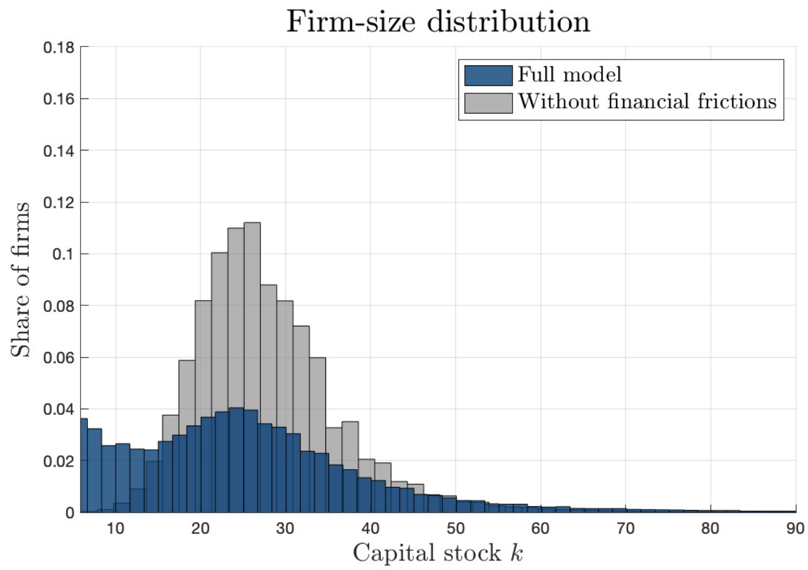
Notes: cross-sectional statistics from the stationary distribution in the market equilibrium BGP. "Calibrated model" refers to full model calibrated as in main text. "Frictionless model" refers to model in which all firms follow the unconstrained policies $k^*(z)$ and $i^*(z)$ from Proposition 1. "TFP losses" are equal to the cumulated growth rate in the frictionless model relative to the cumulated growth rate in the calibrated model.

In order to do so, we compare our calibrated model to the *frictionless model* in which there are no financial frictions. In this alternative model, all firms follow the unconstrained policies $k^*(z)$ and $i^*(z)$ from Proposition 1.

Table 6 shows that removing financial frictions would raise the aggregate growth rate by 45 basis points per year. Cumulated over long horizons, this difference implies substantial TFP losses from lower innovation; for example, after forty years, financial frictions have reduced aggregate TFP by nearly 20%. Hence, an important cost of financial frictions on the economy is lower economic growth. This finding complements the existing literature which studies how financial frictions may distort the allocation of capital and, therefore, lower the level of TFP (see e.g. Buera, Kaboski and Shin (2011) or Midrigan and Xu (2014)). While that misallocation also occurs in our model, over long horizons the growth effects account for the majority of the costs of financial frictions in our model.

Looking across firms, financial frictions depress innovation primarily in small, financially constrained firms for whom the return to capital is relatively high. For example, in our calibrated model, approximately 80% of economic growth comes from the top 20% of firms. This force thickens the right tail of the firm size distribution relative its mode. We illustrate this mechanism in Figure 7, which compares the distribution of detrended capital stocks in the two BGPs; given the differences in growth rates, direct comparisons across the two economies are not valid, but comparisons within each economy are still meaningful. From this perspective, the size distribution in our full model has more mass in both the left and right tails than does the distribution without financial frictions. The thickness of the left tail reflects the fact that it takes new entrants longer to grow, while the thickness of the right

Figure 7: Firm-Size Distribution



Notes: distribution of capital along the balanced growth path. Capital stocks have been detrended in order to compute a stationary distribution, but the resulting distribution has the same cross-sectional properties as the raw distribution (see Appendix B). "Full model" refers to our calibrated model. "Without financial frictions" refers to the version of the model in which firms follow the unconstrained policies $k^*(z)$ and $i^*(z)$ from Proposition 1.

tail reflects the fact that unconstrained firms who survive follow a random growth process with exogenous death.[21]

Because most innovation is done by financially unconstrained firms, Appendix E shows that temporary financial shocks θ_t do not have a particularly persistent effect in our model (despite the sizeable effects of *permanent* differences in θ described above). This result contrasts with the stylized fact that financial shocks have more persistent negative economic effects in the data (e.g., Cerra and Saxena, 2008). Based on representative firm models, some have argued that this persistence is driven by tighter financial constraints reducing innovation and therefore medium-term growth. However, in our heterogeneous firm model,

[21] In fact, these random growth with death dynamics in the right tail generate a Pareto tail (see, e.g., Jones and Kim (2018)). Unfortunately, the model's tail is thinner than in the data because the expected size of successful innovations must be relatively small to match the average size of investment spikes in the data. However, we can thicken the tail by incorporating heterogeneity in the size of successful innovations, which would create heterogeneity in the expected growth rates (again in the spirit of Jones and Kim (2018)). In this extension, the average of these growth rates would still be pinned down by the average size of investment spikes, but the thickness of the right tail would be driven by firms with higher realized growth.

the majority of innovation at a given time is performed by unconstrained firms, as described above. These firms are not directly affected by the shock and therefore face no impulse to lower innovation; in fact, since the real interest rate falls in general equilibrium, aggregate innovation *rises* following the shock.

7 Policy Implications

The equilibrium studied above will generally not be socially efficient because firms do not internalize the non-rivalry of the ideas they produce, motivating policy intervention. In order to better understand the implications of this externality, Section 7.1 studies the problem of a constrained-efficient planner's problem who internalizes the externality. Section 7.2 uses these results to study how two commonly-used policies, innovation subsidies and investment tax cuts, address the non-rivalry of ideas.

7.1 Financial Frictions and the Non-Rivalry of Ideas

We characterize the problem of a constrained-efficient social planner who faces the same financial constraints as private firms but internalizes the non-rivalry of ideas.[22] In principle, this planner may also want to change the private allocation due to pecuniary externalities through the real interest rate (which may affect welfare due to market incompleteness). Appendix F shows how incorporating these pecuniary externalities would affect our normative results. We exclude them from the main text because they do not affect the long-run choices of the planner and have already been extensively studied in the literature (see, for example, Geanakoplos and Polemarchakis, 1986; Lorenzoni, 2008; Dávila and Korinek, 2018).

Planner's Problem Appendix B formulates the planner's problem recursively. The problem is technically challenging because the state variable is the entire distribution of firms and the control variables are entire functions of the firms' individual states. We overcome

[22]The constrained-efficient approach is common in models with incomplete markets because it does not endow the planner with the power to arbitrarily complete markets in a way that the private sector cannot. Conversely, this approach takes as given that the planner cannot impact whatever underlying frictions lead to these missing markets. To the extent that the planner can indeed alleviate those frictions, our results provide a lower bound for the true welfare gains from the optimal policy.

this challenge by solving the planner's problem in the function space following Lucas and Moll (2014). We arrive at the following natural characterization of the planner's allocation:

***Proposition* 2.** *In the constrained-efficient equilibrium, individual allocations solve the augmented Bellman equation*

$$\omega_t^{cont}(z,n) = \max_{k',b',i} n - k' - A_t z i + \frac{b'}{1+r_t} + \Lambda_t z + \frac{1}{1+r_t}\mathbb{E}_t[\omega_{t+1}(z',n')] \ \ s.t.\ \ d \geq 0 \ \ and \ \ b' \leq \theta k' \tag{14}$$

where Λ_t is the planner's shadow value of the non-rivalry externality:

$$\Lambda_t = \left[a \left(\int z_{jt} dj \right)^{a-1} \right] \times \left[\int (1+\lambda_{jt}) \left((1-\alpha) A_t^{-\alpha} z_{jt}^{1-\alpha} k_{jt}^{\alpha} - z_{jt} i_{jt} \right) dj \right]. \tag{15}$$

Proof. See Appendix B. ∎

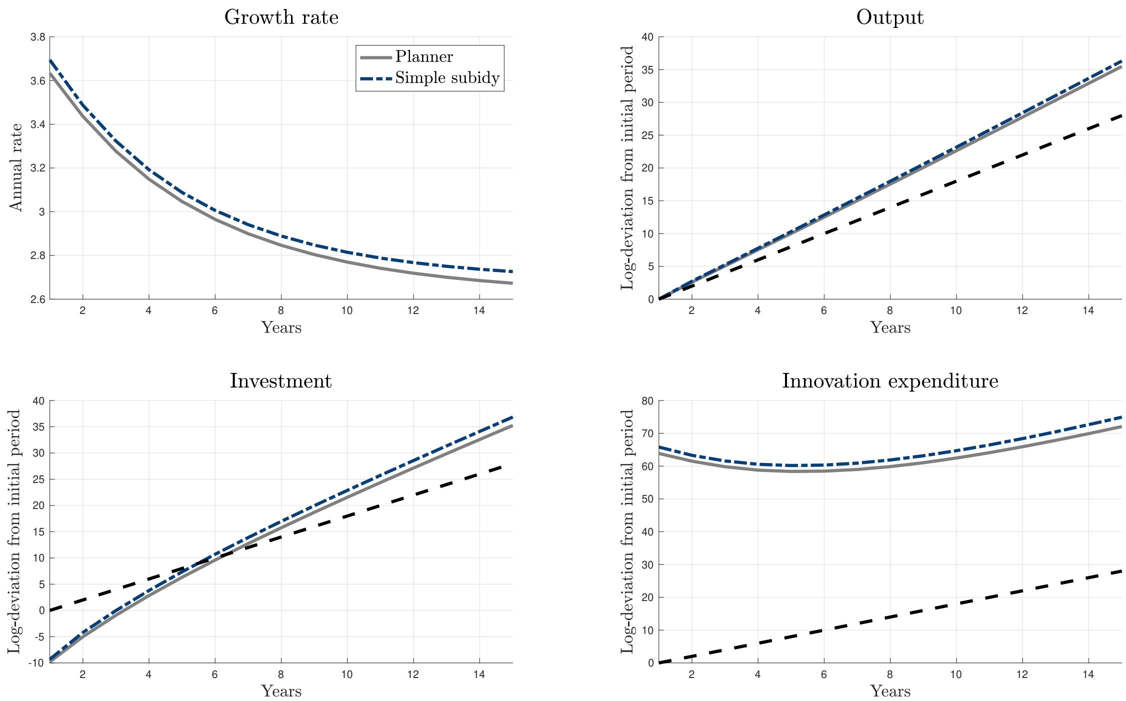
The only difference between the private Bellman equation (5) and the planner's augmented Bellman equation (14) is the shadow value of the non-rivalry externality, Λ_t. Equation (15) shows that this shadow value is the product of two terms: the marginal impact of an individual firm's productivity, z_{jt}, on aggregate productivity times the marginal social benefit of higher aggregate productivity.[23] This object is itself an integral of a product of two firm-level objects: the marginal increase in production net of innovation costs, $(1-\alpha)A_t^{-\alpha} z_{jt}^{1-\alpha} k_{jt}^{\alpha} - z_{jt} i_{jt}$, times the firms' shadow value of funds, $1+\lambda_{jt}$.

Proposition 2 shows that financial frictions amplify the positive externality from the non-rivalry of ideas; all else equal, a higher shadow value of funds $1+\lambda_{jt}$ increases the social value of higher aggregate productivity through the product in (15). This amplification occurs because higher production raises firms' cash flows and therefore alleviates the no-equity issuance constraint, which the planner values at the shadow price λ_{jt}.

Planner's Allocation The characterization of the planner's allocation in Proposition 2 contains two important implications that successful policies must address. First, as is com-

[23] Consistent with our focus on incumbent firms, we also assume that the planner takes as given the distribution of new entrants, i.e. does not take into the positive externality through imitation. We make this simplifying assumption because we take the entry process into our model as exogenous; incorporating this margin would only further increase the positive externality of innovation.

Figure 8: Aggregate Transition Paths in Constrained-Efficient Allocation



Notes: aggregate transition paths chosen by planner (grey lines) and generated by the simple 13% innovation subsidy (dashed blue lines). Growth rate in top rate is in percentage points per year. Aggregate output, investment, and innovation expenditures in the remaining panels are in log-deviations from initial period. Dashed black lines are the growth trajectory in the initial market BGP.

mon in the literature, the planner prefers higher innovation because private firms do not internalize the positive externalities of their innovation ($\Lambda_t > 0$). Second, and more subtle, is the fact that the planner faces a tradeoff in terms of investment. On the one hand, higher innovation requires less investment for constrained firms due to their flow-of-funds constraint, i.e. investment and innovation are substitutes for constrained firms. On the other hand, higher innovation incentivizes more investment for unconstrained firms due to the complementarity between TFP and capital in production, i.e. investment and innovation are complements for unconstrained firms. In order to characterize this tradeoff, we solve for the transition path chosen by the planner starting from the equilibrium BGP.

Figure 8 shows that the planner's balance between the substitutability vs. complementarity of innovation and investment changes over the course of the transition. Early on, the substitutability dominates in the sense that aggregate investment falls. This result occurs for two reasons. First, more firms are financially constrained early in the transition, implying more firms are in the substitutable region of the state space illustrated above. Second, the planner requires especially high innovation early on in the transition, implying constrained firms need to substantially reduce their investment. The planner values high innovation early on because more firms are constrained, which amplifies the planner's shadow value of the non-rivalry externality Λ_t as described above.

Over time, the complementarity between investment and innovation begins to dominate in the sense that aggregate investment eventually increases. This occurs because higher innovation raises net worth, implying that more firms are unconstrained and therefore in the complementary region of the state space. In addition, the planner's desired innovation falls over time as the shadow value of the externality falls as well.

7.2 Evaluating Innovation Subsidies and Investment Tax Cuts

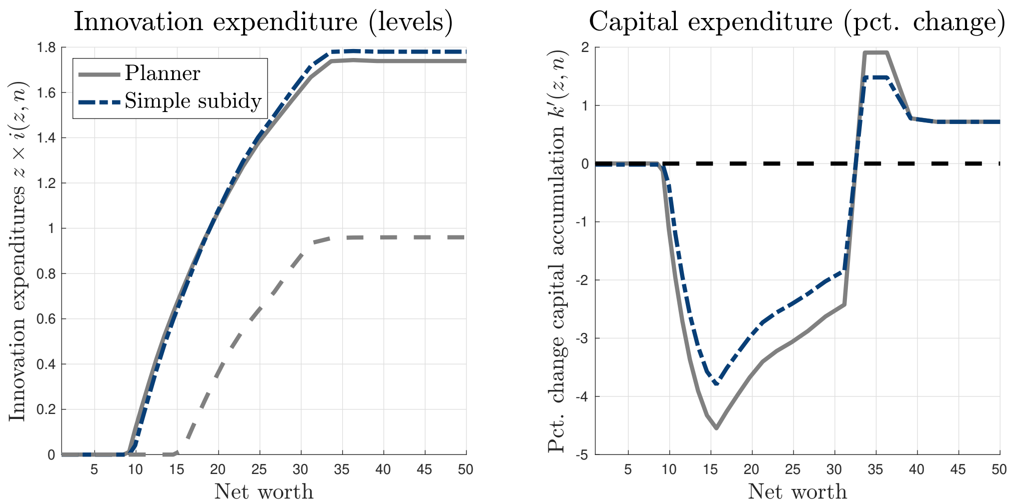
The planner's allocation seems difficult to implement in practice because one has to get both the allocation of innovation *and investment* correct and the relevant tradeoffs vary across both firms and time.[24] In this section, we study how two simple, commonly-used policies perform with respect to these goals: an innovation subsidy and an investment tax cut.

Innovation Subsidy Figure 8 shows a striking finding: a simple innovation subsidy, constant across both firms and time, almost exactly replicates the aggregate paths chosen by the planner. We choose a 26% subsidy to generate the same long-run growth rate in the new BGP with the subsidy as the in the planner's BGP. Despite the fact that the subsidy is constant over time, the economy endogenously "front-loads" innovation early on in the transition, as desired by the planner.

Figure 9 shows that the innovation subsidy even gets the allocation of investment and innovation across firms approximately right. The left panel compares firms' policy rules for innovation as a function of net worth in the market equilibrium, in the planner's allocation,

[24]The planner's augmented Bellman equation (14) suggests one possible implementation: a time-varying transfer to firms proportional to individual productivity. This transfer would have to vary over time to mirror changes in the planner's shadow value Λ_t and vary across firms according to their productivity z_{jt}. Both of these objects are unobservable to policymakers in practice.

FIGURE 9: Planner's Decision Rules vs. Simple Subsidy, Initial Period



Notes: decision rules in the market equilibrium compared with the constrained-efficient allocation (grey lines) or to the policies under the simple innovation subsidy (blue lines). Left panel plots innovation expenditures $A_t z i_t(z,n)$ for a firm with average productivity. Right panel plots the percentage difference in capital accumulation policies.

and under the innovation subsidy. The simple subsidy also replicates the planner's desired changes to the innovation policy at the firm level; constrained firms endogenously increase their innovation by less than unconstrained firms according to the pattern of financial wedges $\lambda_t(z,n)$ across firms. The right panel shows that these firms also cut their capital expenditures in order to finance these higher innovation expenditures. However, the implied change in capital accumulation policies is not as close to the planner's policies as they are for innovation. Hence, it turns out that an appropriately-chosen innovation subsidy performs quite well, but not perfectly, in getting firms to internalize the non-rivalry externality.

Investment Tax Cut We now show that, to the extent that this nearly-optimal innovation subsidy is not fully available in practice, it can be partially substituted by an investment tax cut. Specifically, we find that cutting taxes on investment successfully increases innovation in the long run, but will also suboptimally increase investment in the short run.

We illustrate the connection between investment tax cuts and innovation using the Tax Cuts and Jobs Act (TCJA 2017) as an example. The presence of this tax system changes the after-tax relative price of investment to be $1 - \tau\zeta_t$, where τ is the corporate tax rate and ζ_t is the present value of tax deductions per unit of investment. Before TCJA 2017, firms had to deduct investment expenditures over time according to the MACRS depreciation schedule, implying a present value $\zeta_t = \zeta^* < 1$ smaller than one due to discounting. The TCJA 2017 now allows firms to fully expense investment from their tax bill immediately, implying that $\zeta_t = 1$. We mirror this policy change in our model by studying a permanent decline in the after-tax price of investment by $-\tau(1 - \zeta^*)$.

Figure 10 shows that, in our model, full expensing increases the long-run growth rate by nearly 20bps per year — a 10% increase in the annual growth rate of the economy. This result occurs for two reasons. First, for unconstrained firms, the complementarity of capital and TFP in production implies that the return to innovation increases with investment. Second, if after-tax capital expenditures fall, constrained firms can afford more innovation our of their current cash flows.

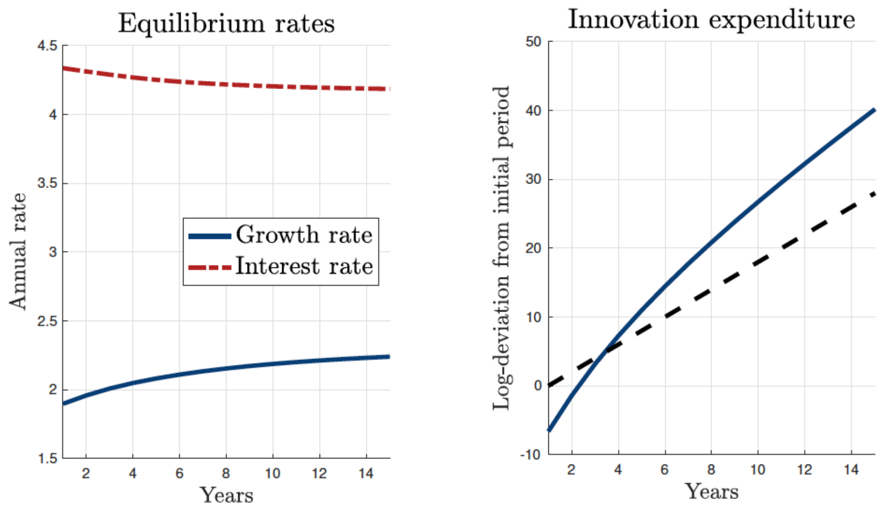
In contrast to our model, investment tax cuts would have no effect on the long-run growth rate in the neoclassical growth model. In the neoclassical model, cutting taxes on investment would increase the capital stock but, due to the diminishing marginal product of capital, that would only lead to an increase in the level of output, not its growth rate.

Figure 10 also shows that — despite raising innovation in the long run — full expensing *lowers* innovation in the first three years after its introduction. Innovation initially falls because the real interest rate r_t rises, lowering the expected returns to innovation. The real interest initially rises because full expensing increases capital demand, but capital supply is partially inelastic in the short run (due to consumption smoothing). Over time, as capital supply catches up to the new long run level of demand, the real interest rate falls and the innovation rate rises. Hence, general equilibrium price effects are important for correctly capturing the dynamic effects of full expensing on innovation over time.

8 Conclusion

In this paper, we have studied the efficiency costs of financial frictions for the macroeconomy. While the existing macroeconomics literature has focused on how financial frictions distort investment decisions and misallocate capital, we focus on how financial frictions distort in-

Figure 10: The Effects of Full Investment Expensing (TCJA 2017)



Notes: transition path following an unexpected, permanent decline in the relative price of capital $1 - \zeta_t$ of the size equivalent to full expensing of investment, starting from the initial market BGP. Dashed lines correspond to the paths of investment, output, and innovation along the initial growth trajectory. Solid lines correspond to their actual paths in response to the change in the relative price of capital. Output and innovation expenditures expressed as log-deviations from initial period.

novation and lower economic growth. We showed these two margins are empirically linked through the pecking order of firm growth. A key contribution of our paper is a new endogenous growth framework with heterogeneous firms and financial frictions that is consistent with this evidence and can be used to draw aggregate implications.

We have purposefully kept our framework as parsimonious as possible in order to focus on the novel mechanisms for our research questions. However, the parsimony of our framework can be leveraged in order to obtain additional insights. For example, it would be natural to characterize the optimal innovation policy given that the private and social returns to innovation differ. We have not done so in this paper because our baseline model only has a positive externality of innovation in the long run, directly implying the efficient allocation requires higher innovation. However, Appendix F extends the model to incorporate a negative pecuniary externality of innovation operating through the labor market. In this extension, innovations from unconstrained firms raise labor demand, which in turn raises labor costs and tightens financial constraints on affected firms. The optimal policy would have to balance the tradeoff between growth (coming from the non-rivalry of ideas) and misallocation (coming from these pecuniary externalities from the wage). Quantifying these insights would require parameterizing the strength of these spillovers and adding an additional state variable to the numerical solution.

Another important extension would relax our assumption that firms cannot sell ideas and instead explicitly model the market for ideas. Given the frictions in this market, it is natural to use a search-and-matching model in the spirit of Lucas and Moll (2014), Perla and Tonetti (2014), or Akcigit, Celik and Greenwood (2016). In this extension, firms would choose between spending some fraction of each time period producing output and the remaining fraction searching in the market for ideas. This extension would provide a third source of firm-level growth, technology adoption. We conjecture that low-productivity firms would be more likely to adopt than innovate because their time cost of foregone output while searching is relatively low and their expected return from matching with other (higher-productivity) firms is relatively high. Conversely, high-productivity firms would be more likely to sell ideas, especially if they are financially constrained and would like to finance capital investment. This extension would also endogenize the non-rivalry of ideas through the composition of idea trades that emerges from the market for ideas.

References

ACEMOGLU, D., U. AKCIGIT, H. ALP, N. BLOOM, AND W. KERR (2018): "Innovation, reallocation, and growth," *American Economic Review*, 108(11), 3450–91.

AGHION, P., AND P. HOWITT (1992): "A Model of Growth Through Creative Destruction," *Econometrica: Journal of the Econometric Society*, pp. 323–351.

AKCIGIT, U., M. A. CELIK, AND J. GREENWOOD (2016): "Buy, keep, or sell: Economic growth and the market for ideas," *Econometrica*, 84(3), 943–984.

AKCIGIT, U., AND W. R. KERR (2018): "Growth through heterogeneous innovations," *Journal of Political Economy*, 126(4), 1374–1443.

ATKESON, A., AND A. T. BURSTEIN (2010): "Innovation, firm dynamics, and international trade," *Journal of political economy*, 118(3), 433–484.

BILAL, A., N. ENGBOM, S. MONGEY, AND G. L. VIOLANTE (2021): "Labor market dynamics when ideas are harder to find," Discussion paper, National Bureau of Economic Research.

BUERA, F. J., J. P. KABOSKI, AND Y. SHIN (2011): "Finance and development: A tale of two sectors," *American economic review*, 101(5), 1964–2002.

CARROLL, C. D. (2006): "The method of endogenous gridpoints for solving dynamic stochastic optimization problems," *Economics letters*, 91(3), 312–320.

CERRA, V., AND S. C. SAXENA (2008): "Growth dynamics: the myth of economic recovery," *American Economic Review*, 98(1), 439–457.

CLEMENTI, G. L., AND B. PALAZZO (2016): "Entry, exit, firm dynamics, and aggregate fluctuations," *American Economic Journal: Macroeconomics*, 8(3), 1–41.

COOPER, R. W., AND J. C. HALTIWANGER (2006): "On the nature of capital adjustment costs," *The Review of Economic Studies*, 73(3), 611–633.

CROUZET, N., J. EBERLY, A. L. EISFELDT, AND D. PAPANIKOLAOU (2022): "Intangible capital, non-rivalry, and growth," *Available at SSRN*.

DÁVILA, E., AND A. KORINEK (2018): "Pecuniary externalities in economies with financial frictions," *The Review of Economic Studies*, 85(1), 352–395.

GARCIA-MACIA, D., C.-T. HSIEH, AND P. J. KLENOW (2019): "How destructive is innovation?," *Econometrica*, 87(5), 1507–1541.

Geanakoplos, J. D., and H. M. Polemarchakis (1986): "Existence, regularity, and constrained suboptimality of competitive allocations," *Essays in Honor of Kenneth J. Arrow: Volume 3, Uncertainty, Information, and Communication*, 3, 65.

Greenwald, D., et al. (2019): *Firm debt covenants and the macroeconomy: The interest coverage channel*, vol. 1. MIT Sloan School of Management.

Grossman, G. M., and E. Helpman (1991): "Quality ladders in the theory of growth," *The review of economic studies*, 58(1), 43–61.

Hopenhayn, H., and R. Rogerson (1993): "Job turnover and policy evaluation: A general equilibrium analysis," *Journal of political Economy*, 101(5), 915–938.

Hopenhayn, H. A. (1992): "Entry, exit, and firm dynamics in long run equilibrium," *Econometrica: Journal of the Econometric Society*, pp. 1127–1150.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 124(4), 1403–1448.

Hurst, E., and B. W. Pugsley (2011): "What do small businesses do?," Discussion paper, National Bureau of Economic Research.

Jones, C. I. (2005): "Growth and ideas," in *Handbook of economic growth*, vol. 1, pp. 1063–1111. Elsevier.

Jones, C. I., and J. Kim (2018): "A Schumpeterian model of top income inequality," *Journal of political Economy*, 126(5), 1785–1826.

Kehrig, M. (2015): "The cyclical nature of the productivity distribution," *Earlier version: US Census Bureau Center for Economic Studies Paper No. CES-WP-11-15*.

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

——— (2013): "Credit shocks and aggregate fluctuations in an economy with production heterogeneity," *Journal of Political Economy*, 121(6), 1055–1107.

Klenow, P. J., and H. Li (2022): "Entry Costs Rise with Growth," .

Klette, T. J., and S. Kortum (2004): "Innovating firms and aggregate innovation," *Journal of political economy*, 112(5), 986–1018.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2017): "Technological innovation, resource allocation, and growth," *The Quarterly Journal of Economics*, 132(2), 665–712.

LIAN, C., AND Y. MA (2021): "Anatomy of corporate borrowing constraints," *The Quarterly Journal of Economics*, 136(1), 229–291.

LORENZONI, G. (2008): "Inefficient credit booms," *The Review of Economic Studies*, 75(3), 809–833.

LUCAS, R. E., AND B. MOLL (2014): "Knowledge growth and the allocation of time," *Journal of Political Economy*, 122(1), 1–51.

LUTTMER, E. G. (2007): "Selection, growth, and the size distribution of firms," *The Quarterly Journal of Economics*, 122(3), 1103–1144.

MELITZ, M. J. (2003): "The impact of trade on intra-industry reallocations and aggregate industry productivity," *econometrica*, 71(6), 1695–1725.

MIDRIGAN, V., AND D. Y. XU (2014): "Finance and misallocation: Evidence from plant-level data," *American economic review*, 104(2), 422–458.

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

NUÑO, G., AND B. MOLL (2018): "Social optima in economies with heterogeneous agents," *Review of Economic Dynamics*, 28, 150–180.

OTTONELLO, P., AND T. WINBERRY (2020): "Financial heterogeneity and the investment channel of monetary policy," *Econometrica*, 88(6), 2473–2502.

PERLA, J., AND C. TONETTI (2014): "Equilibrium imitation and growth," *Journal of Political Economy*, 122(1), 52–76.

RESTUCCIA, D., AND R. ROGERSON (2008): "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic dynamics*, 11(4), 707–720.

WINBERRY, T. (2021): "Lumpy investment, business cycles, and stimulus policy," *American Economic Review*, 111(1), 364–396.

ZWICK, E., AND J. MAHON (2017): "Tax policy and heterogeneous investment behavior," *American Economic Review*, 107(1), 217–248.

A Data Appendix

This appendix provides additional empirical results referenced in the main text.

A.1 Data Construction

Variables For the Compustat sample, we define the variables used in our empirical analysis as follows:

1. *Investment rate*: ratio of capital expenditures (`capx`) to lagged plant, property, and equipment (`ppegt`).

2. *R&D-to-sales*: ratio of research and development expense (`xrd`) to the average of sales (`sale`) in the previous 5 years.

3. *Patents*: Number of patents filed per year (based on the variable `filing_dated`) and market value per patent (based on the variable `xi_real`), constructed from the Kogan et al. (2017) dataset.

4. *Net worth*: defined as sum of plant, property, and equipment and cash and short-term investments (`che`) minus total debt (sum of `dlc` and `dltt`).

5. *Cash flows*: measured as the sum of EBITDA and research and development expense divided by lagged plant, property, and equipment.

6. *Capital-to-employment*: defined at the ratio of lagged plant, property, and equipment to employment (`emp`).

Sample Selection Our empirical analysis excludes:

1. Firms in finance, insurance, and real estate sectors ($\texttt{sic} \in [6000, 6799]$), utilities ($\texttt{sic} \in [4900, 4999]$), nonoperating establishments ($\texttt{sic} = 9995$), and industrial conglomerates ($\texttt{sic} = 9997$).

2. Firms not incorporated in the United States.

Table A.1
Descriptive statistics

	Mean	Median	St dev	95th	Observations
Investment rate	.14	.102	.131	.397	123,954
Investment spike	.198		.4		123,954
Investment rate | spike	.315	.278	.181	.694	23,729
Time since last spike	4.70	2	6.36	18	81,257
R&D-to-sales ratio	.029	0	.057	.170	123,954
Positive R&D expenditure	.446		.497		123,954
R&D-to-sales ratio | positive R&D expenditure	.064	.0345	.070	.223	55,388
Leverage	.258	.229	.215	.659	123,954

Notes: This table shows descriptive statistics for variables used in the empirical analysis of Section 5.1. Investment rate, R&D-to-sales ratio, and leverage are defined in Appendix A.1. *Investment spike* denotes a dummy variable that takes the value of one in periods in which a firm's investment rate is above 20%. *Time since last spike* denotes the number of years since the firm experienced the previous investment spike. *Positive R&D expenditure* denotes a dummy variable that takes the value of one in a period in which a firm's research and development expense (`xrd`) is positive. *Investment rate* | *spike* and *R&D-to-sales ratio* | *positive R&D expenditure* report, respectively, moments for investment rates conditional on periods of investment spikes and of R&D-to-sales ratios conditional on positive R&D expenditure. For sample selection, see Appendix A.1.

3. Firm-year observations that satisfy one of the following conditions, aimed at excluding extreme observations:

   i. Negative assets, sales, capital expenditure, or R&D.

   ii. Low capital values (gross plant, property, and equipment below $5M in 1990 dollars).

   iii. Acquisitions larger than 20% of assets.

   iv. Investment rates higher than 1.

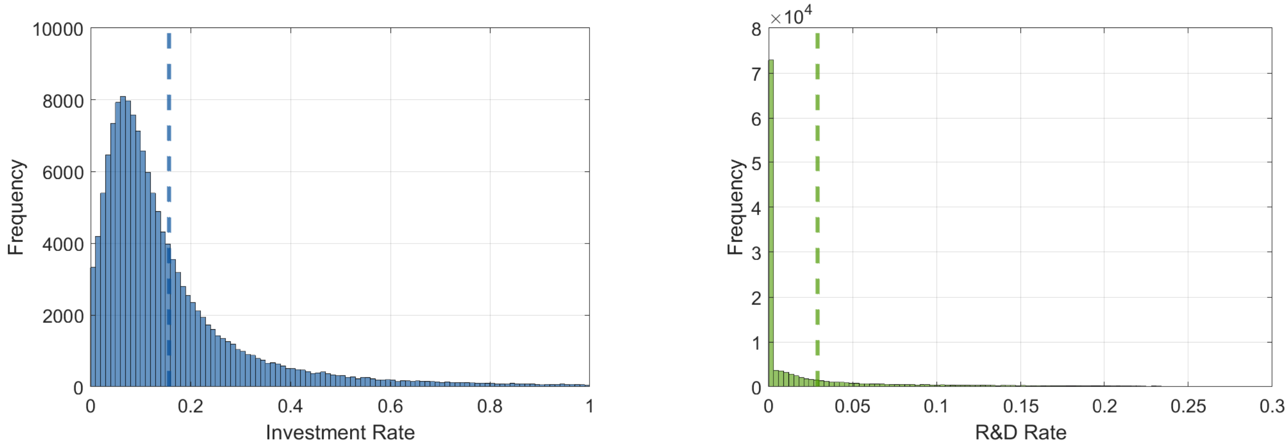
   v. Innovation-to-sales ratios higher than 0.3.

   vi. Gross leverage (defined as the ratio of total debt to total assets) higher than 10 or negative.

Descriptive Statistics Table A.1 contains descriptive statistics of our final analysis sample. Figure A.1 plots the distribution of investment rates and R&D-to-sales ratios in our sample.

Figure A.1: Distribution of Investment Rates and R&D

(a) Investment rates (b) R&D-to-sales



Notes: This figure shows the histogram of investment rates and the R&D-to-sales ratio. Vertical dashed lines represent each variable mean. For variables definitions and sample selection, see Appendix A.1.

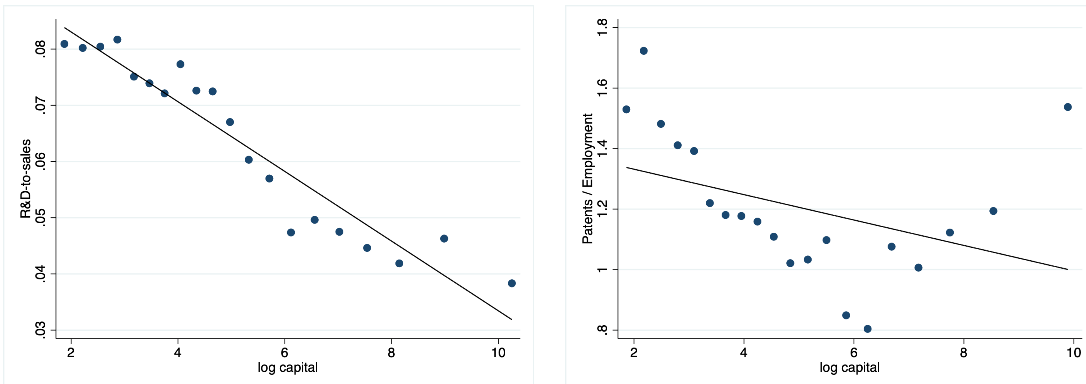
A.2 Additional Results on Role of Within-Firm Variation

This appendix shows additional results on the role of within-firm variation in our comparison to the existing literature from the main text. First, Figure A.2 shows that the difference between using all variation and only using within-firm variation holds using all firms in our sample, not just the "continuous innovative" firms from Figure 2. Second, Table A.2 compares all vs. within-firm variation in a regression context by estimating the regression (1) from the main text for various measures of size s_{jt}. For nearly all specifications, including a firm fixed effect flips the negative relationship between size and innovation, as illustrated by the bin-scatters.[25] The only exception to this pattern is the relationship between patents per employee and employment itself. Table A.2 also shows that for some measures of innovation, such as patent activity and the market value per patent, there is a positive relationship between innovation and size both with or without firm fixed effects.

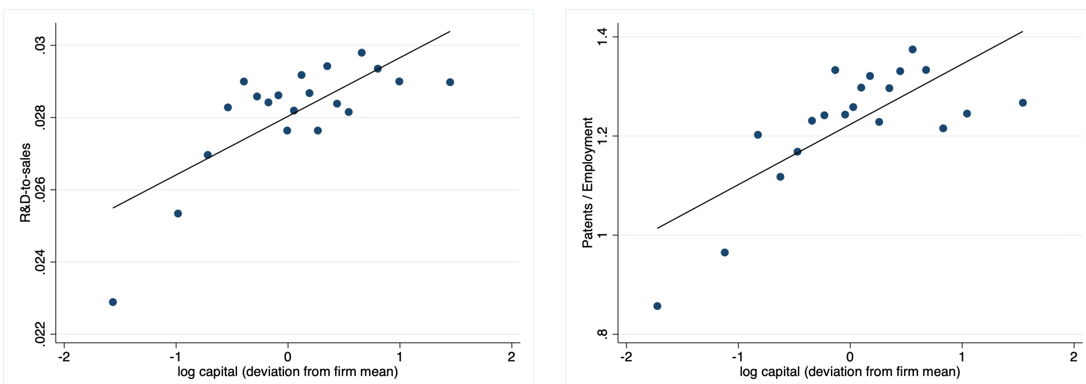
[25]These regression results are a useful way to summarize the various specifications, but the results are similarly visually verified in a collection of bin-scatters similar to Figure 2 in the main text. Results available upon request.

Figure A.2: Role of Within-Firm Variation (All firms)

All variation



Within-firm variation



Notes: Binned scatter plots of R&D to sales and patents per employee by firm size (measured by the log of real capital). All variables are demeaned at the firm level. In order to make the units of the outcome variable more interpretable, we add back in the unconditional mean across all firms of investment rates, share of firms with positive R&D, and share of firms with positive patenting. For variable definitions and sample selection, see Appendix A.

Table A.2
Role of Within-Firm Variation For Various Measures of Size

	R&D activity		R&D-to-sales		Patent activity		Patents per employee		log value per patent	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	No FE	Firm FE	No FE	Firm FE	No FE	Firm FE	No FE	Firm FE	No FE	Firm FE
(a) Unrestricted sample										
log capital	-0.006***	0.026***	-0.006***	0.001**	0.089***	0.072***	-0.091***	0.112*	0.566***	0.492***
	(0.001)	(0.004)	(0.0001)	(0.001)	(0.001)	(0.005)	(0.011)	(0.065)	(0.004)	(0.029)
log employment	0.026***	0.035***	-0.005***	0.002***	0.102***	0.087***	-0.269***	-0.309***	0.573***	0.474***
	(0.001)	(0.005)	(0.0002)	(0.001)	(0.001)	(0.005)	(0.011)	(0.070)	(0.005)	(0.033)
log sales	0.025***	0.026***	-0.004***	0.002***	0.105***	0.071***	-0.072***	0.007	0.619***	0.562***
	(0.001)	(0.004)	(0.0002)	(0.001)	(0.001)	(0.005)	(0.011)	(0.065)	(0.004)	(0.028)
log net worth	0.022***	0.017***	-0.001***	0.002***	0.101***	0.062***	0.193***	0.335***	0.615***	0.390***
	(0.001)	(0.003)	(0.0002)	(0.0004)	(0.001)	(0.004)	(0.012)	(0.056)	(0.004)	(0.025)
(b) Firms with more than 20 years of data										
log capital	0.025***	0.026***	-0.0004**	0.002**	0.103***	0.067***	0.134***	0.270***	0.393***	0.433***
	(0.002)	(0.008)	(0.0002)	(0.001)	(0.002)	(0.008)	(0.016)	(0.102)	(0.010)	(0.081)
log employment	0.045***	0.033***	0.0001**	0.002***	0.128***	0.113***	-0.024	-0.381***	0.401***	0.323***
	(0.003)	(0.011)	(0.0002)	(0.001)	(0.002)	(0.012)	(0.017)	(0.129)	(0.013)	(0.087)
log sales	0.042***	0.025***	0.0004**	0.002***	0.122***	0.075***	0.149***	0.246**	0.438***	0.508***
	(0.002)	(0.008)	(0.0002)	(0.001)	(0.002)	(0.009)	(0.016)	(0.108)	(0.011)	(0.083)
log net worth	0.027***	0.016**	0.001***	0.001**	0.103***	0.051***	0.251***	0.327***	0.393***	0.332***
	(0.002)	(0.007)	(0.0002)	(0.001)	(0.002)	(0.007)	(0.017)	(0.087)	(0.010)	(0.068)
(c) Continuously innovative firms with more than 20 years of data										
log capital	-0.018***	0.029**	-0.003***	0.003**	0.116***	0.031**	0.030	0.405*	0.393***	0.433***
	(0.002)	(0.014)	(0.0003)	(0.001)	(0.003)	(0.014)	(0.025)	(0.219)	(0.010)	(0.081)
log employment	-0.017***	0.037**	-0.003***	0.004***	0.153***	0.142***	-0.298***	-1.024***	0.401***	0.323***
	(0.003)	(0.018)	(0.0003)	(0.001)	(0.003)	(0.020)	(0.028)	(0.286)	(0.013)	(0.087)
log sales	-0.017***	0.025*	-0.003***	0.004**	0.130***	0.048***	0.001	0.387*	0.438***	0.508***
	(0.003)	(0.014)	(0.0003)	(0.002)	(0.003)	(0.015)	(0.026)	(0.230)	(0.011)	(0.083)
log net worth	-0.012***	0.013	-0.0002*	0.003**	0.116***	0.023**	0.218***	0.516***	0.393***	0.332***
	(0.003)	(0.010)	(0.0003)	(0.001)	(0.003)	(0.011)	(0.026)	(0.168)	(0.010)	(0.068)

Notes: Results from estimating the regression $o_{jt} = \alpha_j + \gamma \log s_{jt} + \varepsilon_{jt}$, where o_{jt} is the outcome of interest (indicator for positive R&D, R&D-to-sales, indicator for positive patenting, patent per employee, or log market value per patent), s_{jt} is the measure of size (capital, sales, employment, or net worth), and α_j is a firm fixed effect. We standardize the size measures $\log s_{jt}$ and the log value per patent over the entire sample. Standard errors are clustered at the firm level. Panel (a) reports results using the sample of all firms and periods; panel (b) the sample of firms with at least 20 years of observations; and panel (c) the sample of firms with at least 20 years of observations and that are "continuously innovative" (i.e., firms that have conducted positive R&D or patenting activity over the last five years.)

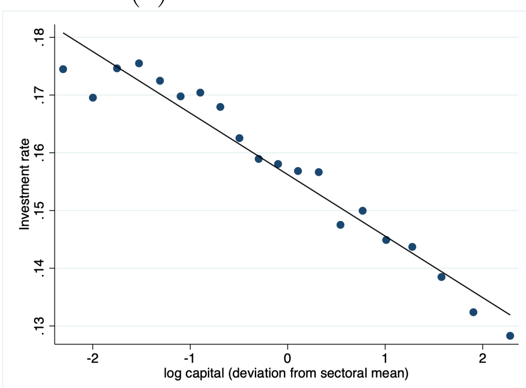
A.3 Robustness of Main Results

This section contains additional robustness analysis referenced in the main text.

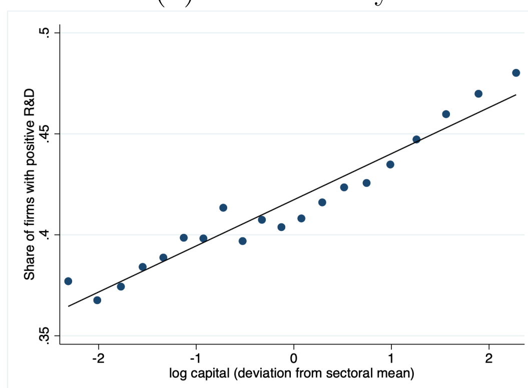
Within-Sector Variation Figure A.3 plots the version of our main bin-scatter Figure 1 except using variation within 4-digit sector rather than only within-firm. If anything, the relationship between investment/innovation and firm size becomes even stronger in this specification.

FIGURE A.3: The Pecking Order of Firm Growth Within Sector

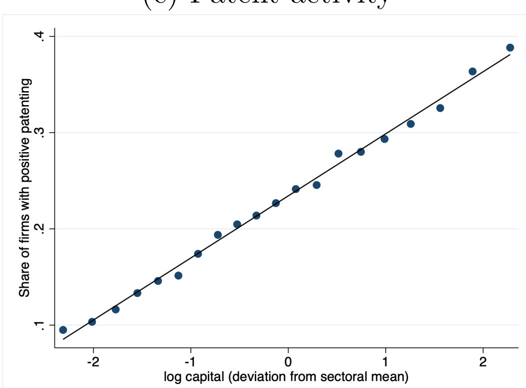
(a) Investment rates



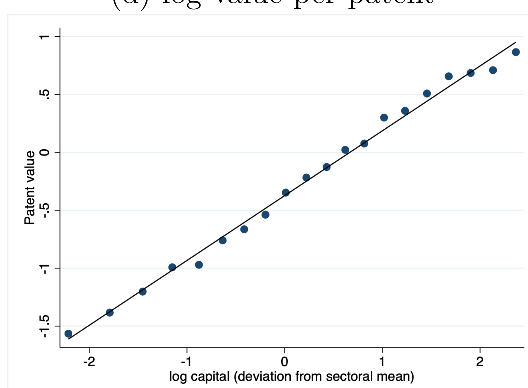
(b) R&D activity



(c) Patent activity



(d) log value per patent

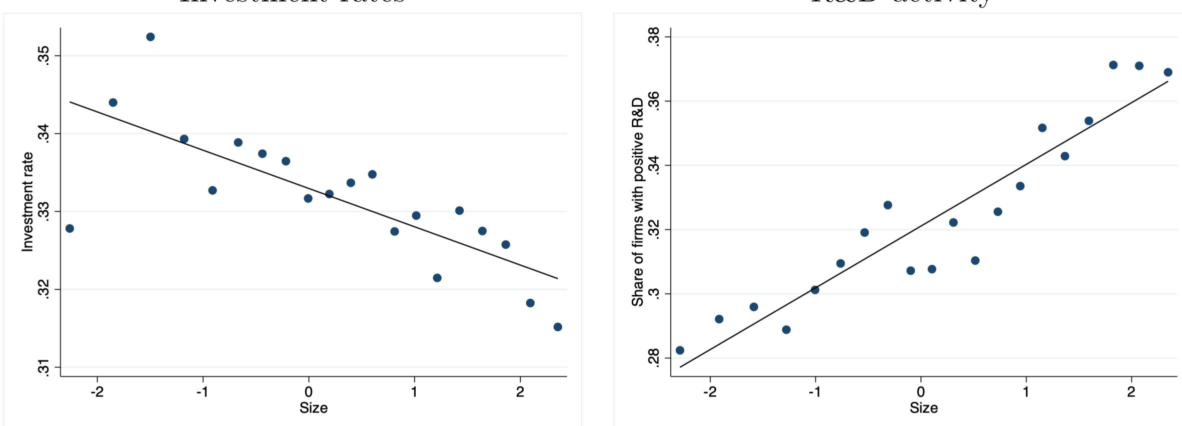


Notes: Binned scatter plots of investment rates, the share of firms with positive R&D, he share of firms with positive patenting, and the market-value weighted change in firms patenting by firm size (measured by the log of real capital). All variables are demeaned at the 4-digit NAICS sector level. In order to make the units of the outcome variable more interpretable, we add back in the unconditional mean across all firms of investment rates, share of firms with positive R&D, and share of firms with positive patenting. For variable definitions and sample selection, see Appendix A.

Orbis Data Figure A.4 shows our empirical results using Orbis, which includes data on privately held firms. In this analysis, we use within 4-digit variation rather than within-firm variation because the number of observations per firm is more limited for the sample of privately held firms (e.g., only 5% of privately held firms have at least 20 years of observations, which is the threshold we use to include firms in our baseline analysis). Similar to our results using Compustat data, we find that investment activity decreases as firms grow larger while innovation activity increases as firms grow larger.

FIGURE A.4: The Pecking Order of Firm Growth in Orbis Data

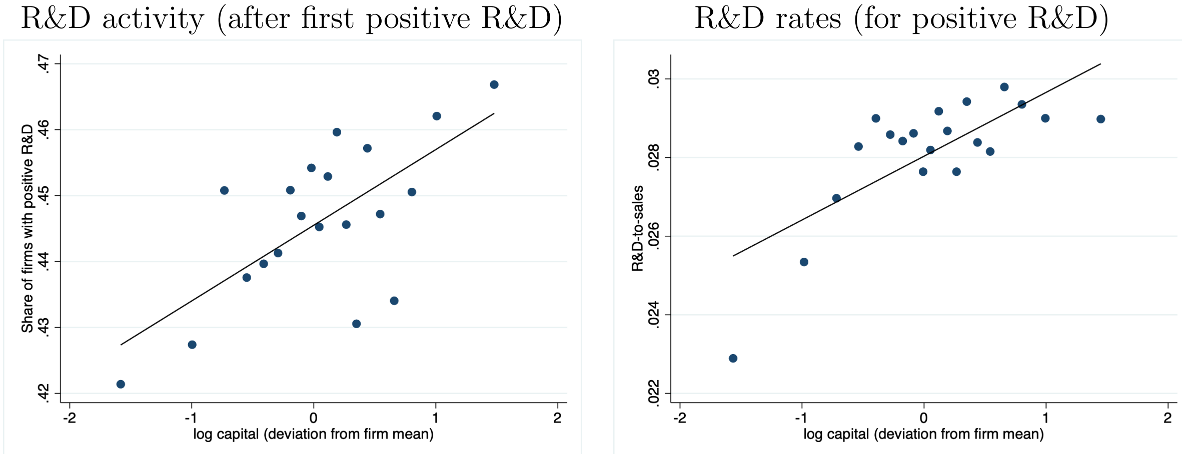
Investment rates | R&D activity



Notes: These figures report binned scatter plots of investment rates and the share of firms with positive R&D by firms' size (measured by the log of real capital) in the Orbis-US dataset. All variables are demeaned at 4-digit-NAICS sector level. To construct the plots for investment rates and the share of firms with positive R&D rates, we add the unconditional mean of each variable to sector-level demeaned variables.

Other Measures of Innovation Figure A.5 shows how our baseline within-firm bin-scatter plot looks similar for two other measures of innovation inputs. First, the left panel plots the share of observations with positive R&D expenditures, conditional on already having reported positive R&D in the past. The hope is that this sample restricts to firms who have already set up the accounting infrastructure to report formal R&D and therefore has less measurement error. Second, the right panel plots the intensive margin, i.e., the R&D-sales-ratio conditional on reporting positive R&D in that year.

Figure A.5: The Pecking Order of Firm Growth for Other Measures of Innovation



Notes: Binned scatter plots of the share of firms with positive R&D conditional on already having an observation with positive R&D (left panel) and the level of R&D-to-sales conditional on having positive R&D (right panel). All variables are demeaned at the firm level. In order to make the units of the outcome variable more interpretable, we add back in the unconditional mean of the outcome variables across all firms. For variable definitions and sample selection, see Appendix A.

Additional Regression Results Table A.3 contains two robustness checks on our regression specification (1). First, panel (b) shows that our results hold using all firms in our Compustat sample, not just those with at least twenty years of data (as in the main text). Second, panel (c) shows that our results are robust to controlling for time fixed effects (which absorb aggregate changes in the composition of investment and innovation).

A.4 Innovation and Investment Spikes

This appendix contains additional results about the relationship between R&D expenditures and investment spikes referenced in Section 5 of the main text. Figure A.4 shows that the results continue to hold with additional lags. Figure A.6 complements the regression results with an event-study analysis around an investment spike.

B Model Characterization and Proofs

This appendix characterizes the model's solution and proves Proposition 1 and Proposition 2 from the main text.

Table A.3
Robustness of the Pecking Order of Firm Growth

	(1) Investment rate	(2) R&D activity	(3) R&D-to-sales	(4) Patent activity	(5) log value per patent
(a) Baseline					
$\hat{\gamma}$	-0.088***	0.026***	0.002**	0.067***	0.433***
	(0.003)	(0.008)	(0.001)	(0.008)	(0.081)
N	50464	46439	46439	54419	9668
R^2	0.273	0.856	0.873	0.632	0.760
(b) Unrestricted sample					
$\hat{\gamma}$	-0.084***	0.026***	0.001**	0.072***	0.492***
	(0.002)	(0.004)	(0.001)	(0.005)	(0.029)
N	156030	165706	132444	165706	38918
R^2	0.322	0.873	0.884	0.604	0.832
(c) Time fixed effects					
$\hat{\gamma}$	-0.041***	0.040***	0.003***	0.146***	0.172*
	(0.005)	(0.013)	(0.001)	(0.013)	(0.098)
N	50464	46439	46439	54419	9668
R^2	0.319	0.856	0.873	0.636	0.792
Mean	0.13	0.44	0.02	0.34	0

Notes: Panels (a) and (b) report results from estimating the regression $o_{jt} = \alpha_j + \gamma \log k_{jt} + \varepsilon_{jt}$, where o_{jt} is the outcome of interest (investment rate, indicator for positive R&D, R&D-to-sales ratio, indicator for positive patenting, or log market value per patent computed following Kogan et al. (2017)), k_{jt} is the capital stock, and α_j is a firm fixed effect. We standardize the log of capital and the log of value per patent over the entire sample. Standard errors are clustered at the firm level. Panel (a) reports results for our baseline specification, for the sample of firms with at least 20 years of observations, and panel (b) those using the sample of all firms and periods. Panel (c) reports the results from estimating the regression $o_{jt} = \alpha_j + \alpha_t + \gamma \log k_{jt} + \varepsilon_{jt}$ α_t is a time fixed effect. Standard errors are two-way clustered by firms and year.

B.1 Positive Analysis: Firms' Decision Rules and the BGP

This subsection characterizes the individual firm's decisions and defines a balanced growth path. We proceed in three steps. First, we detrend the problem in order to work with a stationary Bellman equation for which the usual numerical tools apply. Second, we characterize the solution of the detrended problem and show that it results in Proposition 1 in the main text. Finally, we use these results to show that all decisions and macroeconomic aggregates scale with the growth rate g in a balanced growth path.

Table A.4
Investment spikes and Innovation: Robustness

	(1)	(2)	(3)	(4)	(5)
$\frac{i_{jt-1}}{\tilde{y}_{jt-1}}$	1.12	0.67	1.03	1.12	1.06
	(0.15)	(0.13)	(0.14)	(0.16)	(0.16)
$\frac{\text{cf}_{jt}}{k_{jt}}$	0.11	0.07	0.11	0.12	0.12
	(0.02)	(0.01)	(0.02)	(0.02)	(0.04)
years since spike$_{t-1}$	0.003	0.002	0.003	0.003	0.003
	(0.0008)	(0.0005)	(0.0008)	(0.0008)	(0.0008)
$\frac{k_{jt}}{n_{jt-1}}$	-0.016	-0.01	-.0139	-0.017	-0.014
	(0.003)	(0.002)	(0.003)	(0.004)	(0.003)
Measure of spikes	Absolute	Sectoral	Absolute	Absolute	Absolute
Lags	4	4	3	5	4
Additional controls	No	No	No	No	Size, sales growth, current assets
Observations	53,577	53,577	58,066	49,116	52,292
Adj. R^2	0.282	0.186	0.285	0.279	0.30

Notes: Results from estimating alternative versions of $\mathbb{1}\{\frac{x_{jt}}{k_{jt}} \geq \chi_s\} = \alpha_j + \alpha_{st} + \sum_{h=1}^{H} \beta_h \left(\frac{\widehat{i}_{jt-h}}{\tilde{y}_{jt-h}}\right) + \Gamma' X_{jt} + \varepsilon_{jt}$, where $\frac{x_{jt}}{k_{jt}}$ denotes the investment rate of firm j in period t; χ_s is a threshold defining investment spikes; $\frac{\widehat{i}_{jt}}{\tilde{y}_{jt}}$ the R&D-to-sales ratio; α_j and α_{st} firm and time by sector fixed effects; $X_{j,t}$ is a vector of firm-level controls; and ε_{jt} is a random error term. Column (1) reports estimates for the baseline specification of Table 2, with $\chi_s = 0.2$, $H = 1$, and the vector X_{jt} including cash flows ($\frac{\text{cf}_{jt}}{k_{jt}}$) and the lumpy-investment controls (years since the last investment spike, years since spike$_{t-1}$, and the standardized capital-output ratio, $\frac{k_{jt}}{n_{jt-1}}$). Column (2) uses a "sectoral" threshold for investment spikes, where $\chi_{ts} = 0.2$ is the mean plus one standard deviation of the distribution of investment rates of sector s (at 2-digit NAICS level). Columns (3) and (4) report results for alternative lags of the R&D-to-sales ratio: $H = 3$ and $H = 5$. Column (5) includes additional control variables: size (measured with the log of real plant, property, and equipment), sales growth, and the share of current assets. For variable definitions and descriptive statistics, see Appendix A.

B.1.1 Detrending

We will scale the problem by average productivity $Z_t = A_t \int z_{jt} dj = \left(\int z_{jt} dj\right)^{1+a}$. To that end, let $\widetilde{n} = \frac{n}{Z_t}$, $\widetilde{k} = \frac{k}{Z_t}$ denote variables relative to Z_t. The only except is that we will define $\widetilde{z} = \frac{z}{\int z_{jt} dj}$. Divide the Bellman equation (5) by Z_t to get

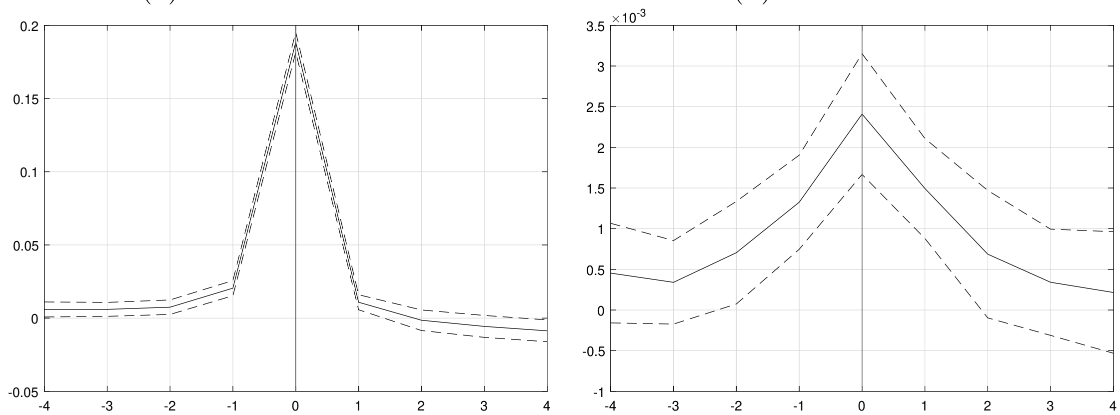
$$\frac{v_t^{\text{cont}}(z,n)}{Z_t} = \max_{k',i,b'} \frac{n}{Z_t} - \frac{k'}{Z_t} - \frac{A_t z i}{Z_t} + \frac{b'}{Z_t(1+r_t)} + \frac{1}{1+r_t}\mathbb{E}_t\left[\pi_d \frac{n'}{Z_t} + (1-\pi_d)\frac{v_{t+1}^{\text{cont}}(z',n')}{Z_t}\right], \quad (16)$$

where we have expanded $\mathbb{E}_t[v_{t+1}(z',n')] = \pi_d \mathbb{E}_t[n'] + (1-\pi_d)\mathbb{E}_t[v_{t+1}(z',n')]$.

Our goal is to write (16) in terms of the detrended variables and the growth rate $g_t = \frac{Z_{t+1}}{Z_t}$

Figure A.6: Event Study Analysis of Investment Spikes

(a) Investment rates (b) R&D-to-sales



Notes: This figure shows the dynamics of investment rates and R&D-to-sales around investment spike episodes. The figure reports the coefficients β_h from estimating $y_{jt} = \alpha_j + \alpha_{st} + \sum_{h=-4}^{4} \beta_h \frac{x_{jt}}{k_{jt-g}} \geq 0.2 + \Gamma' X_{j,t-1} + \varepsilon_{jt}$, where y_{jt} denotes the investment rate ($\frac{x_{jt}}{k_{jt}}$) or R&D-to-sales ratio ($\frac{\widehat{i}_t}{\widetilde{y}_t}$); α_j and α_{ts} firm and time by sector fixed effects; and ε_{jt} is a random error term. For variable definitions and descriptive statistics, see Appendix.

only. To that end, note that $\frac{k'}{Z_t} = \frac{k'}{Z_{t+1}} \frac{Z_{t+1}}{Z_t} = (1+g_t)\widetilde{k}'$. Similarly, $\frac{b'}{Z_t} = (1+g_t)\widetilde{b}'$. Now multiply and divide the continuation value by $\frac{Z_{t+1}}{Z_{t+1}}$ to get

$$\frac{v_t^{\text{cont}}(z,n)}{Z_t} = \max_{k',i,b'} \widetilde{n} - (1+g_t)\widetilde{k}' - \widetilde{z}i + \frac{(1+g_t)\widetilde{b}'}{(1+r_t)} + \frac{1+g_t}{1+r_t}\mathbb{E}_t\left[\pi_d \widetilde{n}' + (1-\pi_d)\frac{v_{t+1}^{\text{cont}}(z',n')}{Z_{t+1}}\right].$$

Define $\widetilde{v}_t(\widetilde{z},\widetilde{n}) = \frac{v_t^{\text{cont}}(z,n)}{Z_t}$ to arrive at our final detrended Bellman equation:

$$\widetilde{v}_t(\widetilde{z},\widetilde{n}) = \max_{\widetilde{k}',i,\widetilde{b}'} \widetilde{n} - (1+g_t)\widetilde{k}' - \widetilde{z}i + \frac{(1+g_t)\widetilde{b}'}{(1+r_t)} + \frac{1+g_t}{1+r_t}\mathbb{E}_t\left[\pi_d \widetilde{n}' + (1-\pi_d)\widetilde{v}_{t+1}(\widetilde{z}',\widetilde{n}')\right]. \quad (17)$$

Finally, we detrend the constraints and consistency conditions of this problem. Clearly, we have $\widetilde{d} \geq 0$, $\widetilde{b}' \leq \theta \widetilde{k}'$, and $\widetilde{n}' = (\widetilde{z}')^{1-\alpha}(\widetilde{k}')^{\alpha} + (1-\delta)\widetilde{k}' - \widetilde{b}'$. In terms of the law of motion for z, in the event of a successful innovation, we have

$$\log \frac{z}{\int z_{jt+1} dj} = \log \frac{z}{\int z_{jt+1} dj} + \Delta + \varepsilon_{jt+1} = \log \frac{z}{\int z_{jt} dj} \frac{\int z_{jt} dj}{\int z_{jt+1} dj} + \Delta + \varepsilon_{jt+1}$$

which implies

$$\log \widetilde{z}' = \log \frac{\widetilde{z}}{1+\widetilde{g}_t} + \Delta + \varepsilon_{jt+1}$$

where $\widetilde{g}_t = \frac{\int z_{jt+1} dj}{\int z_{jt} dj}$ is the growth rate of firm-specific productivity. Since $Z_t = \left(\int z_{jt} dj\right)^{1+a}$, we have $1 + g_t = (1 + \widetilde{g}_t)^{1+a}$.

B.1.2 Proof of Proposition 1

Our characterization in Proposition 1 is similar to Khan and Thomas (2013), extended to include the innovation decision. We proceed in three steps. First, we set up the Lagrangian and take the associated first-order conditions. Second, we use those first-order conditions to derive the partition of the state space from the first part of Proposition 1. Finally, we un-detrend those first-order conditions to get the system of equations in the second part of Proposition 1.

Lagrangian The Lagrangian of the detrended Bellman equation (17) is

$$\mathcal{L} = (1 + \lambda_t(\widetilde{z}, \widetilde{n})) \left(\widetilde{n} - (1 + g_t)\widetilde{k}' - \widetilde{z}i + \frac{(1 + g_t)\widetilde{b}'}{(1 + r_t)} \right) + (1 + g_t)\mu_t(\widetilde{z}, \widetilde{n}) \left(\theta\widetilde{k}' - \widetilde{b}' \right) \tag{18}$$
$$+ \chi_t(\widetilde{z}, \widetilde{n})i + \frac{1 + g_t}{1 + r_t}\mathbb{E}_t \left[\pi_d \widetilde{n}' + (1 - \pi_d)\widetilde{v}_{t+1}(\widetilde{z}', \widetilde{n}') \right],$$

where $\lambda_t(\widetilde{z}, \widetilde{n})$ is the multiplier on the no-equity issuance constraint $\widetilde{d} \geq 0$, $\mu_t(\widetilde{z}, \widetilde{n})$ is the multiplier on the collateral constraint $\widetilde{b}' \leq \theta\widetilde{k}'$, and $\chi_t(\widetilde{z}, \widetilde{n})$ is the multiplier on the nonnegativity constraint on innovation $i \geq 0$.

The first-order condition for borrowing $\widetilde{b}'$ is

$$(1 + \lambda_t(\widetilde{z}, \widetilde{n}))\frac{1 + g_t}{1 + r_t} = (1 + g_t)\mu_t(\widetilde{z}, \widetilde{n}) - \frac{1 + g_t}{1 + r_t}\mathbb{E}_t \left[\pi_d \frac{\partial \widetilde{n}'}{\partial \widetilde{b}'} + (1 - \pi_d)\frac{\partial \widetilde{v}_{t+1}(\widetilde{z}', \widetilde{n}')}{\partial \widetilde{n}'}\frac{\partial \widetilde{n}'}{\partial \widetilde{b}'} \right].$$

From the envelope condition, we have $\frac{\partial \widetilde{v}_t(\widetilde{z}, \widetilde{n})}{\partial \widetilde{n}'} = 1 + \lambda_t(\widetilde{z}, \widetilde{n})$. Use that together with $\frac{\partial \widetilde{n}'}{\partial \widetilde{b}'} = -1$ to get

$$(1 + \lambda_t(\widetilde{z}, \widetilde{n}))\frac{1 + g_t}{1 + r_t} = (1 + g_t)\mu_t(\widetilde{z}, \widetilde{n}) + \frac{1 + g_t}{1 + r_t}\mathbb{E}_t \left[\pi_d + (1 - \pi_d)(1 + \lambda_{t+1}(\widetilde{z}', \widetilde{n}')) \right].$$

Note that $\pi_d + (1 - \pi_d)(1 + \lambda_{t+1}(\widetilde{z}, \widetilde{n})) = 1 + (1 - \pi_d)\lambda_{t+1}(\widetilde{z}, \widetilde{n})$. Use that fact, multiply by

$\frac{1+r_t}{1+g_t}$, and subtract 1 from both sides to finally arrive at

$$\lambda_t(\widetilde{z},\widetilde{n}) = (1+r_t)\mu_t(\widetilde{z},\widetilde{n}) + (1-\pi_d)\mathbb{E}_t\lambda_{t+1}(\widetilde{z}',\widetilde{n}'). \tag{19}$$

Hence, the financial wedge $\lambda_t(\widetilde{z},\widetilde{n})$ is the expected value of current and all future Lagrange multipliers on the collateral constraint $\mu_t(\widetilde{z},\widetilde{n})$, discounted by the exit probability.

The first-order condition for capital accumulation $\widetilde{k}'$ is

$$(1+g_t)(1+\lambda_t(\widetilde{z},\widetilde{n})) = \theta(1+g_t)\mu_t(\widetilde{z},\widetilde{n}) + \frac{1+g_t}{1+r_t}\mathbb{E}_t\left[\pi_d\frac{\partial\widetilde{n}'}{\partial\widetilde{k}'} + (1-\pi_d)\frac{\partial\widetilde{v}_{t+1}(\widetilde{z}',\widetilde{n}')}{\partial\widetilde{n}'}\frac{\partial\widetilde{n}'}{\partial\widetilde{k}'}\right].$$

Note that $\frac{\partial\widetilde{n}'}{\partial\widetilde{k}'} = MPK(\widetilde{z}',\widetilde{k}') + (1-\delta)$, where $MPK(\widetilde{z}',\widetilde{k}') = \alpha\left(\frac{\widetilde{z}'}{\widetilde{k}'}\right)^{1-\alpha}$ is the marginal product of capital. Using very similar steps to above, the terms in the continuation value can be collected to yield

$$1+\lambda_t(\widetilde{z},\widetilde{n}) = \theta\mu_t(\widetilde{z},\widetilde{n}) + \frac{1}{1+r_t}\mathbb{E}_t\left[\Big(MPK(\widetilde{z}',\widetilde{k}') + (1-\delta)\Big)\left(\pi_d + (1-\pi_d)\lambda_{t+1}(\widetilde{z}',\widetilde{n}')\right)\right]. \tag{20}$$

The first-order condition for innovation i is

$$(1+\lambda_t(\widetilde{z},\widetilde{n}))\widetilde{z} = \chi_t(\widetilde{z},\widetilde{n}) + \frac{1+g_t}{1+r_t}\frac{\partial}{\partial i}\mathbb{E}_t\left[\pi_d\widetilde{n}' + (1-\pi_d)\widetilde{v}_{t+1}(\widetilde{z}',\widetilde{n}')\right].$$

Consider the term in the continuation value in the case where the firm exits in the next period. We can write this expectation as $\mathbb{E}_t[\widetilde{n}'] = \eta(i)\mathbb{E}^{\varepsilon}[\widetilde{n}'|\text{success}] + (1-\eta(i))\mathbb{E}^{\varepsilon}[\widetilde{n}'|\text{failure}]$ where $\mathbb{E}^{\varepsilon}$ denotes the expectation over the idiosyncratic shocks ε. Hence, we have $\frac{\partial\mathbb{E}_t[\widetilde{n}']}{\partial i} = \eta'(i)\left(E^{\varepsilon}[\widetilde{n}'|\text{success}] - E^{\varepsilon}[\widetilde{n}'|\text{failure}]\right)$. By a similar argument,

$$\frac{\partial\mathbb{E}_t[\widetilde{v}_{t+1}(\widetilde{z}',\widetilde{n}')]}{\partial i} = \eta'(i)\left(E^{\varepsilon}[\widetilde{v}_{t+1}(\widetilde{z}',\widetilde{n}')|\text{success}] - E^{\varepsilon}[\widetilde{v}_{t+1}(\widetilde{z}',\widetilde{n}')|\text{failure}]\right).$$

Putting these all together yields

$$(1+\lambda_t(\widetilde{z},\widetilde{n}))\widetilde{z} \geq \frac{1+g_t}{1+r_t}\eta'(i)\begin{bmatrix}\pi_d\left(E^{\varepsilon}[\widetilde{n}'|\text{success}] - E^{\varepsilon}[\widetilde{n}'|\text{failure}]\right) + \\ (1-\pi_d)\left(E^{\varepsilon}[\widetilde{v}_{t+1}(\widetilde{z}',\widetilde{n}')|\text{success}] - E^{\varepsilon}[\widetilde{v}_{t+1}(\widetilde{z}',\widetilde{n}')|\text{failure}]\right)\end{bmatrix}, \tag{21}$$

with equality if $i > 0$.

To summarize, the firm's optimal decisions are characterized by the first-order conditions (19), (20), and (21) together with the complementarity slackness conditions:

$$\mu_t(\widetilde{z},\widetilde{n})(\theta\widetilde{k}' - \widetilde{b}') = 0 \text{ with } \mu_t(\widetilde{z},\widetilde{n}) \geq 0, \text{ and}$$
$$\lambda_t(\widetilde{z},\widetilde{n})\widetilde{d} = 0 \text{ with } \lambda_t(\widetilde{z},\widetilde{n}) \geq 0.$$

Partition of State Space We now use these first order conditions to derive the partition of the state space in the first part of Proposition 1.

Unconstrained Firms: We define a financially unconstrained firm as one for whom the financial wedge $\lambda_t(z, n) = 0$. From (19), these firms have zero probability of a binding collateral constraint in the future, so $\mu_{jt+s} = \lambda_{jt+s} = 0$ for all $s \geq 0$; that is, being unconstrained is an absorbing state. We will guess and verify that these firms decisions are independent of net worth and are characterized by a set of objects $\widetilde{b}_t'^*(\widetilde{z})$, $\widetilde{k}_t'^*(\widetilde{z})$, $i_t^*(\widetilde{z})$, and $\widetilde{v}_t^*(\widetilde{z})$. We now characterize these objects.

First, because $\lambda_t(\widetilde{z},\widetilde{n}) = \mu_t(\widetilde{z},\widetilde{n}) = 0$, they are indifferent over any combination of b' and d which leaves them financially unconstrained. Following Khan and Thomas (2013), we resolve this indeterminacy by assuming firms accumulate the most debt (or, if $b' < 0$, do the least amount of savings) which leaves them financially unconstrained. Khan and Thomas (2013) refer to this policy $b_t'^*(\widetilde{z})$ as the *minimum savings policy*. In order to derive a characterization of it, note that if the firm adopts $b_t'^*(\widetilde{z})$ in period t, then its dividends in the next period $t+1$, conditional on a particular realized state $\widetilde{z}'$, are

$$\widetilde{d}_{t+1}(\widetilde{z}') = (\widetilde{z}')^{1-\alpha}(\widetilde{k}_t'^*(\widetilde{z}))^\alpha + (1-\delta)\widetilde{k}_t'^*(\widetilde{z}) - \widetilde{b}_t'^*(\widetilde{z}) - \widetilde{z}' i_{t+1}^*(\widetilde{z}') - (1+g_{t+1})\widetilde{k}_{t+1}'^*(\widetilde{z}') + \frac{1+g_{t+1}}{1+r_{t+1}}\widetilde{b}_{t+1}'^*(\widetilde{z}')$$

In order to be financially unconstrained, it must be the case that $\widetilde{d}_{t+1}(\widetilde{z}') \geq 0$ for all $\widetilde{z}'$ which have a positive probability. The minimum savings policy $\widetilde{b}_t'^*(\widetilde{z})$ is the largest level of debt which satisfies this constraint with probability one:

$$\widetilde{b}_t'^*(\widetilde{z}) = \min_{\widetilde{z}'}(\widetilde{z}')^{1-\alpha}(\widetilde{k}_t'^*(\widetilde{z}))^\alpha + (1-\delta)\widetilde{k}_t'^*(\widetilde{z}) - \widetilde{z}' i_{t+1}^*(\widetilde{z}') - (1+g_{t+1})\widetilde{k}_{t+1}'^*(\widetilde{z}') + \frac{1+g_{t+1}}{1+r_{t+1}}\widetilde{b}_{t+1}'^*(\widetilde{z}') \quad (22)$$

Note that this policy implies dividends are zero at a minimizer of the RHS of (22) and strictly positive otherwise.

Next, we define $\widetilde{v}_t^*(\widetilde{z})$ to be the value of a firm starting right after they adopt the unconstrained policies:

$$\widetilde{v}_t^*(\widetilde{z}) = -(1+g_t)\widetilde{k}_t'^*(\widetilde{z}) - \widetilde{z}i_t^*(\widetilde{z}) + \frac{(1+g_t)\widetilde{b}_t'^*(\widetilde{z})}{1+r_t} + \frac{1}{1+r_t}\mathbb{E}_t\left[\pi_d\widetilde{n}' + (1-\pi_d)\widetilde{v}_{t+1}^*(\widetilde{z}')\right], \quad (23)$$

where $\widetilde{n}' = (\widetilde{z}')^{1-\alpha}(\widetilde{k}_t'^*(\widetilde{z}))^\alpha + (1-\delta)\widetilde{k}_t'^*(\widetilde{z}) - \widetilde{b}_t'^*(\widetilde{z})$ is independent of $\widetilde{n}$. Since the financial constraints never bind for unconstrained firms, their value function is linearly separable in net worth. Therefore, the total value of a firm who becomes unconstrained in period t is $\widetilde{v}_t(\widetilde{z},\widetilde{n}) = \widetilde{n} + \widetilde{v}_t^*(\widetilde{z})$.

Given this characterization of the value function, the first-order conditions for capital and innovation (20) and (21) become

$$1 = \frac{1}{1+r_t}\mathbb{E}_t[MPK(\widetilde{z}',\widetilde{k}') + (1-\delta)] \quad (24)$$

$$1 \geq \frac{\eta'(i)}{\widetilde{z}}\frac{1+g_t}{1+r_t}\mathbb{E}_t\begin{bmatrix}\pi_d\left(E^\varepsilon[\widetilde{n}'|\text{success}] - E^\varepsilon[\widetilde{n}'|\text{failure}]\right) + \\ (1-\pi_d)\left(E^\varepsilon[\widetilde{v}_{t+1}^*(\widetilde{z}')|\text{success}] - E^\varepsilon[\widetilde{v}_{t+1}^*(\widetilde{z}')|\text{failure}]\right)\end{bmatrix}. \quad (25)$$

Note that the innovation policy implicitly enters the first-order condition for capital (24) through the expectations operator. Nevertheless, one can verify from (20) and (21) that these policies are independent of current net worth $\widetilde{n}$ given that both $\widetilde{n}'$ and $\widetilde{v}_{t+1}^*(\widetilde{z}')$ are themselves independent of net worth.

Finally, note that if it is feasible to follow these policies, then it will also be optimal because they solve the firm's profit maximization problem with an expanded choice set. In turn, it is feasible to follow these policies if the firm can adopt them without violating the no-equity issuance constraint:

$$\widetilde{n} - (1+g_t)k_t'^*(\widetilde{z}) - \widetilde{z}i_t^*(\widetilde{z}) + \frac{1+g_t}{1+r_t}\widetilde{b}_t'^*(\widetilde{z}) \geq 0. \quad (26)$$

This condition is satisfied if and only if $\widetilde{n} \geq \overline{n}_t(\widetilde{z}) \equiv (1+g_t)\widetilde{k}_t'^*(\widetilde{z}) + \widetilde{z}i_t^*(\widetilde{z}) - \frac{(1+g_t)\widetilde{b}_t'^*(\widetilde{z})}{1+r_t}$.

Constrained Firms: We define financially constrained firms as those for whom $\lambda_t(z,n) >$

0, i.e. there is a positive probability of facing a binding collateral constraint. These firms' decision rules are characterized by the full system of first-order conditions (19), (20), and (21), and therefore depend on net worth. We divide these firms into two cases: (i) *currently constrained* firms currently face a binding collateral constraint, i.e., $\mu_t(\widetilde{z},\widetilde{n}) > 0$, and (ii) *potentially constrained* firms who do not currently face a binding collateral constraint, i.e., $\mu_t(\widetilde{z},\widetilde{n}) = 0$.

To derive the threshold $\underline{n}_t(\widetilde{z},\widetilde{n})$ from the proposition, let $i_t^{\text{p}}(\widetilde{z},\widetilde{n})$, $\widetilde{k}_t'^{\text{p}}(\widetilde{z},\widetilde{n})$, and $\widetilde{b}_t'^{\text{p}}(\widetilde{z},\widetilde{n})$ denote the policy rules of the currently constrained firms. If these choices are feasible, then they are also optimal because they solve a relaxed version of the full problem. The policies are feasible as long as

$$\widetilde{n} \geq \underline{n}_t(\widetilde{z},\widetilde{n}) \equiv \widetilde{z}i_t(\widetilde{z},\widetilde{n}) + (1+g_t)\widetilde{k}_t'^{\text{p}}(\widetilde{z},\widetilde{n}) - \frac{(1+g_t)b_t'^{\text{p}}(\widetilde{z},\widetilde{n})}{1+r_t}.$$

Un-Detrending the Conditions We now show that the detrended first-order conditions (19), (20), and (21) derived above imply the conditions (6), (7), and (8) from the main text.

We start with the first-order condition for capital. First note that, from the chain rule,

$$\frac{\partial v_t(z,n)}{\partial n} = Z_t \frac{\partial \widetilde{v}_t(\widetilde{z},\frac{n}{Z_t})}{\partial n} = \frac{Z_t}{Z_t}\frac{\partial \widetilde{v}_t(\widetilde{z},\widetilde{n})}{\partial \widetilde{n}} \implies 1+\lambda_t(z,n) = 1+\lambda_t(\widetilde{z},\widetilde{n}),$$

i.e. the financial wedge is the same in the detrended and un-detrended problems. Next, note that

$$MPK_{t+1}(z',k') = \alpha\left(\frac{A_{t+1}z'}{k'}\right)^{1-\alpha} = \alpha\left(\frac{Z_t}{Z_t}\frac{A_{t+1}\widetilde{z}'}{\widetilde{k}'}\right)^{1-\alpha} = MPK(z',k').$$

Hence, the detrended first-order condition (20) directly implies the undetrended first-order condition (6) (where $\mu_t(z,n) = \mu_t(\widetilde{z},\widetilde{n})$ as well).

Next, consider the detrended first-order condition for innovation (21). Plugging in the fact that $1+g_t = \frac{Z_{t+1}}{Z_t}$ and rearranging gives

$$(1+\lambda_t(z,n))Z_t\widetilde{z} \geq \frac{\eta'(i_t(z,n))}{1+r_t} Z_{t+1}\mathbb{E}_t\left[\begin{array}{l}\pi_d\left(E^\varepsilon[\widetilde{n}'|\text{success}] - E^\varepsilon[\widetilde{n}'|\text{failure}]\right) + \\ (1-\pi_d)\left(E^\varepsilon[\widetilde{v}_{t+1}(\widetilde{z}',\widetilde{n}')|\text{success}] - E^\varepsilon[\widetilde{v}_{t+1}(\widetilde{z}',\widetilde{n}')|\text{failure}]\right)\end{array}\right]$$

By definition of the detrended variables, this equation is the same as the un-detrended condition (7) from the main text.

The nonnegativity constraint for dividends (8) follows directly from our detrending of the problem.

B.1.3 Balanced Growth Path

In this subsection, we characterize a balanced growth path of the model. In order to do so, we must first explicitly write out the law of motion for the distribution of firms. We find it easier to work with the distribution over de-trended state variables, $\Phi_t(\widetilde{z}, \widetilde{n})$. Heuristically, its evolution is given by

$$\widetilde{\Phi}_{t+1}(\widetilde{z}', \widetilde{n}') = (1-\pi_d) \int\int\int \begin{pmatrix} \eta(i_t(\widetilde{z},\widetilde{n})) \left[\mathbb{1}\{\widetilde{z}' = \frac{\widetilde{z}e^{\Delta}e^{\varepsilon}}{1+\widetilde{g}_t}\} \times \mathbb{1}\{n'(\frac{\widetilde{z}e^{\Delta}e^{\varepsilon}}{1+\widetilde{g}_t}, k_t'(\widetilde{z},\widetilde{n}), b_t'(\widetilde{z},\widetilde{n}))\}\right] \\ +(1-\eta(i_t(\widetilde{z},\widetilde{n}))) \left[\mathbb{1}\{\widetilde{z}' = \frac{\widetilde{z}e^{\varepsilon}}{1+\widetilde{g}_t}\} \times \mathbb{1}\{n'(\frac{\widetilde{z}e^{\varepsilon}}{1+\widetilde{g}_t}, k_t'(\widetilde{z},\widetilde{n}), b_t'(\widetilde{z},\widetilde{n}))\}\right] \end{pmatrix}$$
$$\times\, p(\varepsilon) d\varepsilon \widetilde{\Phi}_t(\widetilde{z},\widetilde{n}) d\widetilde{z} d\widetilde{n} + \pi_d \widetilde{\Phi}^0(\widetilde{z},\widetilde{n}), \tag{27}$$

where $\widetilde{n}' = (\widetilde{z}')^{1-\alpha}(\widetilde{k}_t'(\widetilde{z},\widetilde{n}))^{\alpha} + (1-\delta)\widetilde{k}_t'(\widetilde{z},\widetilde{n}) - \widetilde{b}_t'(\widetilde{z},\widetilde{n})$ is the law of motion for detrended state variables induced by the policy rules.[26]

We are now ready to define a **balanced growth path** as the limiting behavior of the model when $\frac{Z_{t+1}}{Z_t} = 1+g$ for all t. Using the results in the previous subsections, we have shown that the firm value function and decision rules are all scaled by Z_t in the sense that their detrended analogs $\widetilde{v}(\widetilde{z},\widetilde{n})$ are time-invariant. In addition, the distribution of detrended state variables $\widetilde{\Phi}(\widetilde{z},\widetilde{n})$ is constant and equal to the stationary distribution implied by (27). Finally, it is easy to see that aggregate consumption is stationary because can be written as the integral of the policy rules, which scale with Z_t, against the stationary distribution:

$$C = \int \widetilde{z}^{1-\alpha}\widetilde{k}^{\alpha} d\widetilde{\Phi}(\widetilde{z},\widetilde{k},\widetilde{b}) - (1-\pi_d)\int \Big(((1+g)\widetilde{k}_t'(\widetilde{z},\widetilde{k},\widetilde{b}) - (1-\delta)\widetilde{k}) + \widetilde{z} i_t(\widetilde{z},\widetilde{k},\widetilde{b})\Big) d\widetilde{\Phi}(\widetilde{z},\widetilde{k},\widetilde{b})$$
$$- \pi_d \int \widetilde{k} d\widetilde{\Phi}^0(\widetilde{z},\widetilde{k},\widetilde{b}),$$

where (abusing notation somewhat) $\widetilde{\Phi}(\widetilde{z},\widetilde{k},\widetilde{b})$ denotes the stationary distribution over $(\widetilde{z},\widetilde{k},\widetilde{b})$.

[26] This description is heuristic because the true transition function for the distribution should be defined over measurable sets of $(\widetilde{z}', \widetilde{n}')$. One can view the heuristic evolution (27) as the generator of that transition function if one interprets the indicator functions $\mathbb{1}$ as Dirac delta functions.

B.2 Normative Analysis: Proof of Proposition 2

We formulate the planner's problem recursively. For notational convenience, let $s = (z, k, b)$ denote a firm type. The planner's state variable is the distribution of firms, $\Phi(s)$. The planner's value function solves the Bellman equation

$$W_t(\Phi) = \max_{k'(\cdot),i(\cdot),b'(\cdot)} \frac{C^{1-\sigma}-1}{1-\sigma} + \beta W_{t+1}(T(\Phi; k'(\cdot), i(\cdot), b'(\cdot))) \text{ such that} \tag{28}$$

$$C = \int \left[(Az)^{1-\alpha}k^{\alpha} + (1-\delta)k\right]\Phi(s)ds - (1-\pi_d)\int \left[k'(s) + Azi(s)\right]\Phi(s)ds \tag{29}$$

$$- \pi_d \int k'\Phi^0(z', k', b')dz'dk'db'$$

$$(Az)^{1-\alpha}k^{\alpha} + (1-\delta)k - b - k'(s) - Azi(s) + \frac{b'(s)}{1+r_t} \geq 0 \text{ for all } s \tag{30}$$

$$b'(s) \leq \theta k'(s) \text{ for all } s \tag{31}$$

$$A = \left(\int z\Phi(s)dz\right)^{a} \tag{32}$$

$$T(\Phi; k'(\cdot), i(\cdot), b'(\cdot))(z', k', b') = \pi_d\Phi^0(z', k', b') \tag{33}$$

$$+ (1-\pi_d)\int \begin{bmatrix} (\mathbb{1}\{k' = k'(s)\} \times \mathbb{1}\{b' = b'(s)\})\times \\ (\eta(i(s))\mathbb{1}\{z' = ze^{\Delta}e^{\varepsilon}\} + (1-\eta(i(s)))\mathbb{1}\{z' = ze^{\varepsilon}\}) \end{bmatrix} p(\varepsilon)\Phi(s)ds,$$

where $p(\varepsilon)$ is the p.d.f. of ε and $T(\Phi; k'(\cdot), i(\cdot), b'(\cdot))$ is the transition function for the distribution. We denote the entire decision rule function using, e.g., $k'(\cdot)$, and the function evaluated at a particular using $k'(s)$.

The planner's problem (28) is a *functional equation* because both the state variable and choice variables are functions of the individual state s. Nuño and Moll (2018) provide conditions under which Lagrangian methods apply using Gâteaux derivatives, which we assume hold in our model as well. These derivatives are the natural extension of partial derivatives into the function space. For example, $\frac{\delta W}{\delta \Phi(s)}(\Phi)$ denotes the Gâteaux derivative with respect to the mass of households at point s, which itself is a function of the entire distribution Φ.[27] For notational simplicity we will often omit the dependence on Φ and the time subscripts that denote the dependence on the path of rates.

[27] A more explicit analogy with partial derivatives may be useful. Suppose that the state space s lay on a finite grid with N points. Then the distribution $\Phi(s)$ would be an $N \times 1$ vector, and the value function $W(\Phi) : \mathbb{R}^N \to 1$. In this case, the partial derivative $\frac{\partial W}{\partial \Phi(s_i)} : \mathbb{R}^N \to 1$ is a function of Φ as well.

We will use these tools to solve the planner's problem (28) using Lagrangian methods. Let $\lambda(s)$ denote the multiplier on the no-equity issuance constraint (45), $\mu(s)$ denote the multiplier on the collateral constraint (46), and Λ denote the multiplier on the non-rivalry externality (47). We will directly plug in the definitions of consumption (44) and the transition function for the distribution (33). With all this notation in hand, the Lagrangian is

$$\mathcal{L} = \frac{C^{1-\sigma}-1}{1-\sigma} + \int \lambda(s)\left((Az)^{1-\alpha}k^{\alpha} + (1-\delta)k - b - k'(s) - Azi(s) + \frac{b'(s)}{1+r_t}\right)ds$$
$$+ \int \mu(s)\left(\theta k'(s) - b'(s)\right)ds + \Lambda\left[\left(\int z\Phi(s)ds\right)^{a} - A\right] + \beta W(T(\Phi; k'(\cdot), i(\cdot), b'(\cdot))),$$

where, again, it is understood that C and $T(\Phi; k'(\cdot), i(\cdot), b'(\cdot))$ stand in for (44) and (33).

We proceed in two steps. First, subsection B.2.1 takes the first-order conditions with respect to all the planner's choices. Second, subsection B.2.2 characterizes those choices in terms of the marginal social value function from Proposition 2 in the main text.

B.2.1 First Order Conditions

We analyze each first-order condition separately.

Aggregate productivity The FOC with respect to aggregate productivity is

$$C^{-\sigma}\left[\int (1-\alpha)A^{-\alpha}z^{1-\alpha}k^{\alpha}\Phi(s)ds - (1-\pi_d)\int zi(s)\Phi(s)ds\right]$$
$$+ \int \lambda(s)\left[(1-\alpha)A^{-\alpha}z^{1-\alpha}k^{\alpha} - zi(s)\right]ds = \Lambda. \tag{34}$$

Going forward, it will be convenient to work with the transformed multipliers $\widetilde{\lambda}(s) = \frac{\lambda(s)}{\Phi(s)(1-\pi_d)C^{-\sigma}}$ and $\widetilde{\Lambda} = \frac{\Lambda}{C^{-\sigma}}$.[28] Plugging these in and simplifying yields

$$\widetilde{\Lambda} = \pi_d \underbrace{\int (1-\alpha)A^{-\alpha}z^{1-\alpha}k^{\alpha}\Phi(s)ds}_{=\widetilde{\Lambda}^{\text{exit}}} + (1-\pi_d)\underbrace{\int (1+\widetilde{\lambda}(s))\left[(1-\alpha)A^{-\alpha}z^{1-\alpha}k^{\alpha} - zi(s)\right]\Phi(s)ds}_{=\widetilde{\Lambda}^{\text{cont}}}. \tag{35}$$

[28] Of course, this transformed multiplier $\widetilde{\lambda}(s)$ is only defined for points with a positive mass of firms.

Innovation The FOC with respect to innovation at a particular point $i(s)$ is

$$C^{-\sigma}(1-\pi_d)Az\Phi(s)+\lambda(s)Az=\beta\int\frac{\delta W(\Phi')}{\delta\Phi'(s')}\frac{\delta T(s')}{\delta i(s)}ds'.$$

The LHS is the planner's marginal cost of higher innovation $i(s)$, which reduces consumption and tightens the no-equity issuance constraint for firm-type s. The RHS is the marginal benefit, which captures how higher innovation affects the distribution of productivity in the next period. To keep the notation manageable, we denote $T(s')=T(\Phi;k'(\cdot),i(\cdot),b'(\cdot))(s')=\Phi'(s')$. The integral is the functional-derivative extension of the chain rule: a change in $i(s)$ affects the mass of firms at each point in the state space in the next period $T(s')$, and each of those marginal changes affects the social welfare function $W(\Phi')$.

We can simplify the $\frac{\delta T(s')}{\delta i(s)}$ terms using the definition of the transition function (33). In particular, marginal changes in $i(s)$ only affect the transition function through changing the probability of success, not changing the value of the state conditional on success. Therefore, we have

$$\frac{\delta T(s')}{\delta i(s)}=\left\{\begin{matrix}(1-\pi_d)\eta'(i(s))p(\varepsilon)\Phi(s)\text{ if }s'=(ze^{\Delta}e^{\varepsilon},k'(s),b'(s)),\\ -(1-\pi_d)\eta'(i(s))p(\varepsilon)\Phi(s)\text{ if }s'=(ze^{\varepsilon},k'(s),b'(s))\\ 0\text{ otherwise}\end{matrix}\right\}$$

Plugging this into the FOC gives

$$C^{-\sigma}(1-\pi_d)Az\Phi(s)+\lambda(s)Az=\beta(1-\pi_d)\eta'(i(s))\Phi(s)\begin{bmatrix}\int\frac{\delta W(\Phi')}{\delta\Phi'(ze^{\Delta}e^{\varepsilon},k'(s),b'(s))}p(\varepsilon)d\varepsilon\\ -\int\frac{\delta W(\Phi')}{\delta\Phi'(ze^{\varepsilon},k'(s),b'(s))}p(\varepsilon)d\varepsilon\end{bmatrix}$$

Finally, dividing by $C^{-\sigma}(1-\pi_d)\Phi(s)$ and using our definition of $\widetilde{\lambda}(s)$ from above gives

$$Az(1+\widetilde{\lambda}(s))=\frac{\beta}{C^{-\sigma}}\eta'(i(s))\left[\int\frac{\delta W(\Phi')}{\delta\Phi'(ze^{\Delta}e^{\varepsilon},k'(s),b'(s))}p(\varepsilon)d\varepsilon-\int\frac{\delta W(\Phi')}{\delta\Phi'(ze^{\varepsilon},k'(s),b'(s))}p(\varepsilon)d\varepsilon\right].\tag{36}$$

Investment The FOC for capital accumulation at a particular point $k'(s)$ is

$$C^{-\sigma}(1-\pi_d)\Phi(s)+\lambda(s)=\theta\mu(s)+\beta\int\frac{\delta W(\Phi')}{\delta\Phi'(s')}\frac{\delta T(s')}{\delta k'(s)}ds'.$$

The derivatives of next period's value functions are more complicated than for innovation because a marginal change in $k'(s)$ affects the value of the state s' in the next period. Assuming we can swap the order of differentiation, we can write

$$\int \frac{\delta W(\Phi')}{\delta \Phi'(s')} \frac{\delta T(s')}{\delta k'(s)} ds' = \int \frac{\delta}{\delta k'(s)} \frac{\delta W(\Phi')}{\delta \Phi'(s')} T(s') ds'.$$

Plugging in the definition of the transition function, noting only the part of the transition function from incumbents will matter for the derivatives, and swapping the order of integration gives

$$\int \frac{\delta}{\delta k'(s)} \frac{\delta W(\Phi')}{\delta \Phi'(s')} T(s') ds' = (1-\pi_d) \int\int\int \frac{\delta}{\delta k'(s)} \frac{\delta W(\Phi')}{\delta \Phi'(s')} \begin{bmatrix} (\mathbb{1}\{k' = k'(s)\} \times \mathbb{1}\{b' = b'(s)\}) \times \\ (\eta(i(s))\mathbb{1}\{z' = ze^{\Delta} e^{\varepsilon}\} + (1-\eta(i(s)))\mathbb{1}\{z' = ze^{\varepsilon}\}) \end{bmatrix} p(\varepsilon)\Phi(s) ds ds' d\varepsilon.$$

Using only the initial state s under consideration and eliminating the values of the future state variables s' with zero probability, the integral becomes

$$(1-\pi_d)\left[\eta(i(s)) \int \frac{\delta}{\delta k'(s)} \frac{\delta W(\Phi')}{\delta \Phi'(ze^{\Delta}e^{\varepsilon}, k'(s), b'(s)))} p(\varepsilon) d\varepsilon + (1-\eta(i(s))) \int \frac{\delta}{\delta k'(s)} \frac{\delta W(\Phi')}{\delta \Phi'(ze^{\varepsilon}, k'(s), b'(s)))} p(\varepsilon) d\varepsilon\right] \Phi(s).$$

Finally, we will plug this into the FOC, and as usual divide by $C^{-\sigma}(1-\pi_d)\Phi(s)$ to get

$$1 + \widetilde{\lambda}(s) = \theta\widetilde{\mu}(s) + \frac{\beta}{C^{-\sigma}} \begin{bmatrix} \eta(i(s)) \int \frac{\delta}{\delta k'(s)} \frac{\delta W(\Phi')}{\delta \Phi'(ze^{\Delta}e^{\varepsilon}, k'(s), b'(s)))} p(\varepsilon) d\varepsilon + \\ (1-\eta(i(s))) \int \frac{\delta}{\delta k'(s)} \frac{\delta W(\Phi')}{\delta \Phi'(ze^{\varepsilon}, k'(s), b'(s)))} p(\varepsilon) d\varepsilon \end{bmatrix} \tag{37}$$

where $\widetilde{\mu}(s) = \frac{\mu(s)}{C^{-\sigma}(1-\pi_d)\Phi(s)}$.

Borrowing The FOC for borrowing at a particular point $b'(s)$ is

$$\frac{\lambda(s)}{1+r_t} = \mu(s) - \beta \int \frac{\delta W(\Phi')}{\delta \Phi'} \frac{\delta T(s')}{\delta b'(s)} ds'$$

As with capital, we can write the integral term as

$$\int \frac{\delta}{\delta b'(s)} \frac{\delta W(\Phi')}{\delta \Phi'(s')} T(s') ds' = (1-\pi_d) \int\int\int \frac{\delta}{\delta b'(s)} \frac{\delta W(\Phi')}{\delta \Phi'(s')} \begin{bmatrix} (\mathbb{1}\{k' = k'(s)\} \times \mathbb{1}\{b' = b'(s)\}) \times \\ (\eta(i(s))\mathbb{1}\{z' = ze^{\Delta}e^{\varepsilon}\} + (1-\eta(i(s)))\mathbb{1}\{z' = ze^{\varepsilon}\}) \end{bmatrix} p(\varepsilon)\Phi(s) ds ds' d\varepsilon.$$

And as in the case with capital, this integral becomes

$$(1-\pi_d) \left[\eta(i(s)) \int \frac{\delta}{\delta b'(s)} \frac{\delta W(\Phi')}{\delta \Phi'(ze^{\Delta}e^{\varepsilon}, k'(s), b'(s)))} p(\varepsilon) d\varepsilon + (1-\eta(i(s))) \int \frac{\delta}{\delta b'(s)} \frac{\delta W(\Phi')}{\delta \Phi'(ze^{\varepsilon}, k'(s), b'(s)))} p(\varepsilon) d\varepsilon \right] \Phi(s).$$

Plugging this into the FOC and dividing by $C^{-\sigma}(1-\pi_d)\Phi(s)$ yields

$$\frac{\widetilde{\lambda}(s)}{1+r_t} = \widetilde{\mu}(s) - \frac{\beta}{C^{-\sigma}} \begin{bmatrix} \eta(i(s)) \int \frac{\delta}{\delta b'(s)} \frac{\delta W(\Phi')}{\delta \Phi'(ze^{\Delta}e^{\varepsilon}, k'(s), b'(s)))} p(\varepsilon) d\varepsilon + \\ (1-\eta(i(s))) \int \frac{\delta}{\delta b'(s)} \frac{\delta W(\Phi')}{\delta \Phi'(ze^{\varepsilon}, k'(s), b'(s)))} p(\varepsilon) d\varepsilon \end{bmatrix} \tag{38}$$

B.2.2 Marginal Social Value Functions

The optimal choices to the planner's problem are given the FOCs (35), (36), (37), and (38), together with the complementarity slackness conditions. In order to arrive at the results in Proposition 2, we now use the envelope theorem to get a recursive expression for the marginal social value function $\frac{\delta W(\Phi)}{\delta \Phi(s)}$.

Differentiating the RHS of the planner's objective at the optimal policies results in

$$\begin{aligned} \frac{\delta W(\Phi)}{\delta \Phi(s)} =& C^{-\sigma} \left[(Az)^{1-\alpha} k^{\alpha} + (1-\delta)k - (1-\pi_d)\left(k'(s) + Azi(s)\right) \right] + \Lambda a \left(\int z\Phi(s) ds \right)^{a-1} z \\ &+ \beta \int \frac{\delta W(\Phi')}{\delta \Phi'(s')} \frac{\delta T(s')}{\delta \Phi(s)} ds'. \end{aligned}$$

From the definition of the transition function (33), we have

$$\int \frac{\delta W(\Phi')}{\delta \Phi'(s')} \frac{\delta T(s')}{\delta \Phi(s)} ds' = (1-\pi_d) \left[\eta(i(s)) \int \frac{\delta W(\Phi')}{\delta \Phi'(ze^{\Delta}e^{\varepsilon}, k'(s), b'(s))} p(\varepsilon) d\varepsilon + (1-\eta(i(s))) \int \frac{\delta W(\Phi')}{\delta \Phi'(ze^{\varepsilon}, k'(s), b'(s))} p(\varepsilon) d\varepsilon \right].$$

We now define $\omega(s;\Phi) = \frac{\delta W(\Phi)}{\delta \Phi(s)}$ to be the marginal social value function in the direction of $\Phi(s)$. Plugging this into the two equations above and slightly rearranging, we have

$$\omega(s;\Phi) = \pi_d C^{-\sigma}\left[(Az)^{1-\alpha}k^{\alpha} + (1-\delta)k\right] + (1-\pi_d)C^{-\sigma}\left[(Az)^{1-\alpha}k^{\alpha} + (1-\delta)k - k'(s) - Azi(s)\right]$$
$$+ \Lambda a\left(\int z\Phi(s)ds\right)^{a-1} z + \beta(1-\pi_d)\mathbb{E}^{\varepsilon}\left[\eta(i(s))\omega(s';\Phi') + (1-\eta(i(s)))\omega(s';\Phi')\right],$$

where $\mathbb{E}^{\varepsilon}[\omega(s';\Phi')] = \int \omega(s';\Phi')p(\varepsilon)d\varepsilon$ takes the expectation over idiosyncratic shocks ε.

We now define $\widetilde{\omega}(s;\Phi) = \frac{\omega(s;\Phi)}{C^{-\sigma}}$. Plugging this into the equation above yields

$$\widetilde{\omega}(s;\Phi) = \pi_d\left[(Az)^{1-\alpha} + (1-\delta)k + \widetilde{\Lambda}^{\text{exit}} a\left(\int z\Phi(s)ds\right)^{a-1} z\right] + \tag{39}$$
$$+ (1-\pi_d)\begin{bmatrix} (Az)^{1-\alpha}k^{\alpha} + (1-\delta)k - k'(s) - Azi(s) + \widetilde{\Lambda}^{\text{cont}} a\left(\int z\Phi(s)ds\right)^{a-1} z \\ + \beta\left(\frac{C'}{C}\right)^{-\sigma}\mathbb{E}^{\varepsilon}\left[\eta(i(s))\widetilde{\omega}(s';\Phi') + (1-\eta(i(s)))\widetilde{\omega}(s';\Phi')\right] \end{bmatrix}.$$

We are finally in a position to derive the equations in Proposition 2 from the main text. Let time subscripts denote the optimal value and policy functions conditional on the optimal path of the distribution $\Phi(s)$. Then, let

$$\widehat{\omega}_t(s) = \widehat{\omega}(s;\Phi_t) - b'_{t-1}(s) + (1-\pi_d)\frac{b'_t(s)}{1+r_t} + \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma}(1-\pi_d)\left(-b'_t(s) + (1-\pi_d)\frac{b'_{t+1}(s)}{1+r_{t+1}}\right) + \ldots$$

be the planner's social marginal value function plus the path of borrowing and debt repayments starting from period t. Plugging this into (39) gives the *augmented Bellman equation*

$$\widehat{\omega}_t(s) = \pi_d\left[(Az)^{1-\alpha} + (1-\delta)k - b + \widetilde{\Lambda}^{\text{exit}} a\left(\int z\Phi(s)ds\right)^{a-1} z\right] + \tag{40}$$
$$+ (1-\pi_d)\begin{bmatrix} (Az)^{1-\alpha}k^{\alpha} + (1-\delta)k - b - k'(s) - Azi(s) + \widetilde{\Lambda}^{\text{cont}} a\left(\int z\Phi(s)ds\right)^{a-1} z \\ + \frac{b'(s)}{1+r_t} + \beta\left(\frac{C'}{C}\right)^{-\sigma}\mathbb{E}^{\varepsilon}\left[\eta(i(s))\widehat{\omega}_t(s') + (1-\eta(i(s)))\widehat{\omega}_t(s')\right] \end{bmatrix}.$$

To keep notation even simpler, define $\widehat{\Lambda}^{\text{exit}} = \widetilde{\Lambda}^{\text{exit}} a\left(\int z\Phi(s)ds\right)^{a-1}$ and similarly $\widehat{\Lambda}^{\text{cont}} = \widetilde{\Lambda}^{\text{cont}} a\left(\int z\Phi(s)ds\right)^{a-1}$. In addition, let $\mathbb{E}_t$ denote the expectation over both the innovation shock and the idiosyncratic ε shocks, as in the main text. Finally, let $\widehat{\omega}_t^{\text{exit}}$ denote the terms inside the first set of brackets in (40) and let $\widehat{\omega}_t^{\text{cont}}$ second set of brackets in (40). Then we have $\widehat{\omega}_t(s) = \pi_d \widehat{\omega}_t(s)^{\text{exit}} + (1-\pi_d)\widehat{\omega}_t(s)^{\text{cont}}$, where

$$\widehat{\omega}_t^{\text{cont}}(s) = (Az)^{1-\alpha}k^{\alpha} + (1-\delta)k - b - k'(s) - Azi(s) + \frac{b'(s)}{1+r_t} + \widehat{\Lambda}^{\text{cont}}z + \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma}\mathbb{E}_t\left[\widehat{\omega}_{t+1}(s')\right] \quad (41)$$

This Bellman-like equation (41) is similar to the augmented Bellman equation (14) from Proposition 2 except that (41) is evaluated at the planner's optimal policies. Therefore, it remains to show that the planner's policies maximize the RHS of Bellman operator implied by the RHS of (41) subject to the constraints $d \geq 0$ and $b' \leq \theta k'$. But inspection of the FOCs we derived above shows that this is the case.

C Solution Algorithm

This appendix describes our numerical solution algorithm. This algorithm may be of interest to other researchers because it is extremely efficient by avoiding numerical optimizer or equation-solver.

Balanced Growth Path We first describe how we solve for a balanced growth path and then describe how we solve for a transition path starting from an arbitrary initial condition away from the BGP. Our algorithm for solving the balanced growth path iterates over candidate growth rates g^*. For each candidate growth rate, the most difficult part is solving for the individual decisions.

We solve for the individual decision rules in two steps. First, we solve for the decisions of the financially unconstrained firms. The key step in this process is iterating over the unconstrained policies $\widetilde{k}'^*_{(\text{it})}(\widetilde{z})$, $i^*_{(\text{it})}(\widetilde{z})$, and $\widetilde{v}_{(\text{it})}(\widetilde{z})$, where (it) indexes the iteration. Given the current iteration of these objects, we perform the following:

(i) Update the investment policy from (20), which becomes $\widetilde{k}'^*_{(\text{it})+1}(\widetilde{z}) = \left(\alpha\frac{\mathbb{E}_t[(\widetilde{z}')^{1-\alpha}]}{r^*-\delta}\right)^{\frac{1}{1-\alpha}}$, where $r^* = \frac{1}{\beta}(1+g^*)^{\sigma} - 1$ is the real interest rate associated with the growth rate g^*. Note that we use the previous iteration of the innovation policy $i^*_{(\text{it})}(\widetilde{z})$ to evaluate the expectation.

(ii) Update the innovation policy from (21), which can also be evaluated in closed form:

$$i^*_{(\text{it})+1}(\widetilde{z}) = \max\{0, \frac{1}{\eta_0}\log\left(\frac{\eta_0}{\widetilde{z}}\frac{1+g^*}{1+r^*}\mathbb{E}_t\begin{bmatrix}\pi_d\left(E^\varepsilon[\widetilde{n}'|\text{success}] - E^\varepsilon[\widetilde{n}'|\text{failure}]\right) + \\ (1-\pi_d)\left(E^\varepsilon[\widetilde{v}^*_{(\text{it})}(\widetilde{z}')|\text{success}] - E^\varepsilon[\widetilde{v}^*_{(\text{it})}(\widetilde{z}')|\text{failure}]\right)\end{bmatrix}\right)\}.$$

We use the new iteration of the capital policy $k'^*_{(\text{it})+1}(\widetilde{z})$ to evaluate the evolution of net worth. Note that the minimum savings policy drops out of this difference and is therefore not necessary for this computation.

(iii) Update the value function $\widetilde{v}^*_{(\text{it})+1}(\widetilde{z})$ by iterating on the Bellman operator implied by (23).

Given these unconstrained objects, we can solve for the minimum savings policy by iterating on the operator implied by (22). Finally, we can recover the unconstrained net worth cutoff $\overline{n}(\widetilde{z})$ from (26).

With these unconstrained policies in hand, we can now solve for he decision rules for all firms over the entire state space $(\widetilde{z}, \widetilde{n})$. We do so by iterating on $\widetilde{k}'_{(\text{it})}(\widetilde{z}, \widetilde{n})$, $\widetilde{b}'_{(\text{it})}(\widetilde{z}, \widetilde{n})$, $i_{(\text{it})}(\widetilde{z}, \widetilde{n})$, $\lambda_{(\text{it})}(\widetilde{z}, \widetilde{n})$, and $v_{(\text{it})}(\widetilde{z}, \widetilde{n})$:

(i) If a particular state $(\widetilde{z}, \widetilde{n})$ satisfies $\widetilde{n} > \overline{n}(\widetilde{z})$, then use the unconstrained policies and value derived above.

(ii) Solve for the policy rules assuming the collateral constraint is not binding:

- Update the capital accumulation policy from (20), which can be computed in closed form:

$$\widetilde{k}'_{(\text{it})+1}(\widetilde{z}, \widetilde{n}) = \left(\alpha\frac{\mathbb{E}_t[(\widetilde{z}'(1+1-\pi_d)\lambda_{(\text{it})}(\widetilde{z}', \widetilde{n}')]}{(1+r^*)(1+\lambda_{(\text{it})}(\widetilde{z}, \widetilde{n})) - (1-\delta)\mathbb{E}_t[(1+1-\pi_d)\lambda_{(\text{it})}(\widetilde{z}', \widetilde{n}')]}\right)^{\frac{1}{1-\alpha}},$$

where we compute the law of motion for net worth $\widetilde{n}$ and the expectation using the current iteration (it) of the policy rules.

- Update the implied $\widetilde{b}'_{(\text{it})+1}$ from the $\widetilde{d} = 0$ constraint:

$$\widetilde{b}'_{(\text{it})+1}(\widetilde{z}, \widetilde{n}) = \frac{1+r^*}{1+g^*}\left(\widetilde{z}i_{(\text{it})}(\widetilde{z}, \widetilde{n}) + (1+g^*)\widetilde{k}'_{(\text{it})+1}(\widetilde{z}, \widetilde{n}) - \widetilde{n}\right).$$

(iii) For each point in the state space $(\widetilde{z},\widetilde{n})$, which if the collateral constraint is binding at these candidate solutions, i.e. if $\widetilde{b}'_{(\text{it})+1}(\widetilde{z},\widetilde{n}) > \theta\widetilde{k}'_{(\text{it})+1}(\widetilde{z},\widetilde{n})$. If so, compute the policies with a binding collateral constraint:

- Update the capital accumulation policy from the $\widetilde{d}=0$ constraint with $\widetilde{b}'=\theta\widetilde{k}'$:

$$\widetilde{k}'_{(\text{it})+1} = \frac{\widetilde{n}-\widetilde{z}i_{(\text{it})}(\widetilde{z},\widetilde{n})}{(1+g^*)(1-\frac{\theta}{1+r^*})}.$$

- Set $\widetilde{b}'_{(\text{it})+1}(\widetilde{z},\widetilde{n}) = \theta\widetilde{k}'_{(\text{it})+1}(\widetilde{z},\widetilde{n})$.
- Recover the Langrange multiplier on the collateral constraint $\mu_{(\text{it})+1}(\widetilde{z},\widetilde{n})$ from the capital Euler equation (20).

(iv) Update the innovation policy (21) given this new iteration of the investment and borrowing policies:

$$i^*_{(\text{it})+1}(\widetilde{z}) = \max\{0, \frac{1}{\eta_0}\log\left(\frac{\eta_0}{\widetilde{z}}\frac{1+g^*}{1+r^*}\mathbb{E}_t\begin{bmatrix}\pi_d\left(E^\varepsilon[\widetilde{n}'|\text{success}]-E^\varepsilon[\widetilde{n}'|\text{failure}]\right)+\\(1-\pi_d)\left(E^\varepsilon[\widetilde{v}^*_{(\text{it})}(\widetilde{z}')|\text{success}]-E^\varepsilon[\widetilde{v}^*_{(\text{it})}(\widetilde{z}')|\text{failure}]\right)\end{bmatrix}\right)\}$$

where we evaluate the law of motion for net worth using $\widetilde{k}'_{(\text{it})+1}(\widetilde{z},\widetilde{n})$ and $\widetilde{b}'_{(\text{it})+1}(\widetilde{z},\widetilde{n})$.

(v) Update the value function $\widetilde{v}_{(\text{it})+1}(\widetilde{z},\widetilde{n})$ by iterating on the Bellman operator from (17).

(vi) Update the financial wedge $\lambda_{(\text{it})+1}(\widetilde{z},\widetilde{n})$ from (19):

$$\lambda_{(\text{it})+1}(\widetilde{z},\widetilde{n}) = (1+r^*)\mu_{(\text{it})+1}(\widetilde{z},\widetilde{n}) + (1-\pi_d)\mathbb{E}_t[\lambda_{(\text{it})}(\widetilde{z}',\widetilde{n}')].$$

While we do not have a formal proof that this iteration will converge, we find that it robustly converges for the parameterizations that we have explored. Given these policy rules, we compute the stationary distribution $\widetilde{\Phi}(\widetilde{z},\widetilde{n})$ implied by (27). We now need to compute the aggregate growth rate implied by these decision rules. By definition, the growth rate is $1+\widehat{g} = (1+g_z)^{1+a}$, where g_z is the growth rate of firm-level productivity z. We compute g_z

using the definition

$$1 + g_z = \frac{(1-\pi_d)\int z' p(\varepsilon)\Phi(s) d\varepsilon ds + \pi_d(1+g_z)\int z\Phi(s)ds}{\int z\Phi(s)ds}$$

where $s = (z, n)$ denotes the individual state vector and $Phi(s)$ is the p.d.f. of incumbent firms. The second term in the numerator reflects our assumption that the average productivity of initial entrants is equal to the average productivity of incumbents. Rearranging this expression gives

$$1 + g_z = \frac{\int z' p(\varepsilon)\Phi(s) d\varepsilon ds}{\int z\Phi(s)ds}.$$

The numerator in this integral is

$$\begin{aligned}&\int \left[\eta(i(s))e^{\Delta}e^{\varepsilon}z + (1-\eta(i(s)))e^{\varepsilon}z\right] p(\varepsilon)\Phi(s)d\varepsilon ds\\ &= e^{\sigma_\varepsilon^2/2}\left[\int z\Phi(s)ds + \int \eta(i(s))(e^{\Delta}-1)z\Phi(s)ds\right]\end{aligned}$$

where the second line uses the fact that ε is log-normally distributed independent of s. Collecting terms, we have

$$1 + g_z = e^{\sigma_\varepsilon^2/2}\left[1 + (e^{\Delta}-1)\frac{\int \eta(i(s))z\Phi(s)ds}{\int z\Phi(s)ds}\right] \quad \Longrightarrow \; 1 + \widehat{g} = (1+g_z)^{1+a_0}.$$

The above procedure defines a mapping from the current guess of the growth rate, g^*, to a new guess $\widehat{g} = f(g^*)$. A balanced growth path is a fixed point of this mapping. We compute that fixed point using the bisection method.

Transition Path We can solve for the transition path starting at an arbitrary initial distribution $\widetilde{\Phi}_0(\widetilde{z}, \widetilde{n})$ using a nonlinear equation solver. Specifically, we assume the economy converges to the balanced growth path by some finite period T and define the transition path as a sequence of $\{g_t, r_t\}_{t=0}^T$ which solves $f(\{g_t, r_t\}) = 0$, where f performs the following:

(i) Given the sequence $\{g_t, r_t\}_{t=0}^T$, solve for the individual decisions using backward iteration in the scheme described above for computing the BGP.

(ii) Given these policies and the initial distribution, $\widetilde{\Phi}_0(\widetilde{z}, \widetilde{n})$, simulate forward to get the path of distributions $\{\widetilde{\Phi}_t(\widetilde{z},\widetilde{n})\}_{t=1}^T$.

(iii) The elements of $f(\{g_t, r_t\})$ are then the aggregate consistence conditions:

$$(e^{\Delta}-1)\int \eta(i_t(\widetilde{z},\widetilde{n}))d\widetilde{\Phi}_t(\widetilde{z},\widetilde{n}) - g_t = 0$$
$$\frac{1}{\beta}\left(\frac{C_{t+1}}{C_t}\right)^{\sigma} - (1+r_t) = 0.$$

D Model Extensions

This appendix provides details of two model extensions described in the main text: adding labor as a factor of production and allowing for corporate taxes.

D.1 Labor

Adding labor extends the model in two ways. First, as discussed in the main text, the production function becomes $y_{jt} = (A_t z_{jt})^{1-\alpha} k_{jt}^{\alpha} \ell_{jt}^{\nu}$, where ℓ_{jt} is the labor used in production by firm j and $\alpha+\nu < 1$. Second, we incorporate labor supply into the household's preferences by assuming that the utility function is

$$\sum_{t=0}^{\infty} \beta^t \left[\log C_t - \chi \frac{L_t^{1+\psi}}{1+\psi}\right],$$

where χ is a scale parameter and ψ^{-1} is the Frisch elasticity of labor supply.[29]

Adding labor does not significantly alter our positive results; it simply leads to a re-interpretation of the production function in the main text. To see this, note that firms' optimal labor demand is purely static and is therefore independent of their net worth:

$$\max_{\ell_{jt}} (A_t z_{jt})^{1-\alpha} k_{jt}^{\alpha} \ell_{jt}^{\nu} - w_t \ell_{jt} \quad \Longrightarrow \quad \ell_{jt} = \left(\frac{\nu (A_t z_{jt})^{1-\alpha} k_{jt}^{\alpha}}{w_t}\right)^{\frac{1}{1-\nu}}$$

Now define variable profits $\pi_{jt} = y_{jt} - w_t \ell_{jt}$. Plugging in the above expression for optimal

[29] Given these additively separable preferences over consumption and labor supply, balanced growth requires log utility over consumption.

labor demand and simplifying yields

$$\pi_{jt} = \widetilde{\nu} \left(A_t z_{jt}\right)^{\frac{1-\alpha}{1-\nu}} w_t^{-\frac{\nu}{1-\nu}} k_{jt}^{\frac{\alpha}{1-\nu}}.$$

where $\widetilde{\alpha} = \frac{\alpha}{1-\nu}$ and $\widetilde{\nu} = \nu^{\frac{\nu}{1-\nu}} - \nu^{\frac{1}{1-\nu}}$.

The firm's problem in this extended model is isomorphic to our previous model using the new definition of net worth: $n_{jt} = \pi_{jt} + (1-\delta)k_{jt} - b_{jt}$. Importantly, net worth still grows with Z_t, facilitating the same detrending as in our baseline model. Specifically, it is easy to guess and verify that the real wage w_t scales with Z_t, which implies that the first two terms grow with $Z_t^{\frac{1-\alpha-\nu}{1-\nu}} = Z_t^{1-\frac{\alpha}{1-\nu}}$. But since capital grows with Z_t, the term involving capital grows with $Z_t^{\frac{\alpha}{1-\nu}}$. Putting these two observations together, variable profits grows with $Z_t^{1-\frac{\alpha}{1-\nu}} Z_t^{\frac{\alpha}{1-\nu}} = Z_t$.

The equilibrium of this extended model is the same as in our baseline model, except that we add the real wage w_t as another equilibrium price and add the labor market as another market clearing condition:

$$\left(\frac{w_t C_t^{-1}}{\chi}\right)^{\frac{1}{\psi}} = \int \ell_{jt} dj.$$

D.2 Taxes

We model the structure of the U.S. corporate tax code before the Tax Cuts and Jobs Act (TCJA 2017), and then consider the long-run effects of implementing the TCJA 2017. We assume firms pay a linear tax rate τ on their revenues net of tax deductions. Firms can fully deduct innovation expenditures in the period in which they occur, but investment expenditures must be gradually deducted over time according to the tax depreciation schedule.[30] Following Winberry (2021), we assume the tax deduction schedule follows a geometric depreciation process with tax depreciation rate $\widehat{\delta}$ (which may differ from economic depreciation δ). Each period, firms inherit a stock of depreciation allowances $\widehat{k}_{jt}$ from past investments and deduct the fraction $\widehat{\delta}$ of those depreciation allowances from their tax bill. In addition, firms deduct the same fraction $\widehat{\delta}$ of new investment $k_{jt+1} - (1-\delta)k_{jt}$ from their tax bill as

[30]Empirically, R&D expenditures are typically fully deducted because they primarily reflect labor costs.

well. Therefore, their total tax bill in a given period is

$$\tau \times \left(y_{jt} - A_t z_{jt} i_{jt} - \widehat{\delta}\left[\widehat{k}_{jt} + (k_{jt+1} - (1-\delta)k_{jt})\right]\right).$$

The firm carries the un-deducted portion of its investments into the next period: $\widehat{k}_{jt+1} = (1-\widehat{\delta})\left[\widehat{k}_{jt} + (k_{jt+1} - (1-\delta)k_{jt})\right]$.

In principle, we would need two new state variables, $\widehat{k}_{jt}$ and k_{jt}, in order to forecast the evolution the stock of depreciation allowances $\widehat{k}_{jt+1}$. However, we are able to bypass these additional states using the following simplifying assumption.

***Proposition* 3.** *Suppose that firms can borrow against future tax deductions at the risk-free rate r_t. Then the tax depreciation schedule only affects firm decisions through the present value of tax deductions per unit of investment:*

$$\widehat{\zeta}_t = \sum_{s=0}^{\infty}\left(\prod_{p=0}^{s}\frac{1}{1+r_{t+p}}\right)(1-\widehat{\delta})^s. \tag{42}$$

This present value alters the effective after-tax price of capital:

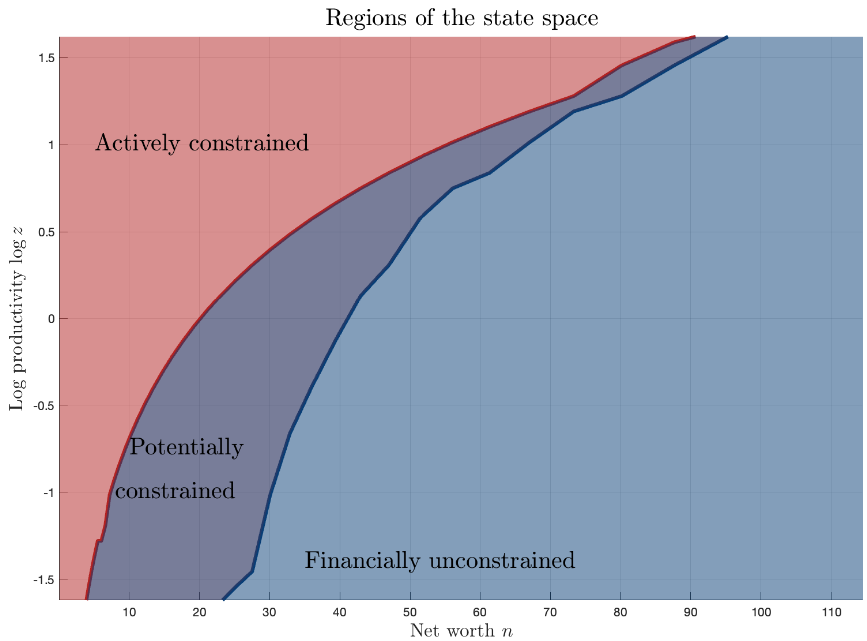
$$v_t^{cont}(z,n) = \max_{k',i,b'} n - (1-\tau\widehat{\zeta}_t)k' - (1-\tau)A_t z i + \frac{b'}{1+r_t} + \frac{1}{1+r_t}\mathbb{E}_t\left[v_{t+1}(z',n')\right] \ \text{s.t. } d \geq 0 \text{ and } b' \leq \theta k',$$

where $n' = (1-\tau)(A_t z')^{1-\alpha}(k')^{\alpha} + (1-\tau\widehat{\zeta}_t)(1-\delta)k - b'$.

Proof. The key insight of our proof is that borrowing against the stream of future tax deductions is equivalent to selling a claim on this stream to households. Since the claim is risk-free, the household is willing to pay its present value $\tau\widehat{\zeta}_t \times (k_{jt+1} - (1-\delta)k_{jt})$. Hence, each unit of investment produces $\tau\widehat{\zeta}_t$ of additional resources to the firm, lowering its after-tax price by that amount.

The financially constrained firms from Proposition 1 (with a positive financial wedge $\lambda_t(z,n) > 0$) will strictly prefer to sell the claim because their shadow value of funds is higher than the household's value of funds. However, financially unconstrained firms (with no financial wedge $\lambda_t(z,n) = 0$) will be indifferent between selling the claim or not because they value funds the same as the household. However, one can show that in this case, the

FIGURE E.1: Partition of the State Space



Notes: partition of the state space from Proposition 1 in the market BGP. Net worth n and log productivity $\log z$ have been detrended following Appendix B.

present value of the tax deductions affects firms decisions because they are indifferent over the timing (technically, their value function is linearly separable in the tax deductions; see Winberry (2021)). ■

This proposition allows us to model both temporary investment tax incentives and permanent tax reforms using changes in the present value $\widehat{\zeta}_t$.

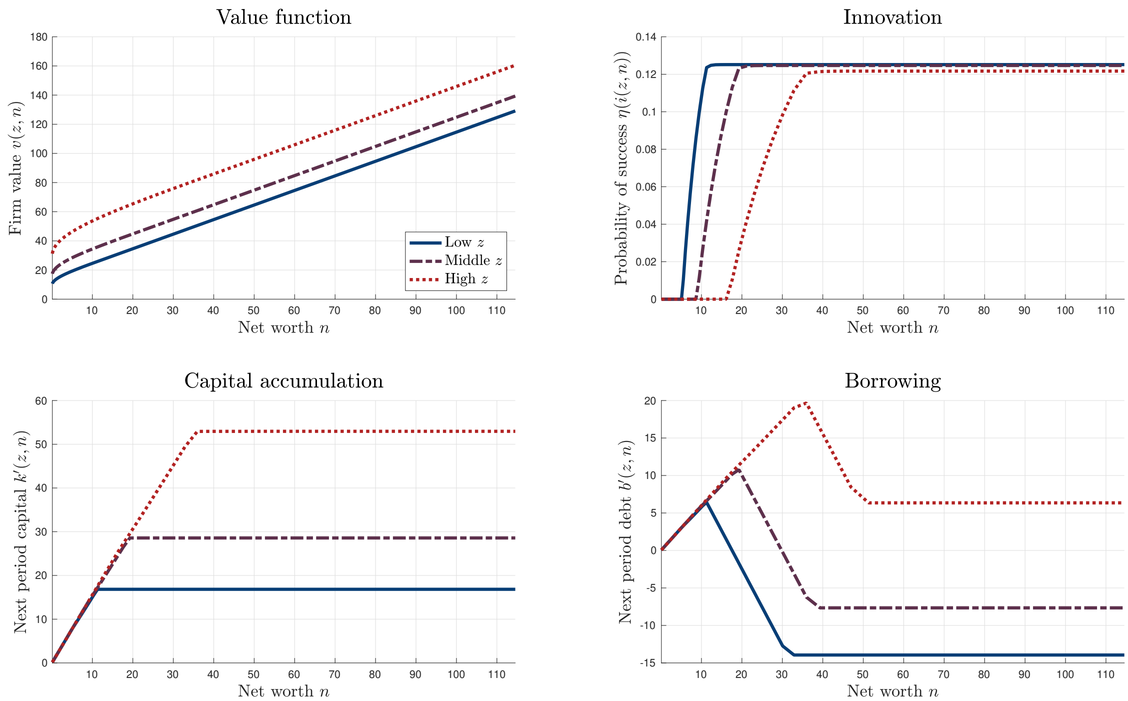
E Additional Quantitative Results

This appendix provides details of results described in the main text.

E.1 Sources of Firm Heterogeneity

Figure E.1 visualizes the partition of the state space characterized in Proposition 1. The red isocurve implicitly defines the constrained cutoff $\underline{n}(z, n)$; firms above this curve are

Figure E.2: Decision Rules



Notes: firm decision rules in the market BGP. All variables have been detrended following Appendix B.

actively constrained. The level of net worth below which firms are constrained is increasing in productivity z because higher productivity firms have a higher optimal scale of capital $k^*(z)$ and therefore a greater incentive to borrow. The blue isocurve implicitly defines the unconstrained cutoff $\overline{n}(z)$; firms below this curve are financially unconstrained. Firms in between these two isocurves are potentially unconstrained.

Decision Rules Figure E.2 plots firms' value functions and decision rules as a function of net worth n for different levels of productivity z. Consistent with the pecking order of firm growth from Section 4, firms with low net worth spend all their available resources on investment and do not innovate. The level of net worth at which firms begin innovating is increasing in their productivity because higher-productivity firms have a higher marginal product of capital and, therefore, a higher opportunity cost of innovations. While constrained, firms accumulate debt until they reach their optimal scale $k^*(z)$, at which point they use additional net worth to pay down their debt (and potentially engage in financial saving). Once firms become financially unconstrained, they adopt the minimum savings policy described in Proposition 1. Unconstrained firms' capital varies substantially, but all unconstrained firms have the same innovation rate because the cost of innovation is scaled by productivity.

Figure E.3 plots the "cash flow sensitivities" of investment and innovation, defined as $\frac{\partial k'(z,n)}{\partial n}$ and $\frac{\partial i(z,n)}{\partial n}$. Of course, unconstrained firms have sensitivities of zero because their decision rules are independent of net worth (see Figure E.2). Among constrained firms, those that do not innovate simply put all additional net worth toward investment. We can explicitly compute the resulting investment-cash flow sensitivity by differentiating the flow of funds constraint (8) with innovation $i(z,n) = 0$ and borrowing $b' = \theta k'$

$$k'(z,n) = n + \frac{\theta k'(z,n)}{1+r} \implies \frac{\partial k'(z,n)}{\partial n} = \left(1 - \frac{\theta}{1+r}\right)^{-1} \approx 1.75,$$

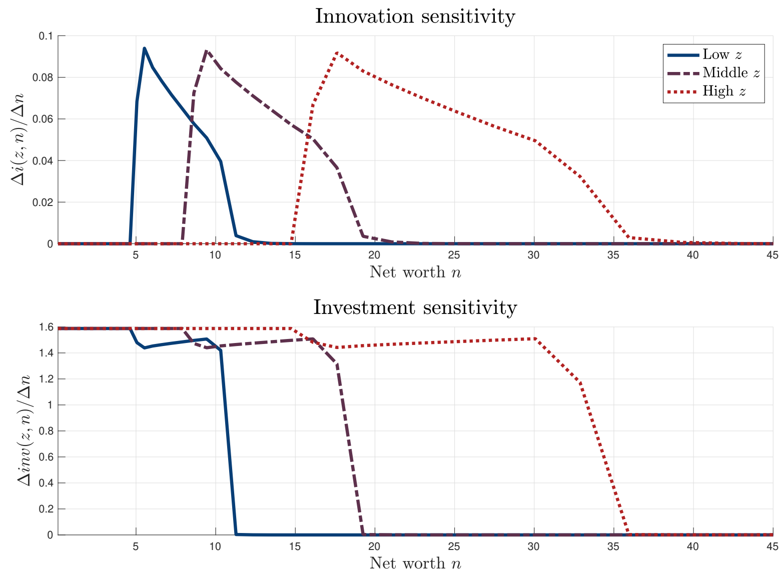
where the last approximation uses our calibrated values of $\theta = 0.45$ and $r = 0.04$. Since firms can lever up investment with borrowing, their investment-cash flow sensitivities are above one. Constrained firms with positive innovation have a smaller investment-cash flow sensitivity because they put some of the additional funds toward innovation as well:

$$k'(z,n) + A_t z i(z,n) = n + \frac{\theta k'(z,n)}{1+r} \implies \frac{\partial k'(z,n)}{\partial n} = \left(1 - \frac{\theta}{1+r}\right)^{-1} \left(1 - A_t z \frac{\partial i(z,n)}{\partial n}\right).$$

Quantitatively, Figure E.3 shows that the innovation-cash flow sensitivities are an order of magnitude smaller than the investment-cash flow sensitivities.

Lifecycle Dynamics Figure E.4 plots a sample lifecycle for a firm that enters the economy at time $t = 0$. In order to highlight the role of innovation, we assume that the firm receives no idiosyncratic productivity shocks $\varepsilon_{jt} = 0$ over this sample path. In its first few years of life, the firm has a very high investment rate and does not innovate. But as the firm ages, it exhausts its marginal product of capital, reducing its investment rate and increasing its innovation rate. These dynamics are consistent with the descriptive evidence from Figure 1 in the main text. In this particular sample path, the firm receives two successful innovations: one in year 17 and the other in year 27. Both of these successful innovations are accompanied

Figure E.3: Cash Flow Sensitivities



Notes: cash flow sensitivities computed as $\frac{\partial k'(z,n)}{\partial n}$ and $\frac{\partial i(z,n)}{\partial n}$. Derivatives computed using finite differences.

by investment spikes.

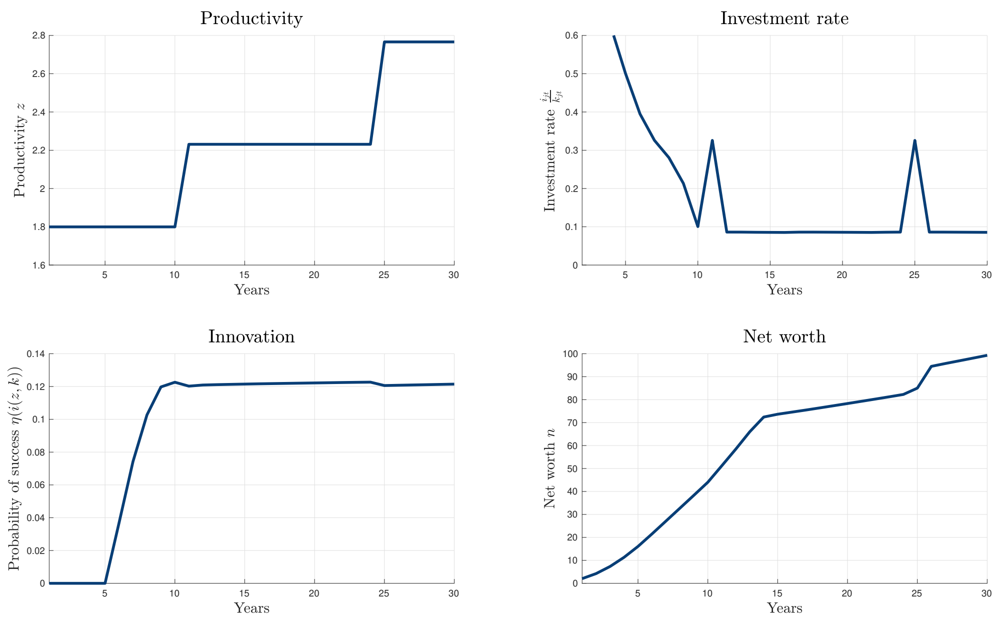
E.2 Distribution of Investment, Innovation, and Leverage

Table E.1 compares a number of moments of the stationary distribution of investment, R&D, and leverage from our model to their counterparts in the Compustat data. The model endogenously matches the average investment rate fairly well even though they it was not directly targeted in the calibration. The model's average R&D-to-sales ratio is about three as high as in the data. We chose not to target R&D spending because it is well-known to under-report total innovation expenditures. Finally, our model captures most of the dispersion in leverage and also fits the first two moments of the gross leverage distribution fairly well.

E.3 Small Growth Effects of Financial Shocks

Figure E.5 plots the effects of a transitory tightening of the financial constraint θ_t in partial equilibrium, i.e. holding the real interest rate $r_t = r^*$ fixed at its initial value. The path of the shock is in the top plotted in the top left graph. The shock reduces available financing,

Figure E.4: Sample Firm Lifecycle



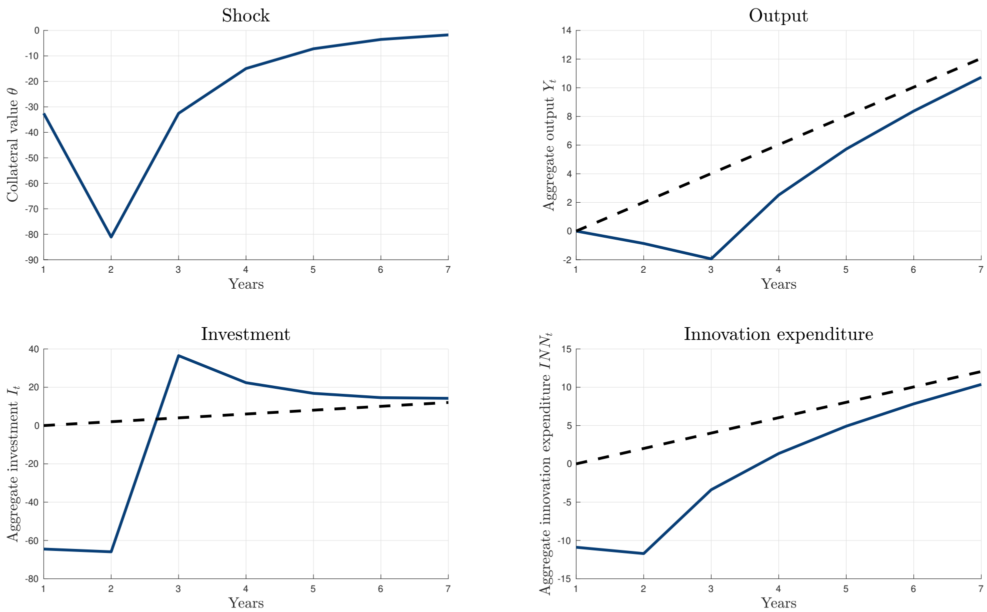
Notes: sample lifecycle profile for a firm without idisosyncratic shocks $\varepsilon_{jt} = 0$ for all j. Initially endowed with approximately average productivity and net worth among new entrants .

Table E.1
Distribution of Investment, Innovation, and Leverage

Statistic	Data	Model
Investment spending		
$\sigma(x_{jt}/k_{jt})$ (targeted)	0.13	0.11
$\mathbb{E}[x_{jt}/k_{jt}]$	0.14	0.18
$\mathbb{E}[x_{jt}/k_{jt}\|\text{spike}]$ (targeted)	0.32	0.32
R&D spending		
$\mathbb{E}[i_{jt}/y_{jt}]$	0.03	0.10
$\sigma(i_{jt}/y_{jt})$	0.06	0.03
Leverage		
$\mathbb{E}[b_{jt}/k_{jt}]$ (targeted)	0.13	0.13
$\sigma(b_{jt}/k_{jt})$	0.25	0.30
$\mathbb{E}[b_{jt}/k_{jt}]$ (gross)	0.26	0.19
$\sigma(b_{jt}/k_{jt})$ (gross)	0.22	0.17

Notes: cross-sectional statistics from stationary distribution of firms. As in the maint text, x_{jt} denotes investment, k_{jt} denotes capital, i_{jt} denotes innovation, y_{jt} denotes sales, and b_{jt} denotes borrowing. We compute gross borrowing in the model as $\max\{b_{jt}, 0\}$.

FIGURE E.5: Aggregate Transition Paths, Financial Shock (Partial Equilibrium)



Notes: aggregate transition paths following an unexpected tightening of the collateral constraint θ_t. Top left panel plots the path of θ_t. Remaining panels plot aggregate output, investment, and innovation expenditures in log-deviations from the initial period. Dashed black lines are the growth trajectory in the initial market BGP. Real interest rate $r_t = r^*$ is kept fixed at its value in the initial market BGP.

which directly lowers investment and, to some extent, innovation. However, once the shock reverts back to the steady state, investment and innovation recover relatively quickly.

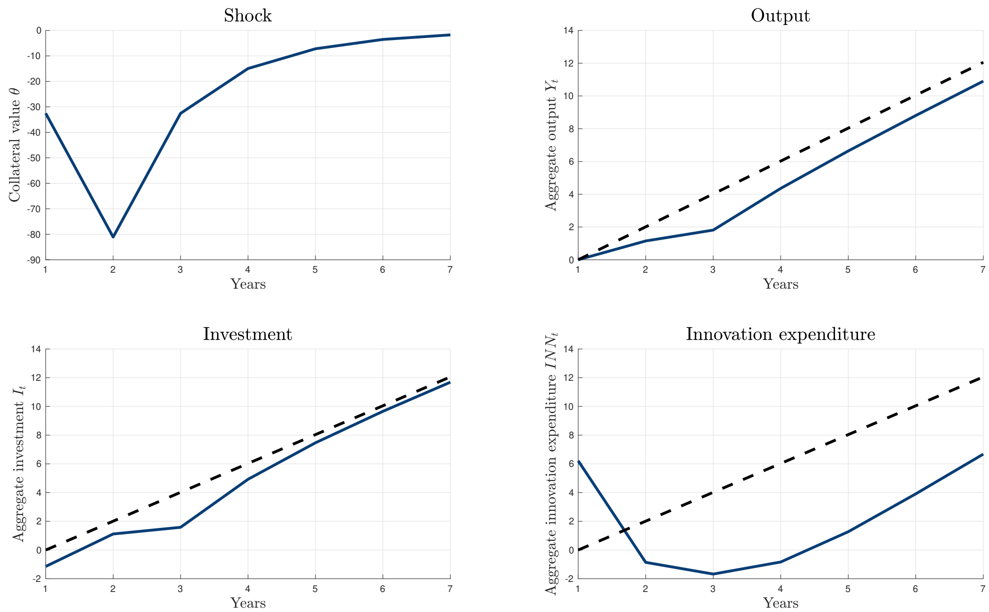
Figure E.6 performs the same exercise in general equilibrium, i.e. when the real interest rate r_t is endogenously determined. The real interest rate r_t declines given the fall in investment demand, which dampens the fall in investment. More strikingly, by raising the relative return to innovation, the lower real interest rate actually *raises* the innovation rate. Hence, the effects of the financial shock are even more short-lived than in partial equilibrium.

F Pecuniary Externalities

This appendix extends the planner's problem from Appendix B to incorporate pecuniary externalities from the real interest rate as well. We do so in the context of the model with labor to illustrate the tradeoff between growth and misallocation discussed in Section 8.

We set up the planner's problem analogously to that in Appendix B, but extend that

FIGURE E.6: Aggregate Transition Paths, Financial Shock (General Equilibrium)



Notes: aggregate transition paths following an unexpected tightening of the collateral constraint θ_t. Top left panel plots the path of θ_t. Remaining panels plot aggregate output, investment, and innovation expenditures in log-deviations from initial period. Dashed black lines are the growth trajectory in the initial market BGP.

treatment to include pecuniary externalities from both the real wage and the real interest rate. Adding the real interest rate requires that the planner includes the household's Euler equation as an additional constraint. We incorporate this constraint using the promised utility approach. In this case, the planner's state variable for any period $t \geq 0$ is the distribution of firms, $\Phi(s)$, and the promised marginal utility of consumption, M (inherited from past consumption and interest rate choices). In the initial period $t = 0$, the planner does not inherit a promised value of marginal utility M.

Planner's Problem The planner chooses $\ell^{\text{exit}}(s), \ell^{\text{cont}}(s), k'(\cdot), b'(\cdot), i(\cdot), C, r, w, A$ in order to maximize

$$W(\Phi, M) = \log C - \chi \frac{L^{1+\psi}}{1+\psi} + \beta W(T(\Phi; k'(\cdot), b'(\cdot), i(\cdot)), \frac{M}{1+r}). \tag{43}$$

Note that the planner may choose different levels of employment for exiting firms $\ell^{\text{exit}}(s)$ and continuing firms $\ell^{\text{cont}}(s)$. The transition function for the distribution of firms $T(\Phi; k'(\cdot), b'(\cdot), i(\cdot))$ is the same as in Appendix B.

The planner faces the following constraints (with associated Lagrange multipliers)

$$C = \pi_d \int \left[(Az)^{1-\alpha} k^{\alpha} \ell^{\text{exit}}(s)^{\nu} + (1-\delta)k\right] \Phi(s)ds + (1-\pi_d) \int \left[(Az)^{1-\alpha} k^{\alpha} \ell^{\text{cont}}(s)^{\nu} + (1-\delta)k\right] \Phi(s)ds$$
$$- (1-\pi_d) \int \left[k'(s) + Azi(s)\right] \Phi(s)ds - \pi_d k_0 \qquad (\times \kappa) \tag{44}$$

$$(Az)^{1-\alpha} k^{\alpha} \ell^{\text{cont}}(s)^{\nu} - w\ell^{\text{cont}}(s) + (1-\delta)k - b - k'(s) - Azi(s) + \frac{b'(s)}{1+r} \geq 0 \forall s \ (\times \mu(s)(1-\pi_d)\Phi(s)) \tag{45}$$

$$b'(s) \leq \theta k'(s) \forall s \qquad (\times \lambda(s)(1-\pi_d)\Phi(s)) \tag{46}$$

$$A = \left(\int z\Phi(s)dz\right)^{a} \qquad (\times \Lambda_A) \tag{47}$$

$$M = C^{-\sigma} \qquad (\times \Lambda_C) \tag{48}$$

$$wC^{-1} = \chi L^{\psi} \qquad (\times \Lambda_L) \tag{49}$$

It is understood that we will substitute $L = \pi_d \int \ell^{\text{exit}}(s)\Phi(s)ds + (1-\pi_d) \int \ell^{\text{cont}}(s)\Phi(s)ds$ where necessary.

***Proposition* 4.** *In the constrained-efficient equilibrium, incumbent firms' decisions solve the augmented Bellman equation*

$$\omega_t^{cont}(s) = \max_{\ell,k',b',i} (Az)^{1-\alpha} k^{\alpha} \ell^{\nu} - \widetilde{w}_t \ell + (1-\delta)k - b - k' - Azi + \frac{b'}{1+r_t}$$
$$+ \widetilde{\Lambda}_{At} z + \frac{1}{1+\widetilde{r}_t} \mathbb{E}_t[\omega_{t+1}(s')] \ \textit{s.t.} \ b' \leq \theta k' \ \textit{and}$$
$$d = (Az)^{1-\alpha} k^{\alpha} \ell^{\nu} - w_t \ell + (1-\delta)k - b - k' - Azi + \frac{b'}{1+r_t} \geq 0.$$

The strength of the non-rivalry externality is analogous to (15) *in the main text:*

$$\widetilde{\Lambda}_{At} = a \left(\int z\Phi_t(s)ds\right)^{a-1} \times \begin{bmatrix} \pi_d \int (1-\alpha) A_t^{-\alpha} z^{1-\alpha} k^{\alpha} \ell_t^{exit}(s)^{\nu} \Phi_t(s)ds \\ + (1-\pi_d) \int (1+\widetilde{\lambda}_t(s)) \left[(1-\alpha) A_t^{-\alpha} z^{1-\alpha} k^{\alpha} n_t^{cont}(s)^{\nu} - zi_t(s)\right] \Phi_t(s)ds \end{bmatrix}$$

where $\widetilde{\lambda}_t(s)$ is the firm's shadow value of funds. The pecuniary externalities are captured in the *social prices $\widetilde{r}_t$ and $\widetilde{w}_t$, which are related to the market prices r_t and w_t through*

$$\widetilde{w}_t = w_t \left(1 + \widetilde{\Lambda}_{Lt} \psi L_t^{-1}\right)$$
$$\frac{1}{1+\widetilde{r}_t} = \frac{1}{1+r_t} \frac{1 + C_{t+1}^{-1}\widetilde{\Lambda}_{Ct+1} - \widetilde{\Lambda}_{Lt+1} w_{t+1} C_{t+1}^{-1}}{1 + C_t^{-1}\widetilde{\Lambda}_{Ct} - \widetilde{\Lambda}_{Lt} w_t C_t^{-1}}$$

where $\widetilde{\Lambda}_{Ct}$ and $\widetilde{\Lambda}_{Lt}$ measure the strength of the pecuniary externalities through

$$\Omega_{t+1} = -\frac{1-\pi_d}{1 + \widetilde{\Lambda}_{Ct} C_t^{-1} - \widetilde{\Lambda}_{Lt} w_t C_t^{-1}} \int \widetilde{\lambda}_t(s) b_t'(s) \Phi_t(s) ds, \text{ where } \Omega_{t+1} \text{ is from}$$
$$\Omega_t = \Lambda_{Ct} + \frac{1}{1+r_t} \Omega_{t+1} \text{ starting from } \Omega_0 = 0$$
$$\widetilde{\Lambda}_{Lt} = \frac{1-\pi_d}{1 + \widetilde{\Lambda}_{Ct} C_t^{-1} - \widetilde{\Lambda}_{Lt} w_t C_t^{-1}} \int \widetilde{\lambda}_t(s) \ell_t^{cont}(s) \Phi_t(s) ds.$$

Proof We now take the first order conditions of the planner's problem with respect to their choice variables. The conditions for aggregate productivity, capital accumulation, innovation, and borrowing are the same as in Appendix B.

The first order condition for the real interest rate r is

$$\frac{\partial W(\Phi', M')}{\partial M'} = -\frac{(1-\pi_d)\kappa}{M} \int \widetilde{\lambda}(s) b'(s) \Phi(s) ds \tag{50}$$

where $\widetilde{\lambda}(s) = \lambda(s)/\kappa$. A higher promised marginal utility lowers the planner's value because it raises the real interest rate, which has a negative pecuniary externality to the extent that firms are financially constrained.

The first order condition for consumption differs for $t = 0$ and $t \geq 1$ because the planner does not inherit a costate M in $t = 0$. For $t \geq 1$, the first order condition is

$$\kappa = C^{-1} \left(1 + C^{-1}\Lambda_C - \Lambda_N w C^{-1}\right). \tag{51}$$

We will show below that $\Lambda_C \leq 0$ and $\Lambda_N \geq 0$, i.e. the planner discount dates in which the pecuniary externalities are more binding.

For $t = 0$, the planner does not face the constraint that $M = C^{-1}$, but understands that its choice of C_0 enters the continuation value state variable through $C_1^{-1} = \frac{C_0^{-1}}{\beta(1+r)}$. The first order condition is

$$\kappa_0 = C_0^{-1}\left(1 - \frac{1}{1+r}\frac{\partial W(\phi_1, M_1)}{\partial M_1} - \Lambda_N w C^{-1}\right). \tag{52}$$

We will show below that these two first order conditions can be collapsed to a single condition with an appropriate initial condition for Λ_C.

The first order condition for the wage is

$$\Lambda_L C^{-1} = \kappa(1-\pi_d)\int \widetilde{\lambda}(s)\ell^{\text{cont}}(s)\Phi(s)ds, \tag{53}$$

where $\widetilde{\Lambda}_L = \Lambda_L/\kappa$. The multiplier $\widetilde{\Lambda}_L$ measures the strength of the pecuniary externality from the wage. It is zero in the model without financial frictions, i.e. if $\widetilde{\lambda}(s) = 0$ for all s.

As described above, exiting firms and continuing firms may have different levels of labor. The first order condition for exiting firms is

$$(Az)^{1-\alpha}k^{\alpha}\nu\ell^{\text{exit}}(s)^{\nu-1} = \frac{\chi L^{\psi}}{\kappa}\left(1+\Lambda_L\psi N^{-1}\right) \equiv \widetilde{w} \tag{54}$$

where we use the notation $\widetilde{w}$ to denote the "shadow wage" to which the planner equates these firms' marginal products. Plugging in the expression for κ from above gives

$$\widetilde{w} = \frac{\chi L^{\psi}}{C^{-1}}\frac{1+\Lambda_L\psi K^{-1}}{1+C^{-1}\Lambda_C - \Lambda_L w C^{-1}} = w \times \frac{1+\Lambda_L\psi L^{-1}}{1+C^{-1}\Lambda_C - \Lambda_L w C^{-1}} \tag{55}$$

where the second equality plugs in the labor supply FOC. Note that, if there are no financial frictions, the planner does not value the pecuniary externalities $\Lambda_L = \Lambda_C = 0$ so $w = \widetilde{w}$, i.e. the shadow wage is equal to the marginal rate of substitution between consumption and leisure. On the other hand, if there are financial frictions so that higher labor imposes pecuniary externalities on constrained firms, then $\Lambda_L > 0$ and (as argued below) $\Lambda_C < 0$, which then implies that $\widetilde{w} > w$. The gap between the market wage and the shadow wage reflects exactly the strength of the pecuniary externalities.

The first order condition for the labor of continuing firms will also feature this shadow wage. In particular, the FOC can be simplified to

$$\widetilde{w} + \widetilde{\lambda}(s)w = (1 + \widetilde{\lambda}(s))(Az)^{1-\alpha}k^{\alpha}\nu\ell^{\text{cont}}(s)^{\nu-1}. \tag{56}$$

Note that this equation is the same as for exiting firms if the firm is financially unconstrained, i.e. $\widetilde{\lambda}(s) = 0$. Otherwise, the presence of financial constraints has two effects: it raises the marginal cost of hiring by tightening the no-equity issuance constraint and raises the marginal benefit of hiring by reducing the no-equity issuance constraint. In the market equilibrium, these forces would cancel so that the firm simply equates the market wage to the marginal product of labor. But the planner's shadow wage differs from the market wage which governs the tightening of financial constraints. This is why the planner chooses different levels of labor for exiting and continuing firms.

We now use the envelope theorem to get expressions for the marginal social value functions. For $t \geq 1$, the derivative of the value function with respect to marginal utility M evaluated at the optimum is

$$\frac{\partial W(\Phi, M)}{\partial M} = \Lambda_C + \frac{1}{1+r}\frac{\partial W(\Phi', M')}{\partial M'}, \tag{57}$$

i.e. the marginal social value of an additional promise of marginal utility is equated to the present value of the pecuniary externalities as summarized by Λ_C. Let $\Omega_t = \frac{\partial W(\Phi_t, M_t)}{\partial M_t}$ denote the partial derivative evaluated along the optimal path. Then the equation above becomes

$$\Omega_t = \Lambda_{Ct} + \frac{1}{1+r_t}\Omega_{t+1} \text{ for } t \geq 1. \tag{58}$$

We will now combine this with the FOCs for aggregate consumption for $t \geq 1$, (51), and the FOC for $t = 0$, (52):

$$\kappa_t = C_t^{-1}\left(1 + C_t^{-1}\Lambda_{Ct} - \Lambda_{Nt}w_tC_t^{-1}\right) \text{ for } t \geq 1$$
$$\kappa_0 = C_0^{-1}\left(1 - \frac{1}{1+r_0}\Omega_1 - \Lambda_N w_t C^{-1}\right) \text{ for } t = 0$$

Note that if we use (58) evaluated at $t = 0$ with the initial condition $\Omega_0 = 0$, then these equations can all be combined into

$$\Omega_t = \Lambda_{Ct} + \frac{1}{1+r_t}\Omega_{t+1} \text{ for } t \geq 0 \text{ with } \Omega_0 = 0 \tag{59}$$

$$\kappa_t = C_t^{-1}\left(1 + C_t^{-1}\Lambda_{Ct} - \Lambda_{Nt} w_t C_t^{-1}\right) \text{ for } t \geq 0 \tag{60}$$

The derivative of the value function with respect to $\Phi(s)$ is

$$\begin{aligned}\frac{\delta W(\Phi, M)}{\delta\phi(s)} &= -\chi L^{\psi}\left(\pi_d \ell^{\text{exit}}(s) + (1-\pi_d)\ell^{\text{cont}}(s)\right) + \Lambda_A a\left(\int z\phi(s)ds\right)^{a-1} z + (1-\pi_d)\mu(s)\left(\theta k'(s) - b'(s)\right)\\ &+ (1-\pi_d)\lambda(s)\left((Az)^{1-\alpha}k^{\alpha}\ell^{\text{cont}}(s)^{\nu} - w\ell^{\text{cont}}(s) + (1-\delta)k - b - k'(s) - Azi(s) + \frac{b'(s)}{1++r}\right)\\ &+ \kappa\left(\pi_d\left[(Az)^{1-\alpha}k^{\alpha}\ell^{\text{exit}}(s)^{\nu} + (1-\delta)k\right] + (1-\pi_d)\left[(Az)^{1-\alpha}k^{\alpha}\ell^{\text{cont}}(s)^{\nu} + (1-\delta)k - k'(s) - Azi(s)\right]\right)\\ &- \Lambda_L\chi\psi L^{\psi-1} + \beta\int\frac{\delta W(\Phi', M')}{\delta\Phi'(s)}\frac{\delta T(s')}{\delta\Phi(s)}p(\varepsilon)d\varepsilon ds.\end{aligned}$$

Rearranging gives

$$\begin{aligned}\frac{\delta W(\Phi, M)}{\delta\Phi(s)} = \pi_d \times &\left\{\begin{array}{l}\kappa((Az)^{1-\alpha}k^{\alpha}\ell^{\text{exit}}(s)^{\nu} + (1-\delta)k) - \chi L^{\psi}(1+\psi L^{-1}\Lambda_L)\ell^{\text{exit}}(s)\\ + \Lambda_A a\left(\int z\phi(s)ds\right)^{a-1} z\end{array}\right\} +\\ (1-\pi_d) \times &\left\{\begin{array}{l}\kappa((Az)^{1-\alpha}k^{\alpha}\ell^{\text{exit}}(s)^{\nu} + (1-\delta)k - k'(s) - Azi(s)) - \chi L^{\psi}(1+\psi L^{-1}\Lambda_L)\ell^{\text{exit}}(s) +\\ \lambda(s)\left((Az)^{1-\alpha}k^{\alpha}\ell^{\text{cont}}(s)^{\nu} - w\ell^{\text{cont}}(s) + (1-\delta)k - b - k'(s) - Azi(s) + \frac{b'(s)}{1+r}\right)\\ \mu(s)(\theta k'(s) - b'(s)) + \Lambda_A a\left(\int z\Phi(s)ds\right)^{a-1} z\\ + \beta\left(\eta(i(s))\mathbb{E}^{\varepsilon}[\frac{\delta W(\phi', M')}{\delta\Phi'(z+\Delta+\varepsilon, k'(s), b'(s))}] + (1-\eta(i(s)))\mathbb{E}^{\varepsilon}[\frac{\delta W(\Phi', M')}{\delta\Phi'(z+\varepsilon, k'(s), b'(s))}]\right)\end{array}\right\}.\end{aligned}$$

Now define $\omega_t(s) = \frac{\delta W(\Phi_t, M_t)}{\delta\phi_t(s)}/\kappa_t$ to be the marginal social value relative to the multiplier κ_t, evaluated along the optimal path. Plugging this definition into the above yields

$$\begin{aligned}\omega_t(s) = \pi_d \times &\left\{((Az)^{1-\alpha}k^{\alpha}\ell^{\text{exit}}(s)^{\nu} + (1-\delta)k - \widetilde{w}_t\ell^{\text{exit}}(s) + \widetilde{\Lambda}_A a\left(\int z\phi(s)ds\right)^{a-1} z\right\} +\\ (1-\pi_d) \times &\left\{\begin{array}{l}(Az)^{1-\alpha}k^{\alpha}\ell^{\text{exit}}(s)^{\nu} + (1-\delta)k - k'(s) - Azi(s) - \widetilde{w}_t\ell^{\text{exit}}(s) +\\ \widetilde{\lambda}(s)\left((Az)^{1-\alpha}k^{\alpha}\ell^{\text{cont}}(s)^{\nu} - w_t\ell^{\text{cont}}(s) + (1-\delta)k - b - k'(s) - Azi(s) + \frac{b'(s)}{1+r_t}\right) +\\ \widetilde{\mu}(s)(\theta k'(s) - b'(s)) + \widetilde{\Lambda}_{At} a\left(\int z\phi(s)ds\right)^{a-1} z +\\ \beta\frac{\kappa_{t+1}}{\kappa_t}\left(\eta(i(s))\mathbb{E}^{\varepsilon}[\omega_{t+1}(z+\Delta+\varepsilon, k'(s), b'(s))] + (1-\eta(i(s)))\mathbb{E}^{\varepsilon}[\omega_{t+1}(z+\varepsilon, k'(s), b'(s))]\right)\end{array}\right\}.\end{aligned}$$

We are almost ready to write this in terms of a convenient dynamic programming problem. The last thing we need to do is write the social discount factor $\beta\frac{\kappa_{t+1}}{\kappa_t}$ in a more economically interpretable form. Plugging in the expression for κ_t from (60) gives

$$\beta\frac{\kappa_{t+1}}{\kappa_t} = \beta\frac{C_{t+1}^{-1}}{C_t^{-1}}\frac{1 + C_{t+1}^{-1}\Lambda_{Ct+1} - \Lambda_{Lt+1}w_{t+1}C_{t+1}^{-1}}{1 + C_t^{-1}\Lambda_{Ct} - \Lambda_{Lt}w_t C_t^{-1}}$$

Note that $\beta\frac{C_{t+1}^{-1}}{C_t^{-1}} = \frac{1}{1+r_t}$, the private discount factor. Hence, we write the social discount factor as

$$\frac{1}{1+\widetilde{r}_t} = \frac{1}{1+r_t}\frac{1 + C_{t+1}^{-1}\Lambda_{Ct+1} - \Lambda_{Lt+1}w_{t+1}C_{t+1}^{-1}}{1 + C_t^{-1}\Lambda_{Ct} - \Lambda_{Lt}w_t C_t^{-1}}. \tag{61}$$

The wedge between the private and social discount factors is again related to the pecuniary externalities. Without financial frictions, and therefore when those two externalities do not matter, then the wedge disappears and the private discount factor equals the social discount factor. If the pecuniary externalities bind more in one period or the next, the social discount factor changes.

Summing Up The planner's allocation can be characterized as the solution to the following system of equations. Incumbent firms' decisions are the solution to the augmented Bellman equation

$$\begin{aligned}\omega_t^{\text{cont}}(s) = \max_{\ell,k',b',i}\,& (Az)^{1-\alpha}k^{\alpha}\ell^{\nu} - \widetilde{w}_t\ell + (1-\delta)k - b - k' - Azi + \frac{b'}{1+r_t} \\ &+ \widetilde{\Lambda}_{At}a\left(\int z\phi(s)ds\right)^{a-1}z + \frac{1}{1+\widetilde{r}_t}\mathbb{E}_t[\omega_{t+1}(s')] \text{ s.t. } d \geq 0, b' \leq \theta k' \end{aligned} \tag{62}$$

where $d = (Az)^{1-\alpha}k^{\alpha}\ell^{\nu} - w_t\ell + (1-\delta)k - b - k' - Azi + \frac{b'}{1+r_t}$. (Note that dividends are computed using the market wage and market interest rate, not the social ones.)

The continuation value is given by $\omega_t(s) = \pi_d\omega_t^{\text{exit}}(s) + (1-\pi_d)\omega_t^{\text{cont}}(s)$, where

$$\omega_t^{\text{exit}}(s) = \max_{\ell}(Az)^{1-\alpha}k^{\alpha}\ell^{\nu} + (1-\delta)k - \widetilde{w}_t\ell + \widetilde{\Lambda}_{At}a\left(\int z\phi(s)ds\right)^{a-1}z. \tag{63}$$

In order to compute these Bellman equations, we need to know the path of six aggregate variables: $\{g_t, w_t, C_t, \widetilde{\Lambda}_{At}, \widetilde{\Lambda}_{Ct}, \widetilde{\Lambda}_{Lt}\}$. From these variables we can compute the social prices

$$\widetilde{w}_t = w_t \left(1 + \widetilde{\Lambda}_{Nt} \psi N_t^{-1}\right) \tag{64}$$

$$\frac{1}{1+\widetilde{r}_t} = \frac{1}{1+r_t} \frac{1 + C_{t+1}^{-1}\widetilde{\Lambda}_{Ct+1} - \widetilde{\Lambda}_{Nt+1} w_{t+1} C_{t+1}^{-1}}{1 + C_t^{-1}\widetilde{\Lambda}_{Ct} - \widetilde{\Lambda}_{Nt} w_t C_t^{-1}} \tag{65}$$

where we implicitly compute L_t from $w_t C_t^{-1} = \chi L_t^{\psi}$ and r_t from $C_t^{-1} = \beta(1+r_t)C_{t+1}^{-1}$.

These six variables $\{g_t, w_t, C_t, \widetilde{\Lambda}_{At}, \widetilde{\Lambda}_{Ct}, \widetilde{\Lambda}_{Lt}\}$ must satisfy the consistency conditions

$$1 + g_t = \left(e^{\sigma_\varepsilon^2/2}\left[1 + (e^{\Delta} - 1)\frac{\int \eta(i_t(s)) z \Phi_t(s) ds}{\int z \Phi_t(s) ds}\right]\right)^{1+a} \tag{66}$$

$$\left(\frac{w_t C_t^{-1}}{\chi}\right)^{\frac{1}{\psi}} = \pi_d \int \ell_t^{\text{exit}}(s)\Phi_t(s)ds + (1-\pi_d)\int \ell_t^{\text{cont}}(s)\Phi_t(s)ds \tag{67}$$

$$C_t = \begin{bmatrix} \pi_d \int \left[(A_t z)^{1-\alpha} k^{\alpha} \ell_t^{\text{exit}}(s)^{\nu} + (1-\delta)k\right] \Phi_t(s)ds + \\ (1-\pi_d)\int \left[(A_t z)^{1-\alpha} k^{\alpha} \ell_t^{\text{cont}}(s)^{\nu} + (1-\delta)k - k_t'(s) - Azi_t(s)\right]\Phi_t(s)ds \end{bmatrix} \tag{68}$$

$$\widetilde{\Lambda}_{At} = \begin{bmatrix} \pi_d \int (1-\alpha) A_t^{-\alpha} z^{1-\alpha} k^{\alpha} \ell_t^{\text{exit}}(s)^{\nu} \Phi_t(s) ds \\ + (1-\pi_d)\int (1+\widetilde{\lambda}_t(s))\left[(1-\alpha)A_t^{-\alpha} z^{1-\alpha} k^{\alpha} \ell_t^{\text{cont}}(s)^{\nu} - z i_t(s)\right]\Phi_t(s)ds \end{bmatrix}$$

$$\Omega_{t+1} = -\frac{1-\pi_d}{1 + \widetilde{\Lambda}_{Ct} C_t^{-1} - \widetilde{\Lambda}_{Nt} w_t C_t^{-1}} \int \widetilde{\lambda}_t(s) b_t'(s) \Phi_t(s) ds, \text{ where } \Omega_{t+1} \text{ is from} \tag{69}$$

$$\Omega_t = \Lambda_{Ct} + \frac{1}{1+r_t}\Omega_{t+1} \text{ starting from } \Omega_0 = 0$$

$$\widetilde{\Lambda}_{Nt} = \frac{1-\pi_d}{1 + \widetilde{\Lambda}_{Ct} C_t^{-1} - \widetilde{\Lambda}_{Lt} w_t C_t^{-1}} \int \widetilde{\lambda}_t(s) \ell_t^{\text{cont}}(s) \Phi_t(s) ds \tag{70}$$