Innovator Heterogeneity, R&D Misallocation, and the Productivity Growth Slowdown

Stephen Ayerst*

July 19, 2023

Abstract

I examine R&D misallocation across firms and over time. I develop a model in which firm innovations differ in expected profitability and knowledge spillovers. Firms invest according to profitability, not knowledge spillovers, misallocating R&D and lowering growth. Using US patent data, I find that highly innovative firms consist of (i) firms that patent frequently and (ii) highly cited firms. Quantifying the model reveals over investment by the former, under investment by the latter, and that R&D misallocation has increased over time. Rising R&D misallocation since the 1970s explains over one third of the observed decline in aggregate productivity growth.

Keywords: Innovation, Growth, Productivity, Misallocation
JEL classification: O11, O14, O33, O41, O43

*International Monetary Fund, 700 19th Street NW, Washington DC, 20431. Email: stephen.b.ayerst@gmail.com. I thank Gadi Barlevy, Murat Celik, Faisal Ibrahim, Gaelan MacKenzie, Marti Mestieri, Scott Orr, Swapnika Rachapalli, Diego Restuccia, Baxter Robinson, Cian Ruane, and Xiaodong Zhu as well as seminar participants at the International Monetary Fund, Federal Reserve Bank of Chicago, the UofT Macro Brown Bag, the 2021 annual meeting of the Canadian Economic Association, the 2021 North American Meeting of the Econometric Society, and the 2021 Asian Meeting of the Econometric Society for useful comments. The views expressed herein are those of the authors and should not be attributed to the IMF, its Executive Board, or its management.

1 Introduction

The United States has experienced coinciding slowdowns in productivity growth (Aghion et al., 2019a) and research productivity (Bloom et al., 2020). I examine R&D misallocation caused by rising profit dispersion (as documented by De Loecker et al., 2020) as a potential driver of both slowdowns. I find that aggregate productivity growth could be 18% higher by reallocating R&D across firms and that rising R&D misallocation has lowered growth by around 11% since the 1970s, over one third of the observed decline.

I examine R&D misallocation caused by mismatches between the private and public returns to innovation. Private returns to innovating, such as market power and profits, reward innovators for incurring R&D costs and are intended to align firm incentives with the public returns to innovating, such as knowledge spillovers and output growth. However, empirical evidence shows that public and private returns often do not coincide, at least at the innovation level (Hall et al., 2005; Kogan et al., 2015; Abrams et al., 2018). In practice, many innovations provide private returns in excess of public returns. For example, evergreening patents—in which a firm extends the effective life of its patent by issuing new patents on minor modifications of the original innovation—allows the innovating firm to extend its market power but may contain almost no public value beyond the original patent.[1] In contrast, many innovations suffer the opposite issue in which public returns outweigh private returns. For example, many innovations were, or led to, breakthrough innovations but were not profitable to the innovating firm, leading these innovations to be abandoned.[2]

At the firm level, R&D becomes misallocated when the relative expected private and public returns to innovation differ across firms. Firms with high relative expected private returns (compared to public returns) over invest while firms with low relative expected private returns under invest. This mismatch could arise, for example, when some innovators are disproportionately capable of exploiting the patent or legal system (such as through evergreening) or when some innovators are more likely to produce breakthrough innovations (such as young firms). R&D misallocation may also be driven by factors unrelated to innovations, such as if large firms disproportionately extract rents through market power or if young firms lack funding to pursue innovations. Rather than take a stance on specific sources of R&D misallocation, I indirectly measure the costs of R&D misallocation through gaps in measured private and public returns.

To start, I develop a model of heterogeneous innovative firms with flexible differences in

[1] Jaffe and Lerner (2004) document many other examples of distortionary practices, such as patenting trivial or well-known innovations without prior patents.

[2] An example is that Hewlett-Packard introduced computers with touchscreen technology in the early 1980s but the technology did not become widely used until the introduction of the iPhone in the 2000s.

the expected private and public returns to innovation. The structure builds on Klette and Kortum (2004), in which a continuum of goods are produced by multi-product firms. Firms produce quality-differentiated varieties of individual goods and hire skilled R&D workers to innovate. An innovating firm invents a new higher quality variety of an existing good, displacing the incumbent producer (i.e., creative destruction). Firm R&D capabilities scale proportionally with the number of goods produced, allowing larger firms to innovate more frequently. Additionally, firms differ in ex-ante types that affect both R&D costs and the expected returns to innovation.

My main departure from the standard model is to allow for a separation of the private and public returns to innovation. Innovations are characterized by quality and knowledge improvement to the currently produced good. The private return (profitability) is determined by the quality improvement. The highest quality firm sets the limit price, based on the second highest quality firm, determining the markup and profitability of the innovated good. The public return (knowledge spillovers) is determined by the knowledge improvement. The productivity of firms is determined by the average embodied knowledge of goods in the economy. However, embodied knowledge is non-exclusionary, preventing firms from profiting from knowledge improvements. The distribution of quality and knowledge improvements differs by firm types, leading to differences in the expected private and public returns to innovation. The structure allows for both well-rewarded innovators where expected profitability matches knowledge spillovers but also poorly-rewarded innovations where expected profitability does not match knowledge spillovers. In contrast, standard innovative firm models (e.g., Klette and Kortum, 2004; Lentz and Mortensen, 2008; Acemoglu et al., 2018; Akcigit and Kerr, 2018) tie public and private returns to the same outcome.

Long-run aggregate productivity growth is driven by the average knowledge spillovers from innovations, which depends on both the innovation technologies of firms and the allocative efficiency of R&D across firms. R&D becomes misallocated because firms invest according to the expected profitability, and not the expected knowledge spillovers, of new innovations. For example, R&D could be misallocated if low knowledge spillover firms are comparatively profitable leading them to crowd out investment by high knowledge spillover firms. R&D could also be misallocated if all firms are equally profitable but differ in expected knowledge spillovers. Consequently, some dispersion in markups, and the resulting output misallocation, may be desirable to improve the R&D allocation by better aligning private and public returns. To characterize R&D misallocation, I follow Hsieh and Klenow (2009) by using the model to construct an R&D wedge that indirectly describes the allocative efficiency of R&D by comparing the the market and growth maximizing R&D allocations. This allows me to describe the growth costs of R&D misallocation without taking a direct stance on the sources of R&D misallocation.

I use administrative data from the US patent and trademark office linked with firm-level financial data to examine R&D misallocation empirically. The empirical relationship between patenting and productivity growth (e.g., Berkes et al., 2022) makes patent-related measures ideal for capturing knowledge spillovers in the model. The public returns to innovation are measured as the average citations received on patents, adjusted for the likelihood patents in different sectors and years are cited. Following the model, the private returns to innovation are measured as the average sales-per-employee of firms. The results are robust to alternative measures of profitability and private returns to innovation (Appendix B), including profit margins and market-based measures of patent value. Using these measures, I find large dispersion in the empirical firm-level R&D wedge—the measured gap between the public and private returns to innovation—that is increasing over time.

The challenge with interpreting the dispersion in R&D wedges is that patenting is relatively infrequent for most firms and so ex-ante differences in innovation outcomes are difficult to characterize. I address this challenge by extending the model to incorporate patenting and similar ex-ante differences in innovation outcomes. As an intermediate step, I construct firm types based on innovation characteristics that are used to calibrate the model. Firms are grouped based on innovation frequency and knowledge spillovers, measured by patenting rate and average citations-per-patent, using a k-means clustering algorithm. Firms are divided into three types: (1) low type (low frequency, low knowledge spillovers) firms; (2) high volume firms, characterized by high patent counts; and (3) high knowledge firms, characterized by high average citations-per-patent. The separation of high volume and high knowledge firms departs from the common assumption that high-ability innovators perform better across all dimensions (as in, for example, Acemoglu et al., 2018). Across firm types, knowledge spillovers are highest for high knowledge firms (even after accounting for sector differences) while profitability is highest for high volume firms. Additionally, the relative profitability of high volume firms has increased over time while the relative knowledge spillovers have remained relatively stable for all three types over time, implying a similar increase in R&D wedge dispersion as with the firm-level measure.[3]

I calibrate the model to match characteristics of the low type, high volume, and high knowledge firms in the late period (1991-2005) where R&D misallocation is more severe. I allow for the possibility that firm types are misclassified in the construction of model moments by simulating firm data and using inferred firm types, based on empirical cutoffs, rather than the true firm types. Relative sales-per-employee and citations-per-patent discipline the

[3]High volume firms also tend to be larger firms in terms of sales and employment, consistent with the rise in market power over this period (De Loecker et al., 2020; Cavenaile et al., 2020).

profitability and knowledge spillovers by firm type while the relative patenting frequency of firm types disciplines the product distribution. Research productivity and R&D wedges point to growth gains from reallocating R&D from high volume to high knowledge firms.

I use the model to measure the potential growth gains from reallocating R&D to match expected knowledge spillovers. Holding the distribution of firms and products fixed, reallocating R&D increases growth by around 18% relative to its initial value. The gain is driven by the increased R&D allocation to high knowledge firms. Growth further increases to almost double its initial value when the distribution of firms endogenously adjusts to the new R&D allocations because high knowledge firms expand into new product lines, increasing their R&D scale. This mechanism is similar to Bento and Restuccia (2017) and Ayerst (2020) finding that distortions amplify the productivity cost of output misallocation because productive firms are discouraged from investing in productivity improvements.

Finally, I use the model to quantify the costs of changing R&D misallocation over time. I consider a simple experiment in which the expected profitability of firm types are adjusted to match the early period (1976-1990) values in the data. The experiment holds R&D technologies and the stock of R&D resources fixed, such that the potential growth capabilities of the benchmark economy and counterfactual economy are the same. Rising R&D misallocation explains a drop in aggregate productivity growth of 11% between the early and late periods, driven primarily by a sharp rise in the profitability of high volume firms. The result explains just over one third of the decline in aggregate productivity growth and around one sixth of the decline in research productivity over this period.[4] The result is also robust to re-calibrating the model to the early period (Appendix D.1).

Related literature. My paper relates to the literature building on Klette and Kortum (2004) to examine innovator heterogeneity (Lentz and Mortensen, 2008; Acemoglu et al., 2016a; Akcigit and Kerr, 2018; Garcia-Macia et al., 2019). In a closely related paper, Acemoglu et al. (2018) examine R&D misallocation across innovative firms with ex-ante differences in research productivity. They focus on a different source of R&D misallocation in which less innovative firms use R&D resources for fixed operating expenses, crowding out investment by more innovative firms. Their measure reflects an empirical gap between firm growth and R&D intensity but assumes that the (private and public) returns to innovating are the same for all firms. In another closely related paper, Akcigit et al. (2019) examine

[4]The contribution of R&D misallocation to productivity growth is much larger than research productivity because research productivity also accounts for inputs. Bloom et al. (2020) document a large increase in effective researchers over this time period, explaining the differences in contribution.

optimal R&D policies from a mechanism design perspective.[5] In their setting, firms differ in both research productivity and knowledge spillovers—although, knowledge spillovers are tightly linked with profitability—while R&D inputs are not perfectly observable by the policymaker. In contrast, I examine R&D misallocation caused by mismatches between the expected profitability and knowledge spillovers of innovations. These mismatches are important because the firms that generate the largest knowledge spillovers are not necessarily the most profitable firms.[6] This channel of R&D misallocation leads to a quantitatively large loss in growth and research productivity that has not previously been documented.

My paper also relates to the misallocation literature. Starting with Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), the literature documents large productivity costs of misallocated factors of production. Edmond et al. (2015), Peters (2019) and MacKenzie (2019) examine output misallocation arising from dispersion in markups across producers. I examine R&D misallocation arising from ex-ante heterogeneity in profits that lowers aggregate productivity growth. I find that some output misallocation may be beneficial to align the private and public returns to innovating.[7] The quantitative welfare cost of lower growth due to R&D misallocation is an order of magnitude larger than from lower productivity due to output misallocation.

Finally, my paper relates to the literature examining the productivity growth slowdown in advanced economies (Gordon, 2016; Aghion et al., 2019a) and coinciding macroeconomic trends, such as declining business dynamism, falling labor share and rising markups, among others (e.g., Aghion et al., 2019b; Akcigit and Ates, 2019; Cavenaile et al., 2020; De Ridder, 2020; Peters and Walsh, 2021). I examine rising R&D misallocation as a novel explanation for the productivity growth slowdown in which the increasing profit dispersion leads to larger dispersion in the mismatches between the private and public returns to innovation. Rising R&D misallocation explains growth falling by 11% relative to its initial value, corresponding to over one third of the growth slowdown in the data. Rising R&D misallocation also provides a micro-foundation for the declining research productivity documented by Bloom et al. (2020), explaining around one sixth of the decline in research productivity. That is,

[5] See also papers examining the relationship between patent rewards and innovation (e.g., Cornelli and Schankerman, 1999; Scotchmer, 1999).

[6] I find a similar pattern of misallocation as Acemoglu et al. (2018) in that R&D should be reallocated from high volume firms (which tend to be older and less innovative) to high knowledge firms (younger and more innovative). However, the comparatively high profitability of the high volume firms suggests that raising incumbent taxation as suggested by Acemoglu et al. (2018) may be ineffective in practice.

[7] Chen et al. (2018) and Konig et al. (2018) examine R&D misreporting by Chinese firms. Additionally, Konig et al. (2018) examine R&D misallocation as arising from output wedges, which has similar elements to my analysis. My analysis differs by focusing on ex-ante firm heterogeneity in innovative capabilities. I also show that eliminating output wedges would not be enough to eliminate R&D wedges and could lead to higher R&D wedges when firms differ in innovative capabilities.

ideas are becoming harder to find, in part, because the firms that generated those ideas in the past generate relatively fewer ideas today.

Outline. The remainder of this paper is organized as follows. Section 2 introduces the model and characterizes R&D misallocation. Section 3 presents the empirical analysis. Section 4 calibrates the model and discusses the model's fit. Section 5 uses the model to quantify R&D misallocation across firms and time. Section 6 concludes.

2 Model

I develop a model to examine R&D misallocation. The structure follows Klette and Kortum (2004) in which firms innovate on existing goods to become the new leading-edge producer. Firms differ in their ex-ante type which determines the relative cost and expected returns of R&D. The key feature of the model is that R&D returns are separated into private returns that benefit only the innovating firm and public returns that benefit the entire economy. R&D misallocation is caused by differences in the expected private and public returns.

2.1 Economic Environment

Time is continuous and indexed by $t \in [0, \infty)$. The economy is populated by an infinitely-lived representative households with a unit mass of unskilled workers and a mass S of skilled workers as well as an endogenously determined mass of firms.

Preferences and final good technology. The representative household has preferences

$$U([C(t)]) = \int_0^\infty e^{-\rho t} \ln C(t) dt,$$

where ρ is the rate that the household discounts future utility. Unskilled workers provide labor for the production of goods and skilled workers provide labor for R&D.[8]

The final good is produced using a continuum of intermediate goods, indexed by $j \in [0, 1]$, that differ in terms of quality q_j and embodied knowledge k_j. The model nests Klette and Kortum (2004) for the case where the embodied knowledge equals quality, $k_j = q_j$. The final

[8]The fixed stock of skilled labor restricts total R&D investment. This implicitly abstracts from aggregate underinvestment in R&D that would occur if R&D used output as an input (as in Romer, 1990). Atkeson and Burstein (2018) find that the elasticity between aggregate R&D spending and growth is quantitatively small across a wide range of parameter values.

good production technology is given by

$$\ln Y(t) = \int_0^1 \ln \left[\sum_{f \in \mathcal{F}_j} \frac{q_{j_f}}{\bar{q}(t)} y_{j_f}(t) \right] dj,$$

where y_{j_f} is the quantity of good j purchased from firm f and $\mathcal{F}_j$ is the set of potential producers of good j. The inclusion of average quality $\bar{q} = \exp \int_0^1 \ln q_j dj$ indicates that the relative quality of goods reallocates preferences without contributing to long-run growth.

Intermediate firms f produce multiple goods and innovate to add new goods to their portfolio. Intermediate goods are produced using a linear technology $y_{j_f} = \bar{k}(t)\ell_{j_f}$, where ℓ_j is unskilled labor hired to produce good j. The average knowledge spillover $\bar{k}(t) = \exp \int_j \ln k_j dj$ captures the spillover from previous innovations to the productivity of the economy (i.e., standing on the shoulders of giants). The separation of knowledge and quality allows for the determinants of profit and productivity growth to be separated. Intuitively, the features of an innovation that make it profitable, captured by q_j, may or may not coincide with the features that make it useful for long-run growth, captured by k_j.

Innovation. Each firm f owns the blueprints to produce a set of goods $\mathcal{J}_f$ with qualities $\{q_j\}_{j \in \mathcal{J}_f}$ and embodied knowledge $\{k_j\}_{j \in \mathcal{J}_f}$. Additionally, firms differ in innovative types $a \in \mathcal{A}$ which determine the expected returns and innovation costs. Firms stochastically transition between types according to a Markov process where a type a firm becomes a type a' firm at rate $\gamma_{a,a'}$, with transition matrix $\Gamma = [\gamma_{a,a'}]$.

Firms invest in R&D to expand the number of goods that they produce. A firm f with type a_f that operates n_f product lines and hires S_f researchers produces new innovations at rate $X_{a_f}(S_f, n_f) = \left(S_f/\psi_{a_f}\right)^{1/\zeta} n_f^{1-1/\zeta}$, where $\zeta > 1$ determines the elasticity of innovations to skilled labor. This implies that the skilled labor per product line S/n is

$$n_f s_{a_f}(x) = n_f \psi_{a_f} x^{\zeta}, \tag{1}$$

where $x = X/n$ is the innovation rate per product line. The assumption that costs scale linearly with the number of product lines implies that Gibrat's law hold in equilibrium (as in Klette and Kortum, 2004).[9] Intuitively, this captures that knowledge gained from production is useful for innovating. In order to receive the increased R&D scale firms must both produce some output and pay a flow cost $w_s(t)B_a$ where the skilled wage $w_s(t)$ is used as a scaling

[9] In the quantitative analysis this assumption dictates the ability of large firms to absorb R&D resources. Akcigit and Kerr (2018) provide evidence that Gibrat's law does not hold in the data with larger firms having lower research productivity. This mechanism would then amplify the degree of R&D misallocation I find as the lower R&D scale of larger firms would increase the gains from reallocating resources to small firms.

term for mathematical convenience.[10]

At rate X, the firm draws an innovation characterized by (λ, η), drawn from a type-specific distribution $\Phi_a(\lambda, \eta)$. The value $\lambda > 1$ determines the quality q'_j of the new variety of good j that the innovating firm produces $q'_j = \lambda q_j$, where q_j is the quality of the currently produced variety of good j. The value $\eta > 1$ determines the embodied knowledge k'_j of the new variety of good j that the firm produces $k'_j = \eta k_j$, where k_j is the embodied knowledge of the currently produced good j. In equilibrium, the quality gap λ determines the profitability of good j to the innovating firm while the knowledge spillover η determines the contribution of the innovation to aggregate growth.

The separation of λ and η allows for flexible relationships between the knowledge spillovers and profitability of innovations, in contrast to standard models (e.g., Klette and Kortum, 2004) where both returns depend on a single value. This allows for some innovations to be useful sources of knowledge spillovers but unprofitable, as in the example of touchscreens in the introduction, and other innovations to be profitable but have low knowledge spillovers, as in the example of evergreening patents. Additionally, while λ and η are distinct, they are not necessarily independent implying that the same features that may contribute to an innovations profitability may also generate knowledge spillovers.

Entry and exit. New firms enter by innovating. Potential entrants hire s_e researchers to create new varieties of goods at a rate $x_e = s_e/\psi_e$. Successful entrants draw their innovative type a from a distribution $\tilde{h}_a$. After drawing its type, an entrant draws a number of entry goods $\bar{n}_a$ and unique innovation characteristics (λ, η) for each good from the distribution $\Phi_a(\lambda, \eta)$. Firms enters at rate $x_{e,a}(n) = x_e \tilde{h}_a / n$ such that the expected rate at which a good is produced by a type a entrant is equal to $\tilde{h}_a x_e$. This implies that the resulting product distribution is equivalent to the case where firms enter with a single product. Multi-product entry helps with the calibration but does not affect the model results.

Firms exit the market when they no longer actively produce any goods. Firms lose goods at rate δ, which is determined by the rate that other firms innovate. Let $\bar{x}_a$ be the product-weighted average innovation rate of type a firms. The creative destruction rate δ is

$$\delta = \sum_{a \in \mathcal{A}} \left[h_a \bar{x}_a + \tilde{h}_a x_e \right], \tag{2}$$

[10] Assumption (A) sets the cost to be equal to the option value of R&D. Higher values of B_a would imply that the leading-edge firm acts as a monopolist because other firms would never enter (similar to the assumption in Akcigit and Kerr, 2018). On the other hand, lower values of B_a would imply that firms may enter and produce near zero output in order to receive the benefit of increased R&D scale, which would be mathematically intractable as the scale of R&D would increase exponentially over time.

where h_a is the share of goods produced by type a firms.

2.2 Market Structure and Equilibrium

I focus on the stationary Balanced Growth Path Equilibrium in which all variables and prices grow at constant rates. I write variables without t where it does not create confusion.

2.2.1 Production and Prices

The final consumption good is taken as the numeraire. The household's problem implies that the Euler equation is $\dot{C}/C = r - \rho$.

Each firm manages a portfolio of goods that they potentially produce and competes with other variety producers of good j in Bertrand competition. Because individual varieties of good j are quality-adjusted perfect substitutes, the final good producer only purchases the variety of good j with the lowest price-per-quality, p_{j_f}/q_{j_f}. The quality leader sets the limit price for good j according to the quality gap between the leader's (q_{j_L}) and follower's (q_{j_F}) goods, $\lambda_j = q_{j_L}/q_{j_F}$, and the marginal cost of production $w_u/\bar{k}$. The price of good j is then $p_j = \lambda_j w_u/\bar{k}$ where the markup is λ_j. Because demand for each good is Y, it follows that firm-level output is $y = Y/(\lambda w_u/\bar{k})$ and demand for unskilled labor is $\ell = Y/\lambda w_u$.

Product and markup distribution. The product distribution h_a describes the distribution of the firm types across the producers of each good. The law of motion is

$$\dot{h}_a = x_e\tilde{h}_a + x_a h_a - \delta h_a + \sum_{a'\in\mathcal{A}} \gamma_{a',a} h_{a'} - \sum_{a'\in\mathcal{A}} \gamma_{a,a'} h_a, \tag{3}$$

where $\dot{h}_a = 0$ in the stationary equilibrium. The first two terms are the probability a type a good is added through entry or incumbent innovation; the third term is the probability that a type a firm loses a good; and the fourth and fifth term are the inflow and outflow of type a goods based on firm transitions across types. The related firm size n and type a distribution is described in Appendix A.1.

The markup distribution $\kappa(\lambda)$ describes the distribution of markups $\lambda_j = q_{j_L}/q_{j_F}$ across goods. The law of motion is

$$\dot{\kappa}(\lambda) = -\delta\kappa(\lambda) + \sum_{a\in\mathcal{A}} [h_a x_a + \tilde{h}_a x_e] \int_\eta \Phi_a(\lambda, d\eta), \tag{4}$$

where $\dot{\kappa}(\lambda) = 0$ in the stationary equilibrium. The first term is the rate at which firms with quality lead less than or equal to λ are replaced by an innovating firm. The second term is the arrival rate of innovations with quality leads less than λ, which depends on the product distribution, the innovation rates, and the marginal distribution of quality improvement.

Allocations, prices, and output. The unskilled labor market clearing condition requires that the total mass of unskilled labor is hired by firms, $1 = \int_j \ell_j dj$. Following the distribution of markups $\kappa(\lambda)$ and demand for unskilled labor, the wage rate that clears the market is

$$w_u = \bar{k} \exp\left(\int_\lambda \ln \lambda^{-1} \kappa(d\lambda)\right).$$

The firm-level prices and allocations can then be written as

$$p(\lambda) = \lambda \int_\lambda \ln \lambda^{-1} \kappa(d\lambda), \qquad y(\lambda) = \bar{k} \frac{\lambda^{-1}}{\int_\lambda \lambda^{-1} \kappa(d\lambda)}, \qquad \ell(\lambda) = \frac{\lambda^{-1}}{\int_\lambda \lambda^{-1} \kappa(d\lambda)}. \tag{5}$$

The prices and allocation depend on the distributions of markups λ and the average stock of embodied knowledge $\bar{k}$ in the economy. The expressions also imply that firm-level sales-per-employee are $p(\lambda)y(\lambda)/\ell(\lambda) \propto \lambda$, which I use in the calibration section to estimates the distribution of private returns λ.

Aggregate output can be used for either consumption or the fixed cost of firm operations, $Y = C + w_s \sum_a B_a h_a$. Aggregate output is equal to

$$Y = \chi \bar{k}, \tag{6}$$

where $\chi = \exp(\int_\lambda \ln \lambda^{-1} \kappa(d\lambda))/\int_\lambda \lambda_i^{-1} \kappa(d\lambda)$ is an aggregate term that captures output misallocation driven by markup dispersion (as in, Peters, 2019). The term χ will be less than one when there is markup dispersion and equal to one when all firms charge the same markup.

The expression for output in (6) shows that output (and consequently productivity) growth depends only on growth in average knowledge $\bar{k}$. In this regard, the production function disentangles the private benefits that firms receive from innovating from the generated knowledge spillovers.[11] However, given that there are no restrictions on the joint distribution $\Phi_a(\lambda, \eta)$ of quality improvements λ and knowledge spillovers η, this should not be taken as placing restrictions on what drives growth. Intuitively, the same features of an innovation that cause it to be profitable may also generate knowledge spillovers.

[11] Appendix A.3 considers a more generalized CES production for the final consumption good. The alternative production function creates a link between knowledge spillovers and profitability as lower cost (high k_j) firms can charge lower prices to increase market share. However, the relative importance of this channel on aggregate productivity growth is quantitatively small.

2.2.2 Research and Development

Innovation problem. The profitability of a good with markup λ is $\Pi(\lambda) = [1 - \lambda^{-1}]Y = \pi Y$ where $\pi = 1 - \lambda^{-1}$ is a scaled measure of profitability. Profits depend on the quality improvement of the innovation λ and not the embodied knowledge improvements of the innovation η or average knowledge $\bar{k}$.

To simplify notation, I refer to firms by their type $a \in \mathcal{A}$, the profitability of actively produced goods $[\pi_i]_{i=1}^n$, and the cardinality of this set n. The dynamic problem of firms is to hire skilled labor to maximize firm value by adding new goods. The value of a firm is

$$\begin{aligned} rV_a([\pi_i]_{i=1}^n, n) - \dot{V}_a([\pi_i]_{i=1}^n, n) = \\ \max_x \sum_{i'=1}^{n} & \begin{bmatrix} \pi_{i'} Y - w_s B_a + \delta \left[V_a([\pi_i]_{i=1}^n / \{\pi_{i'}\}, n-1) - V_a([\pi_i]_{i=1}^n, n) \right] \\ + x \left[\mathbb{E}_a V_a([\pi_i]_{i=1}^n \cup \{\tilde{\pi}\}, n+1) - V_a([\pi_i]_{i=1}^n, n) \right] - w_s s_a(x) \end{bmatrix}, \\ + \sum_{a' \in \mathcal{A}} & \gamma_{a,a'} \left[V_{a'}([\pi_i]_{i=1}^n, n) - V_a([\pi_i]_{i=1}^n, n) \right] \end{aligned} \tag{7}$$

where w_s is the wage of skilled labor. A "/" indicates a set without the following element and a "$\cup$" indicates a set with the following element added. The value of a firm depends on the expected profits of its current portfolio of goods (first line), the expected present value of innovation (second line), and the expected value from changing types (third line).

The entry problem is given by

$$V_e = \max_{x_e} x_e \sum_{a \in \mathcal{A}} \left[\frac{\tilde{h}_a}{\bar{n}_a} \mathbb{E}_a \left[V([\tilde{\pi}_i]_{i=1}^{\bar{n}_a}, \bar{n}_a) \right] - V_e \right] - w_s \psi_e s_e(x_e). \tag{8}$$

The first term is the probability of successfully drawing an innovation x_e multiplied by the expected net value of entry. The second term is the cost of generating a stream of innovations x_e. In equilibrium the entry condition implies that the value V_e of being an entrant is non-positive and equal to zero whenever entry is positive, $x_e > 0$.

Skilled labor is used for incumbent and entrant R&D. The skilled labor market clearing requires

$$S = \psi_e x_e + \sum_a h_a \psi_a x_a^{\zeta}.$$

Growth rate. Along the balanced growth path, the distribution of markups is stationary implying that output growth is driven by growth in the average knowledge stock $\bar{k}$. The growth rate of knowledge $\bar{k}$ is given by

$$g = \frac{d \ln \bar{k}(t)}{dt} = \sum_{a \in \mathcal{A}} [h_a x_a + \tilde{h}_a x_e] \ln \bar{\eta}_a. \tag{9}$$

where $\ln \bar{\eta}_a = \int_\eta \int_\lambda \ln \eta \Phi_a(d\lambda, d\eta)$. The growth rate depends on: (i) the average quality of innovations $\bar{\eta}_a$; and (ii) the frequency with which these ideas occur $h_a x_a + \tilde{h}_a x_e$. In a model without type heterogeneity, the growth rate reduces to the average step size η multiplied by the creative destruction rate δ. The addition of firm heterogeneity implies an additional factor: (iii) the allocation of innovative activities across firm types. Growth is then higher when high expected knowledge spillovers (high $\bar{\eta}_a$) firms invest relatively more. Proposition 1 shows that the relative investment depends on the expected profitability of firm innovations.

2.2.3 BGP Equilibrium

Definition 1. *A balanced growth path equilibrium consists of the values*

$$\{C, Y, w_u, w_s, r, p(\lambda), y(\lambda), \ell(\lambda), x_e, x_a, h_a, g, \delta, \kappa\}$$

for $a \in \mathcal{A}$, $\lambda > 1$, $t \in [0, \infty)$, such that: (i) consumption C maximizes household utility; (ii) the price $p(\lambda)$, output $y(\lambda)$, and labor $\ell(\lambda)$ maximize firm profits for each quality gap λ; (ii) the innovation rate x_a maximizes firm value for each type a; (iii) firm entry x_e maximizes entrant value; (iv) the stationary product and markup distributions follow (3) and (4) for $\dot{h}_a = \dot{\kappa}(\lambda) = 0$; (v) the growth rate g and creative destruction rate δ are given by (9) and (2); (vi) the final and intermediate goods markets clear; (vii) the unskilled and skilled labor markets clear.

Characterization. I focus on the positive entry case and further assume that

$$B_a = \psi_a(\zeta - 1) \left[\frac{\psi_e \bar{\pi}_a}{\psi_a \zeta \left[\sum_{a'} \bar{\pi}_{a'} \tilde{h}_{a'} \right]} \right]^{\frac{\zeta}{\zeta - 1}}, \tag{A}$$

where $\bar{\pi}_a = \int_\eta \int_\lambda \left[1 - \lambda^{-1}\right] \Phi_a(d\lambda, d\eta)$ is the expected profitability of a new good for a type a firm. The assumption sets the fixed costs paid by firms to the option value of innovations. Section 4.5 shows this assumption has a quantitatively small impact on the results.

The equilibrium wage rate of skilled labor normalized by output is denoted by $\omega = w_s/Y$.

Under the above assumption, the normalized wage rate is

$$\omega = \frac{1}{\psi_e(\rho+\delta)} \sum_a \bar{\pi}_a \tilde{h}_a. \tag{10}$$

The wage rate depends on the expected profitability of newly created firms $\sum_a \bar{\pi}_a \tilde{h}_a$, the adjusted discount rate $\rho + \delta$, and the entry technology ψ_e.

Proposition 1 characterizes the solution to the firm's problem.

Proposition 1. *Assume that entry is positive and (A) holds, the firm's problem is solved by the value function*

$$V_a([\pi_i]_{i=1}^n, n) = \sum_{i'=1}^{n} \left[\frac{\pi_{i'}}{\rho+\delta}\right] Y(t) \tag{11}$$

and the innovation rate

$$x_a = \left[\frac{\psi_e \bar{\pi}_a}{\psi_a \zeta \left[\sum_{a'} \bar{\pi}_{a'} \tilde{h}_{a'}\right]}\right]^{\frac{1}{\zeta-1}}, \tag{12}$$

where $\bar{\pi}_a = \int_\eta \int_\lambda [1-\lambda^{-1}] \Phi_a(d\lambda, d\eta)$ *is the expected profits of a new product.*

The values in Proposition 1 and the skilled labor wage rate in (10) express the equilibrium solutions in terms of the creative destruction rate δ, which is a function of the product distribution h_a and entry rate x_e. The product distribution is equal to

$$\mathbf{h} = \left[\mathrm{diag}(1+\delta-x_a) - \Gamma'\right]^{-1} x_e \tilde{\mathbf{h}}, \tag{13}$$

where Γ' is the transpose of the transition matrix Γ and $\mathrm{diag}(1+\delta-x_a)$ is a matrix with diagonal elements $1+\delta-x_a$. The product distribution depends on the entry probability $x_e\tilde{h}_a$ scaled by the relative innovation rates $x_a - \delta$ and the relative transition dynamics $\Gamma' - 1$. The entry rate in equilibrium solves

$$x_e = \frac{S}{\psi_e} - \frac{1}{\psi_e} \sum_a \psi_a x_a^\zeta h_a,$$

where note that h_a depends on the entry rate through $\delta = x_e + \sum_a x_a h_a$.

2.3 R&D Misallocation

The growth rate in (9) highlights the main tension in the model. The R&D allocation depends on the expected profitability of innovations $\bar{\pi}_a$ while the contribution to growth of

R&D resources depends on the expected knowledge spillovers of innovations $\ln \bar{\eta}_a$. R&D is misallocated if there exists a feasible R&D allocation that could increase growth (Definition 2). I focus on the growth maximizing R&D allocation as the benchmark—rather than the socially optimal R&D allocation—because this allocation is more relevant to examining the impact on aggregate productivity growth.[12] Additionally, focusing on the growth maximizing R&D allocation allows me to follow the output misallocation literature.[13]

Definition 2. *Define growth* $g(\mathbf{s};\mathbf{h}) = \sum_{a\in\mathcal{A}} \left[h_a x_a + \tilde{h}_a x_e\right] \ln \bar{\eta}_a$ *for skilled labor allocation* $\mathbf{s} = [s_a]_{a\in\mathcal{A}} \cup \{s_e\}$ *and product distribution* $\mathbf{h} = [h_a]_{a\in\mathcal{A}}$*. For a given* $\mathbf{h}$*, R&D is misallocated if there exists* $\check{\mathbf{s}} \geq 0$ *such that* $g(\check{\mathbf{s}};\mathbf{h}) > g(\mathbf{s};\mathbf{h})$ *and* $S \geq \check{s}_e + \sum_{a\in\mathcal{A}} \check{s}_a h_a$*.*

The growth maximizing allocation of skilled labor is

$$s_a^* = \psi_a \left[\frac{\psi_e \ln \bar{\eta}_a}{\zeta \psi_a \left[\sum_{a'} \ln \bar{\eta}_{a'} \tilde{h}_{a'}\right]}\right]^{\frac{\zeta}{\zeta-1}} \tag{14}$$

and $s_e^* = S - \sum_{a\in\mathcal{A}} s_a^* h_a$. The expression replaces the expected private return $\bar{\pi}_a$ from the market allocation in (12) with the expected public return $\ln \bar{\eta}_a$. I define the R&D wedge between the market and growth maximizing allocations as

$$\theta_a = \frac{s_a}{s_a^*} = \left[\frac{\bar{\pi}_a / \sum_{a'} \bar{\pi}_{a'} \tilde{h}_{a'}}{\ln \bar{\eta}_a / \sum_{a'} \ln \bar{\eta}_{a'} \tilde{h}_{a'}}\right]^{\frac{\zeta}{\zeta-1}} = \tau_a \left[\frac{\sum_{a'} \ln \bar{\eta}_{a'} \tilde{h}_{a'}}{\sum_{a'} \bar{\pi}_{a'} \tilde{h}_{a'}}\right]^{\frac{\zeta}{\zeta-1}}. \tag{15}$$

The wedge measures how far a firm is from the allocation of skilled labor that would maximize the growth rate. Values of θ_a above one indicate that a firm is allocated too much R&D resources in the market allocation while values below one indicate a firm is allocated too few R&D resources. Growth can be increased by reallocating resources from firms with $\theta_a > 1$ to firms with $\theta_a < 1$. The value of τ_a captures the firm-specific component of the R&D wedge whereas $[\sum_{a'} \ln \bar{\eta}_{a'} \tilde{h}_{a'} / \sum_{a'} \bar{\pi}_{a'} \tilde{h}_{a'}]^{\zeta/\zeta-1}$ captures the general equilibrium component that depends on the skilled labor wage rate. The wedge τ_a is directly comparable to the output wedges in Hsieh and Klenow (2009).

Proposition 2 shows how the R&D wedge affects the market allocation growth rate relative to the maximum obtainable growth rate.

Proposition 2. *Assume entry is positive and (A) holds, for a given product distribution* h_a*:*

[12] Appendix D.3 examines the Social Planner's allocation, which accounts for the effect of the R&D allocation on output misallocation (χ) and on the future product distribution h_a.

[13] The growth maximizing R&D allocation also does not rely on model-specific assumption on households preferences or the accumulation of product lines. In this regard, the growth maximizing allocation implied by (15) is consistent with a broader class of models.

1. *the maximum obtainable growth rate is equal to:*

$$g^* = \left[\frac{S}{\psi_e}\sum_a \tilde{h}_a \ln \bar{\eta}_a\right]\left[1 + \sum_a h_a \frac{s_a^*}{S}(\zeta - 1)\right]; \tag{16}$$

2. *The market allocation growth rate is equal to:*

$$g = \underbrace{\frac{1 + \sum_a h_a \frac{s_a^*}{S}(\zeta \theta_a^{\frac{1}{\zeta}} - \theta_a)}{1 + \sum_a h_a \frac{s_a^*}{S}(\zeta - 1)}}_{\textit{Allocative Efficiency}} \times \underbrace{g^*}_{\textit{Max. Growth}}; \tag{17}$$

3. *Growth g is falling in the value of $|\theta_a - 1|$ for any type a;*

4. *$g = g^*$ if and only if $\bar{\pi}_a = \nu \ln \bar{\eta}_a$ for all a and some constant $\nu > 0$.*

The market growth rate g in (17) is equal to the allocative efficiency of R&D multiplied by the maximum growth rate g^*. The maximum growth rate depends on the innovation technologies, captured by R&D costs (ψ_a, ψ_e) and average knowledge spillovers $\ln \bar{\eta}_a$. The insight from (17) is that the decline in research productivity (as documented by Bloom et al., 2020) may be driven by either a decline in either innovation technologies or the allocative efficiency of R&D. The proposition also highlights that some output misallocation from markup dispersion may be desirable to improve the allocative efficiency of R&D. Specifically, the fourth part of the proposition states that growth is maximized only when there is markup dispersion from equating expected profitability and knowledge spillovers of innovations.

Growth in (17) highlights the key moments needed to quantify R&D misallocation: (i) the distribution of products h_a; (ii) the expected profitability of innovations $\bar{\pi}_a$; and (iii) the average knowledge spillovers generated by innovations $\ln \bar{\eta}_a$.

3 Empirical Analysis

The model highlights R&D misallocation caused by mismatches between expected private and public returns to innovation. I use matched patent and firm-level financial data to examine the extent of this mismatch empirically. I also construct innovative firm types that are used in the model calibration.

3.1 Data

NBER USPTO Utility Patent Grant Database. The main data source is the patent and citation dataset from the NBER Patent Database (Hall et al., 2001). I restrict to patents issued by US non-governmental organizations, which I refer to as firms. Citations are adjusted using Hall et al. (2001) weights to account for between-period and between-sector differences in the likelihood to be cited. I restrict the sample to innovative firms that issue three or more patent over the entire sample and drop the final year (2006) in the dataset. I winsorize citations and number of patents by firms at the top 1% to limit the influence of outliers.

Compustat. I use the Compustat database to connect patents with firm-level financial data. I drop firms that experience major mergers or acquisitions in any period and firm-year observations with negative sales.[14] I further restrict to firms that have an innovative type constructed using the patent data (discussed in Section 3.3).[15] The resulting sample contains 14,349 firm-year observations by 1,857 unique firms covering 285,941 patents.[16] I winsorize variables at the top and bottom 2% in each year to limit outliers.

Compustat, and the joint patent data, has been widely used to study aggregate trends (some recent related examples include De Loecker et al., 2020; Cavenaile et al., 2020; De Ridder, 2020). Compustat firms cover around 31% of employment and 41% of sales (Asker et al., 2015), firms from a wide breadth of sectors, and over one quarter of total patents in the final dataset. That said, Compustat data is comprised of publicly traded firms raising concerns about representativeness. With this in mind, I focus the use of Compustat data to measures relative firm profitability, where alternative measures are unavailable, and use the broader-coverage patent data to construct other moments. Additionally, I control for granular sector-level fixed effects in the construction of calibration moments to limit the influence of sectoral representation. Other constructed moments (R&D intensity and average profit margins) fall in line with other datasets (e.g., Acemoglu et al., 2018; Akcigit and Kerr, 2018).

3.2 The Empirical R&D Wedge Distribution

As a starting point, I follow Hsieh and Klenow (2009) by measuring the extent of R&D misallocation using dispersion in firm-level R&D wedges. This is equivalent to assuming that each firm f represents a unique type a. While firm-level measures likely capture many idiosyncratic factors (e.g., measurement error, ex-post heterogeneity in innovation outcomes), relative dispersion in R&D wedges is useful for comparing differences in R&D misallocation

[14] Defined as firms that experience a acquisition or merger valued at more than 5% of the firm's total value in any period of the sample. This is done to prevent issues where a firms patent portfolio or profitability reflects acquisitions.

[15] Including the set of non-innovative firms as a fourth type does not substantially change the results.

[16] The final data only includes firms with a recorded firm type (discussed in Section 3.3), which requires that firms patent over five-year partitions of the data. For example, a firm with data from 1970 to 1990 that applies for two patents in 1980 and 1981 counts as five observations.

between samples. Given this, I divide the sample into an early (1976-1990) and late (1991-2005) period to assess changes in R&D misallocation over time.

The R&D wedge in (15) depends on both a firm-specific component τ_a and a general equilibrium component. The latter can be ignored since it does not affect dispersion in R&D wedges. The firm-specific component of the R&D wedge is

$$\tau_{f,t} = \left[\frac{\text{Expected Profitability}_{f,t}}{\text{Expected Knowledge Spillovers}_{f,t}}\right]^{\frac{\zeta}{\zeta-1}} \approx \left[\frac{\ln \text{Sales-per-Employee}_{f,t}}{\text{Citations-per-Patent}_{f,t}}\right]^{\frac{\zeta}{\zeta-1}}. \tag{18}$$

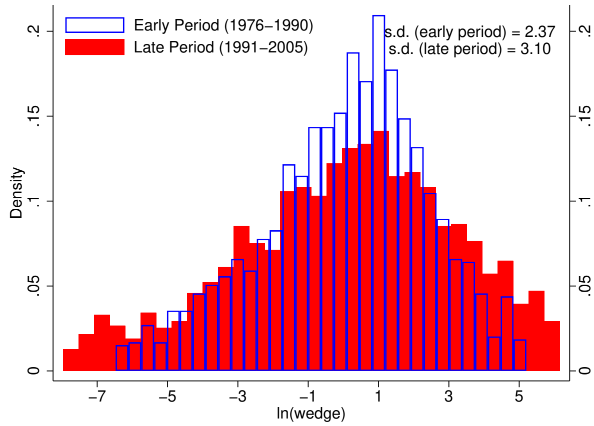
In the model, knowledge spillovers $\ln \eta$ determine the contribution of firm innovations to long-run aggregate productivity growth. Patents are ideal to measure knowledge spillovers because of their empirical relationship with long-run productivity growth (see, for example, Berkes et al., 2022). I measure knowledge spillovers in the data using average firm-level citations-per-patent. Citations are a natural candidate for knowledge spillovers because they capture the usefulness of the innovation to future innovations. In the model, the profitability of a good j can be approximated as $\pi_j = (1 - \lambda_j^{-1}) \approx \ln \lambda_j$ where λ_j is proportional to sales-per-employee, which I use as the empirical measure of firm profitability. Each variable (sales, employees, citations, patents) is scaled by its mean value each year to remove time trends as both patenting and profitability grow over this period (Kortum and Lerner, 1999; De Loecker et al., 2020).[17] I also average firm-level variables over consecutive five-year periods (e.g., 1976-1980, 1981-1985) to reduce measurement error and trim the top and bottom 2% of firms based on measured wedges $\hat{\tau}_{f,t}$ to reduce the influence of outliers.[18] Figure 1 plots the wedge distribution for the early and late periods.

There is a substantial mismatch between the firm-level private and public returns to innovation indicating that firms that produce more highly cited patents are not necessarily more profitable. On its own, the dispersion in the R&D wedges indicates that growth could be increased by aligning the private and public returns. As mentioned, the dispersion in R&D wedges could be driven, in part, by measurement error (e.g., high knowledge spillover patents receiving low citations) or other idiosyncratic factors. The trend in the dispersion of the R&D wedges over time does not suffer from this problem since these underlying factors are present in both periods. The dispersion in R&D wedges has increased substantially between the early and late periods. Appendix B.4 shows that the trend is robust to alternative measures of profitability and constructing wedges as citations divided by total R&D expenditure to follow

[17]Like with output misallocation, the dispersion in variables leads to R&D misallocation. Removing time trends avoids including between-year dispersion in variables in the measure of R&D misallocation.

[18]Dispersion of wedges is also increasing over time if variables are averaged over the full early and late periods or if dispersion in wedges is examined at the five-year sub-period level.

Figure 1: R&D Wedge Distribution by Period



Notes: The figure plots the histogram of the firm-level wedge in (18) over the early (1976-1990) and late (1991-2005) periods. The top and bottom two percent of observations, based on the firm-level wedge, are dropped in each period. The mean of log wedges is normalized to zero in each period.

Hsieh and Klenow (2009).[19]

3.3 Innovator Types

The firm-level R&D wedges in Figure 1 point to large and rising R&D misallocation but suffer from measurement issues. I use the model to address this issue by building similar measurement issues into the model. I estimate the model using firm types $a \in \mathcal{A}$ constructed using firms with similar innovation characteristics. The remainder of the section discusses the construction of firm types and shows that these firm types capture similar dynamics as the firm-level R&D wedges. In the next section, I extend the model to include patenting such that the calibrated model matches the key features of the data, including ex-post heterogeneity in innovation outcomes and the possibility that firm types are misclassified.

I construct firm types using all US firms in the patent database that patent at least three times over the sample. Firm-level variables are averaged over consecutive five-year periods (e.g., 1976-1980, 1981-1985) to provide sufficient time to observe patenting characteristics while still allowing for entry, exit, and transitions between firm types. The final sample includes 83,829 firm-period observations that account for 732,141 patents.

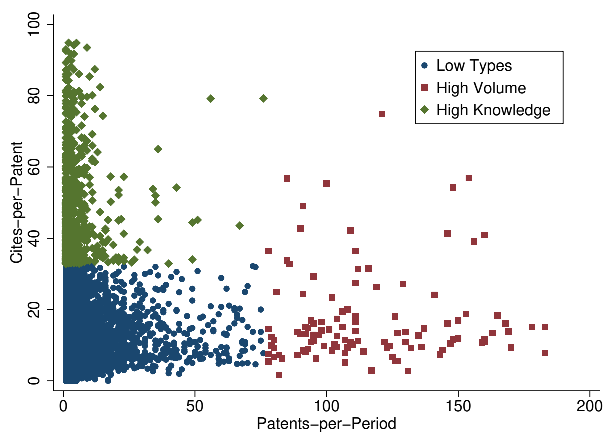
Firms are grouped using a k-means clustering algorithm based on two patenting character-

[19] All three R&D wedges are also positively correlated suggesting the same underlying sources of R&D misallocation. I focus on (18), rather than these alternative measures, because it directly corresponds to the channel described in the model and R&D expenditures appears to be poorly measured in the data, with many firms reporting zero R&D expenditure despite high patenting frequently.

istics.[20] The first characteristic is the firm's average citations-per-patent since this captures knowledge spillovers η in the model. The second characteristic is patents-per-period since innovation frequency is closely related to the costs and private returns of innovations, as in (12). Additionally, the product of the two variables is citation-weighted patents, a common measure of total innovation output (Hall et al., 2001). In this respect, focusing on these two characteristics can be thought of as delving into the implicit heterogeneity within this more traditional measure of innovation output. The analysis implicitly assumes that the two innovation characteristics are different, such that a firm that produces many low-citation patents and a firm that produces few high-citation patents are different.

Figure 2 plots the innovation characteristics for a 10% sample of firm-period observations. Highly innovative firms tend to excel in only one of the two dimensions. That is, firms either over-perform in citations-per-patent firms or in patents-per-period. In contrast, the literature generally treats innovation as unidimensional focusing on quality-adjusted measures of patents-per-period. Returning to the example, the typical approach would imply that a firm that produces many low-citation patents is fundamentally the same as a firm that produces a few high-citation patents. Figure 2 also reports the constructed firm types.

Figure 2: Firm Types



Notes: The figure plots a 10% sample of firm-period observations groups by type. Firm types are constructed using a k-means algorithm with three total types.

I refer to the three types as low types (for low innovation frequency, low knowledge spillovers), high volume, and high knowledge firms. Table 1 reports summary statistics by firm type. High volume and high knowledge firms are better innovators with both types outperforming low types in at least one dimension. Additionally, both high types perform better

[20] Define the vector of innovation characteristic as $\mathbf{z}_f$ for a firm-period observation. The k-means clustering algorithm sets groups $\mathcal{F}_{i=1,2,\dots,k}$ to minimize $\sum_{i=1}^{k} \sum_{f \in \mathcal{F}_i} ||\mathbf{z}_f - \mu_i||^2$ where μ_i is the average of $\mathbf{z}_f$ for $\mathcal{F}_i$.

in the frequency that patents are above the 90th percentile of citations in the current year (freq. >90p) and the frequency that patents receive zero citations (freq. 0 cites). The polarization of the high volume and high knowledge types shows a contrast with unidimensional measures of innovative ability.[21] Appendix B.1 shows that allowing for more granular types results in types that lie between the low type and high volume or high knowledge firms.[22]

Table 1: Descriptive Statistics on Firm Types

	obs	cites/pat.	pat/period	freq. >90p	freq. 0 cites
Low Type	74187.0	10.5	5.5	4.7	14.1
High Vol.	1923.0	15.4	147.6	9.5	13.8
High Know.	7719.0	55.2	5.5	55.6	0.9

A concern is that the types reflect ex-post, rather than ex-ante, heterogeneity in patenting outcomes. I correct for this in the model calibration by allowing similar sources of ex-post heterogeneity and firm type misclassification. Appendix C.3 shows that the model calibrated to a single firm type is unable to replicate the type distribution with ex-post heterogeneity alone. The single-type model generates only around 1% of high volume innovations and under one-third of high knowledge innovations in Table 1.

3.4 Private and Public Returns by Firm Type

Figure 1 shows that there is substantial dispersion in R&D wedges and that this dispersion has increased over time. I use the linked Compustat data to examine whether the constructed firm types capture these trends. I regress firm-level measures of private and public returns to innovation on firm type $a \in \{\text{low type}, \text{high vol}, \text{high know}\}$ indicators. Table 2 reports the results. I divide the results between the early (1976-1990) and late (1990-2005) periods to examine how between-type differences change over time.

For the public returns, the relative ranking of firms is the same as before. High knowledge firms produce the most highly cited patents, followed by high volume firms and then low type firms. This is almost mechanical given the construction of the firm types, but it is reassuring that the ranking holds after controlling for sector and year fixed effects. The results also show that the relative differences in citations remain stable over time.

For the private returns, high volume firms are the most profitable while high knowledge firms are the least profitable, at least in the late period. This ranking is consistent with

[21] Appendix B.2 compares the three firm types with types constructed using only citations-per-period as an aggregate measure of knowledge output. The analysis shows that the group of high knowledge firms tends to be missed while the group of high volume firms tends to be classified as high types in this new grouping.

[22] For example, moving to five types adds mid volume and mid knowledge types to the three baseline types.

Table 2: Private and Public Returns by Firm Type

	Early Period (1976-1990)		Late Period (1991-2005)	
	(1) ln(sales/emp)	(2) cites/pat	(3) ln(sales/emp)	(4) cites/pat
High Vol.	0.0257	0.234**	0.413***	0.358***
	(0.0496)	(0.0959)	(0.0856)	(0.0854)
High Know.	-0.0198	1.116***	-0.0869	1.167***
	(0.0496)	(0.0560)	(0.0648)	(0.0481)
Year FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Observations	6883	4440	7459	5228

*Notes: * p<0.10, ** p<0.05, *** p<0.01. Robust standard errors clustered at the firm level are reported in parentheses. All regressions include year and sector (four-digit SIC) fixed effects. Columns (2) and (4) are estimated using PPML. Columns (2) and (4) have fewer observations than (1) and (3) because not all firm-year observations have patent applications.*

the idea that higher expected profitability drives the innovation frequency (as measured by patents-per-period) of high volume firms. Unlike with public returns, there is a sharp widening of cross-type differences between the periods, with private returns being more egalitarian in the early period. The results are robust to alternative measures of profitability (Appendix B.3.3) and to private and public returns based on the stock market valuations of patents constructed by Kogan et al. (2015) (Appendix B.3.4).

High volume firms are also substantially larger than other firms (Appendix B.3.2). The rise in profitability of high volume firms may be related to other trends about large firms, such as rising market power (e.g., Cavenaile et al., 2020), patent protection (e.g., Akcigit and Ates, 2019) and intangible investment (e.g., De Ridder, 2020).

4 Calibration and Measurement

The empirical evidence points to substantial and rising R&D misallocation. I calibrate the model to the late period (1991-2005) where R&D misallocation is most severe. Appendix D.1 compares the baseline model with the model calibrated to the early period.

4.1 Patents and Citations

Innovation in the data is described by patent outcomes. I extend the model to explicitly describe the mapping from innovations and knowledge spillovers to patenting and citations. The extension allows me to construct model moments directly comparable to data moments.

Firms produce patents at a proportional rate to innovations, such that a firm that innovates at rate x_a produces patents at Poisson rate Λx_a for some constant Λ common to all firm types. In general, more innovative firms patent more, but an innovation does not equate to an exact number of patents. Intuitively, this captures that an innovation could be, for example, a new drug with a single patent or a new car with three patents on the GPS, engine, and electronic control system.

Patent citations are determined by

$$citations = \begin{cases} 0 & \text{with probability } \alpha \\ \tilde{c} & \text{with probability } 1-\alpha \end{cases}, \tag{19}$$

where $\tilde{c} \sim \text{lognormal}(\mu_a, \sigma^2)$. The probability α allows for patents to receive zero citations, as in the data, while the values of μ_a and σ determine the distribution of citations conditional on receiving positive citations. Knowledge spillovers $\ln \bar{\eta}_a$ are proportional to average patent citations, such that the expected knowledge spillovers of an innovation are $\ln \bar{\eta}_a = \nu(1-\alpha) * \exp(\mu_a + \sigma^2/2)$ for constant $\nu > 0$. On average, higher knowledge spillover firms receive more citations on patents, but this may not be the case for individual patents.

The randomness in patent and citation outcomes allows for ex-post heterogeneity in firm outcomes. For example, a low type firm may, through chance, produce many patents and be classified as a high volume firm or produce several highly-cited patents and be classified as a high knowledge firm. This is important for the calibration because it implies that the model captures the same ex-post heterogeneity present in the data.

4.2 Calibration Strategy

The calibration determines the parameters $\{\rho, \zeta, S\}$; the type-specific innovation costs ψ_a; innovation distribution $\Phi_a(\lambda, \eta)$; transition matrix Γ; entry distribution $\tilde{h}_a$; and the patenting parameters $\{\alpha, \mu_a, \sigma, \nu, \Lambda\}$. The types $a \in \mathcal{A}$ are set to match the low type firms ($a = L$), high volume firms ($a = V$) and high knowledge firms ($a = K$) from Section 3.

Externally calibrated parameters. The discount rate is set to $\rho = 0.02$ which implies an interest rate of close to 4% along with the calibrated growth rate. The curvature of the cost function is set to $\zeta = 2$ consistent with the value found by Acemoglu et al. (2018). The mass of skilled labor is set to $S = 0.17$ consistent with the fraction of workers employed in a research capacity in the United States reported by Acemoglu et al. (2018).

To reduce the computational burden of the calibration, I set two sets of moments based on closely related empirical counterparts. First, the number of entry products is set to two for both the low type and high knowledge firms and to 30 for the high volume firms, which is consistent with relative firm sizes (see Table 5). As discussed in Section 2, the number of entry products does not affect the equilibrium product distribution or R&D misallocation. Second, the transition matrix is taken to be the empirical transition matrix,[23] such that

$$\Gamma = \begin{pmatrix} 98.3 & 0.3 & 1.4 \\ 4.5 & 95.2 & 0.3 \\ 17.3 & 0.7 & 82.0 \end{pmatrix}. \tag{20}$$

Appendix C.2 shows that the model-implied transition matrix between observed types is similar to the empirical transition matrix suggesting that this assumption reasonably approximates the data. The model underestimates the persistence of high knowledge firms, which lowers the potential gains from R&D reallocation implying more conservative results. The transition matrix in (20) indicates that high volume and high knowledge firms transition to low type firms over time, consistent with the decline in research productivity of older firms and over time documented by Acemoglu et al. (2018) and Bloom et al. (2020).[24]

The probability that patents receives zero citations is set to $\alpha = 12.8\%$ and the standard deviation of citations is set to $\sigma = 0.93$ based on Table 1. Firms draw ideas (λ, η) from a type-specific distribution where knowledge spillovers η depend on citations as described previously. Private returns are drawn from a degenerate distribution with value $\bar{\lambda}_a$, which is without loss of generality since firms are risk neutral.

Internally calibrated parameters. The remaining 14 parameters $\{\psi_a, \bar{\lambda}_a, \tilde{h}_a, \psi_e, \mu_a, \nu, \Lambda\}$ are jointly chosen to match the moments in Table 3. Ten of the moments describe type-specific outcomes that discipline the extent of cross-type differences, including R&D misallocation. The remaining four moments discipline aggregate characteristics of the economy and are calculated analytically. Given that ν only affects the aggregate growth rate, it is chosen after the other parameters to match this moment.

[23]Specifically, Γ^5 equals the five-year transitions across types found in the data.

[24]High volume firms tend to be older and high knowledge firms tend to be younger than low type firms. However, I do not find evidence of a lifecycle that accounts for transitions from high knowledge to high volume firms over time. A clear indication of this is that high volume firms are more likely to have previously been low type firms than high knowledge firms.

Table 3: Model Moments and Data Targets

Moment	Data	Model
Firm Share (%)	(88.5 , 2.30 , 9.20)	(88.4 , 2.29 , 9.31)
Citations-per-Patent	(10.5 , 15.4 , 55.2)	(10.6 , 15.3 , 56.4)
Patent Count	(5.5 , 147.6 , 5.5)	(5.5 , 147.1 , 5.4)
Rel log(Sales-per-Emp)	(0.00 , 0.41 , −0.09)	(0.00 , 0.38 , −0.11)
R&D Intensity (%)	6.9	7.2
Avg Profit Margin (%)	15.3	15.4
Entry Rate (%)	7.4	7.4
Growth Rate (%)	1.63	1.63

Notes: Moments are ordered corresponding to low type, high volume, and high knowledge firm types. Type-specific moments are calculated using recorded types from data generated from a simulation of 30,000 firms (described in Section 4.2).

Simulation data. Type-specific model moments in Table 3 are based on simulation data of 30,000 firms over a five-year period where firms are weighted based on the equilibrium share of firm types m_a (defined in Appendix A.1).[25] The first-year distribution of firms is initialized using the BGP equilibrium distribution of firms across types a and the number of product lines n.[26] Firms accumulate product lines at rate $x_a n$, lose products at rate δn, and transition between types based on the transition matrix Γ. I calculate model moments using the observed types $\hat{a}$ of firms, which may differ from true types a, based on patent outcomes over the simulation. Firms that patent more than 76 times are recorded as high volume firms while firms that receives at least an average of 32.8 citations-per-patent are recorded as high knowledge firms, where the cutoffs are the lower bound of high knowledge and high volume firms in the data.

The simulation allows moments to be calculated consistent with the data. Importantly, the simulation allows for misclassification such that observed firms types, denoted by $\hat{a}$, may differ from true firm types, denoted by a. Observed types are based on firm outcomes (and cutoffs) while true types determine firm parameters. Allowing for the same sources of misclassification in the simulated data as in the data corrects for systematic biases in the parameters. For example, some low type firms may be highly cited and, incorrectly, recorded

[25]Specifically, 10,000 firms of each type are simulated and moments are calculated assuming that a firm has population weight m_a based on initial firm types. This reduces the computational burden by allowing for relatively more high volume and high knowledge firms to be simulated and ensures that moments are calculated using the correct population weight.

[26]To reduce computational cost, I assume that goods produced by firms in the initial year of the simulation have private return λ_a for the firm's initial type a. This implies an upward bias on the estimate of λ_L given that some low type firms were previously high volume firms. I also assume that all firms produce at least one patent in the first period of the simulation.

as high knowledge firms. If misclassification is not taken into account then this example implies larger parameter differences between low type and high knowledge firms since the model would incorrectly attribute the observed differences entirely to underlying parameters. Building misclassification into the model allows for some of the observed differences between types to be attributed to misclassification. While all types of misclassification occur in the simulation, the most common is for high volume and high knowledge firms to be recorded as low type firms.[27]

4.3 Calibration Moments

I discuss the construction of the model and data moments. While the parameters are highly interdependent and chosen jointly, I also discuss closely related parameters to each moment for intuition. Appendix C.1 provides additional details on the identification and the sensitivity of the model moments with respect to the calibrated parameters.

Firm Share. The model moment is the share of firms with observed type $\hat{a}$, which is based on firm patenting and citation outcomes. Table 1 reports the data moment. The moment is related to the the entry probabilities $\tilde{h}_a$ and the survival rate of firms, which depends on R&D costs ψ_a and returns $\bar{\lambda}_a$ through the innovation rates x_a and creative destruction rate.

Citations-per-Patent. For each firm, citations-per-patent is constructed as the total number of citations received on all of the firm's patents divided by the firm's patent count. The model moment is the average value across observed type $\hat{a}$ firms. Table 1 reports the data moment. The moment is related to the type-specific citation mean μ_a and, consequently, the type-specific knowledge spillovers $\bar{\eta}_a$.

Patent Count. The model moment is constructed as the average total patent count of each observed type $\hat{a}$ firm. The data moment is from Table 1. The moment is closely related to the patent-innovation ratio Λ and the innovation rate x_a, which depends on innovation costs ψ_a and the private returns to innovation $\bar{\lambda}_a$.

Rel log(Sales-per-Emp). For each firm, sales-per-employee is calculated as the total sales divided by total employee-years hired by the firm over the simulation. The model moment is the average log value for observed type $\hat{a}$ firms subtract the value for observed low type

[27] Around 7% of firms are misclassified in the data with around equal mass of high volume and high knowledge firms being classified as low type firms.

firms. The data moment is from Table 2. The value is related to the type-specific private return $\bar{\lambda}_a$, since $p_j y_j/\ell_j \propto \lambda_j$ for good j.

R&D Intensity. The model moment is the average value of R&D spending to sales across firms, $\omega[\sum_a s_a m_a + s_e]$ where m_a is the type a firm share (defined in Appendix A.1). The data moment is the median value of R&D expenditures to sale for firms that report R&D spending less than total sales. The moment is related to the equilibrium wage rate relative to GDP ω, which depends on the entry cost ψ_e, the entry probabilities $\tilde{h}_a$, and the profitability of each type through $\bar{\lambda}_a$.

Avg Profit Margin. The model moment is the average profits across firms, $\sum_a (1-\bar{\lambda}^{-1}) m_a$. The data moment is the average operating income before depreciation divided by sales for firms with positive values. The moment is related to the private returns $\bar{\lambda}_a$ of innovations.

Entry Rate. The model moment is the entry rate divided by the total innovation rate, $\sum_a x_{e,a}(\bar{n}_a)/\delta$. The data moment is the total first patents divided by total patents. The moment is related to the cost of entry ψ_e but also other factors that determine the allocations between entry and incumbent innovation, such as innovation costs ψ_a.

Aggregate Growth. The moment is constructed using (9) and is set to target the productivity growth rate of 1.63% reported by Aghion et al. (2019a) in the final period of their data.[28] The growth rate depends on both the average knowledge spillovers $\bar{\eta}_a$ and R&D allocations across firm types s_a and entry s_e. The knowledge spillover-citation ratio ν is set to match the growth rate after other parameters are determined.

4.4 Parameter Values

Table 4 reports the calibrated model parameters. The estimated parameters show substantial gaps in the underlying characteristics of firm types. As with the empirical evidence, there is substantial variation across firm types in the profitability $(1-\bar{\lambda}_a^{-1})$ and knowledge spillovers $(\ln \bar{\eta}_a)$ of innovations. Relative to low type firms, innovations by high volume firms are around 50% more profitable while innovations by high knowledge firms are around 25% less profitable. In contrast, innovations by high volume firms have around 60% higher knowledge spillovers than low type firms while high knowledge firms have almost four times higher

[28] This ignores the growth acceleration that occurs in the early 1990s, which is generally associated with the computerization of the economy. See, for example, Aghion et al. (2019b), De Ridder (2020) or Ayerst (2021) for models that incorporate features of computerization into growth models.

Table 4: Calibrated Parameters

Parameter		Value
Externally Calibrated Parameters		
Discount Rate	ρ	0.02
Cost Function Curvature	ζ	2
Number of Researchers	S	0.17
Prob of Zero Citation (%)	α	12.8
Standard Dev of Citations	σ	0.93
Internally Calibrated Parameters		
Entry Cost	ψ_e	1.91
Innovation Cost	ψ_a	(11.4 , 8.3 , 5.5)
Entry Probability	$\tilde{h}_a$	(0.52 , 0.24 , 0.23)
Avg Private Return	$\bar{\lambda}_a$	(1.18 , 1.29 , 1.14)
Avg Citation	μ_a	(2.01 , 2.50 , 3.35)
Avg Public Return	$\bar{\eta}_a$	(1.08 , 1.13 , 1.32)
Patent-Innovation Ratio	Λ	4.50
Know Spillover-Cite Ratio	ν	0.0073

Notes: Parameters are ordered corresponding to low type, high volume, and high knowledge firms. The transition matrix Γ is reported in (20).

knowledge spillovers. The gaps in the model parameters are moderately larger than the empirical gaps because misclassification results in understated cross-type differences in the data. For example, some high volume firms are classified as low type firms raising the average profitability of low types leading to the underlying differences in profitability needing to be higher in order to explain the observed empirical gaps.

Despite innovating more frequently, the innovation cost of high volume firms is higher than for high knowledge firms. This is explained by the relatively high profitability of high volume firms justifying large investments into R&D despite the relative innovation.

4.5 Goodness-of-Fit and Other Model Moments

The calibrated model replicates the main data characteristics. Table 5 presents additional moments not directly targeted in the calibration. Panel A reports moments with direct data counterparts as a check of the model's goodness-of-fit. As with the targeted moments, the model moments are based on observed types constructed using simulation data. Panel B reports moments without direct data counterparts that describe cross-type differences, and are based on the true firm types.

The model generates a similar patent distribution across observed types and similar dif-

Table 5: Other Model Moments

A: Comparison with Untargeted Data Moments			
Moment		Data	Model
Patent Share		(55.5 , 37.6 , 6.90)	(55.6 , 38.6 , 5.78)
Cites / Emp		(0.00 , 1.26 , 1.50)	(0.00 , 0.75 , 1.64)
Rel log(Sales)		(0.00 , 2.80 , -0.22)	(0.00 , 2.89 , -0.17)
Rel log(Emp)		(0.00 , 2.35 , -0.12)	(0.00 , 2.82 , -0.15)
B: Additional Model Generated Moments			
Moment			Model
Firm Share (%)	m_a		(81.29 , 5.87 , 12.84)
Product Distribution (%)	h_a		(57.43 , 34.43 , 8.14)
Innovation Rate (%)	x_a		(7.88 , 15.86 , 12.72)
Skilled Labour	s_a		(0.07 , 0.21 , 0.09)
Fixed Cost / Value	$\omega B_a / V_a(\bar{\pi}_a, 1)$		(0.04 , 0.08 , 0.06)

Notes: Moments are ordered corresponding to low type, high volume, and high knowledge firms. Model moments in Panel A are based on the simulation described in Section 4.2. Model moments in Panel B are based on the equilibrium expressions for true firm types described in Section 2.

ferences in citations-per-employee. The patent share is important for R&D misallocation because it is closely related to the product distribution, which is a key determinant of R&D misallocation. Citations-per-employee provides another check on the innovation characteristics of firm types between the model and data. The model also replicates the relative firm sizes, which is a check on the product accumulation of firms.

Panel B summarizes moments for the true firm types that are not directly observed in the data. Both high volume and high knowledge firms innovate more frequently than low type firms. High knowledge firms innovate only moderately less frequently than high volume firms, implying that most of the observed differences in patenting frequency between firm types is due to firm size. Low types produce the majority of products lines, despite innovating less frequently, because of the tendency for firms to transition to low types over time. The final moment shows that the assumed fixed costs B_a account for a relatively small share of firm value, such that Assumption (A) has little impact on the results.

4.6 Research Productivity

The gains from R&D reallocation depend on both the innovation technologies and the distribution of R&D wedges. I conclude this section by discussing the parameter implications for innovation technologies. The aggregate growth contribution of each firm type is

$$g = \underbrace{(x_L h_L + x_e \tilde{h}_L) \ln(\bar{\eta}_L)}_{=26.6\%} + \underbrace{(x_V h_V + x_e \tilde{h}_V) \ln(\bar{\eta}_V)}_{=44.9\%} + \underbrace{(x_K h_K + x_e \tilde{h}_K) \ln(\bar{\eta}_K)}_{=28.4\%}.$$

Despite accounting for only 6% and 13% of firms, high volume and high knowledge firms account for 45% and 28% of growth. The relatively large contributions are explained by: (i) the large product share of high volume firms (around one third of product lines) increasing their R&D scale; (ii) the higher frequency of innovations by high volume and high knowledge firms compared with low type firms (Table 5); and (iii) the relatively higher knowledge spillovers $\ln \bar{\eta}_a$ generated by high volume and high knowledge firm innovations.

The relative contributions does not imply that reallocating from low-type to high-type firms will necessarily increase growth since the relative marginal research productivity of firms may differ. High volume firms innovate more frequently, in part, because they invest relatively more in R&D (Table 5). Table 6 reports measures of the marginal rate of expected knowledge spillovers and profits.[29] The values correspond to the ability of the firms to generate growth and the ability of the firm to generate profits.

Table 6: Research Productivity by Firm Type

	Low Type	High Vol	High Know
Knowledge Spillovers / R&D	8.2	9.1	40.1
Profitability / R&D	17.1	17.1	17.1

Notes: Knowledge Spillovers / R&D is defined as expected knowledge spillovers $x_a \ln(\bar{\eta}_a)$ divided by R&D inputs s_a. Profitability / R&D is defined as expected profits $x_a \bar{\pi}_a$ divided by R&D inputs s_a.

Firms invest to the point where expected marginal profits, not expected marginal knowledge spillovers, are equal to marginal costs. R&D becomes misallocated because knowledge spillovers could be increased by reallocating R&D resources to high knowledge firms that have comparatively high marginal expected knowledge spillovers. Both low type and high volume firms produce around one fifth of the knowledge spillovers of high knowledge firms. The next section explores the growth gains from reallocating R&D as implied by Table 6.

The table highlights the novel feature of the model that is not present in other model of innovative firms (e.g., Klette and Kortum, 2004; Lentz and Mortensen, 2008; Akcigit and Kerr, 2018). When private and public returns are governed by the same parameters (e.g., when $\eta_a = \lambda_a$), the scope for R&D misallocation is limited since firms invest to the point

[29] The equation of the average knowledge spillovers to R&D with the marginal expected knowledge spillovers follows from $x_a \ln \bar{\eta}_a / s_a = (s_a/\psi_a)^{1/\zeta - 1} \ln \bar{\eta}_a$, which is equal to the marginal rate of knowledge spillovers from increasing skill labor s_a. Profits follows a similar argument.

where marginal knowledge spillovers are (approximately) equal, even if research capabilities (e.g., $\psi_a, \bar{\lambda}_a$) differ.[30] Acemoglu et al. (2018) examine differences in average research productivities driven by differences in firm fixed costs that are paid in R&D resources. In this context, it is optimal for some high fixed cost firms to exit to free resources that could be used for variable R&D costs.

5 R&D Misallocation

The previous sections highlight substantial heterogeneity in public and private returns to innovation across firms as well as large cross-type gaps in measured research productivity. I use the calibrated model to quantify the growth cost of R&D misallocation and how the cost has changed over time.

5.1 R&D Misallocation Across Firms

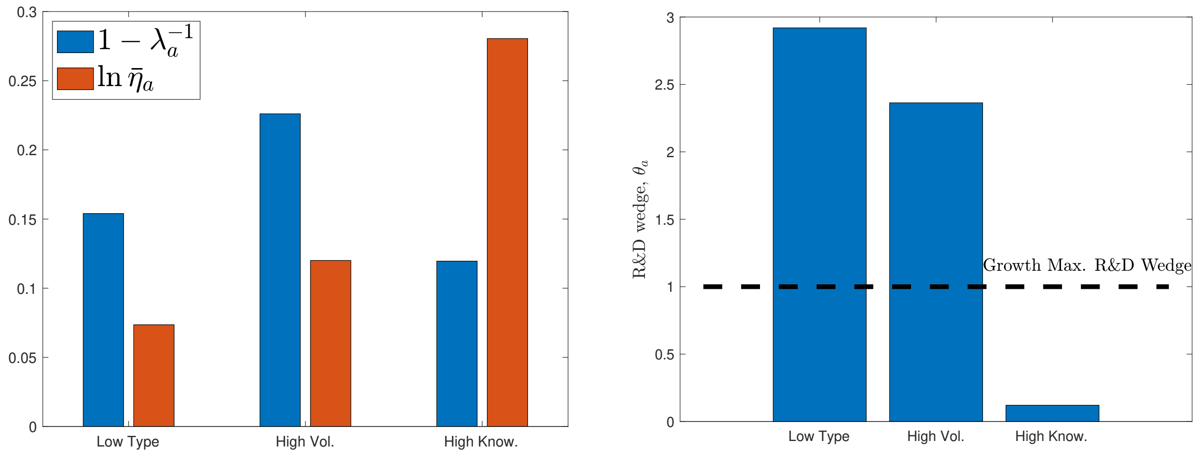
I start by quantifying the cost of R&D misallocation in the late period (1991-2005). Figure 3 reports the R&D wedges in (15) by firm type that describe the relative gap between the market and growth maximizing R&D allocations. For example, an R&D wedge of $\theta_L = 2.9$ indicates that low type firms are allocated 2.9 times as much skilled labor as in the growth maximizing allocation. R&D wedges are independent of the number of product lines firms produce, such that a low type firm invests 2.9 times more than in the growth maximizing allocation regardless of whether it produces one or ten product lines.

The wedges indicate that both the low type and high volume firms invest substantially more, by factors of 2.9 and 2.4, than in the growth maximizing allocation. This is driven by the comparatively small public returns in the case of low type firms and high private returns in the case of high volume firms. In contrast, high knowledge firms invest substantially less than in the growth maximizing allocation, only around one sixth, due to both high public returns and low private returns. Given these differences, the gains from reallocation are driven by reallocating R&D from low and high volume firms to high knowledge firms.

Gains from R&D reallocation. The growth maximizing allocation is achieved when expected profitability is proportional to expected knowledge spillovers, $\bar{\pi}_a^{CF} \propto \ln \bar{\eta}_a$, eliminating the R&D wedge (i.e., $\theta_a = 1$ for all a). Table 7 summarizes output, growth, and

[30] This is only approximately true in most models since private and public returns are usually slightly different functions of the same parameter. For example, in the baseline model setting $\bar{\eta}_a = \bar{\lambda}_a$ does not exactly equate private returns $1 - \bar{\lambda}_a^{-1}$ and public returns $\ln \bar{\eta}_a$. This parameterization of the model would imply small R&D misallocation and be unable to replicate the empirical evidence.

Figure 3: R&D Wedges



(a) Private and Public Returns

(b) R&D Wedges

welfare (reported as a consumption equivalent defined in Appendix A) in two counterfactual experiments. The counterfactual economies correspond to the product distribution h_a being held fixed at the benchmark economy distribution (Fixed) or being determined endogenously based on the new R&D allocations (Flexible). The fixed experiment is equivalent to the short-run growth gains from adjusting R&D allocations and the flexible experiment is the long-run growth gains that are achieved as product lines become reallocated. Appendix D.4 reports the transition dynamics between the fixed and flexible economies. The fixed experiment is also useful as a lower bound on the growth gains from R&D reallocation because it does not require assumptions on the accumulation of product lines.

Table 7: Gains from Reallocation

		Benchmark	Counterfactual	
			Fixed	Flexible
Output Misallocation	χ	0.9987	0.9945	0.9950
Growth Rate (%)	g	1.63	1.91	3.06
Chg in Growth (%)	$g^{CF}/g-1$	-	17.70	88.4
Chg in Welfare (%)	ξ	-	14.99	104.4

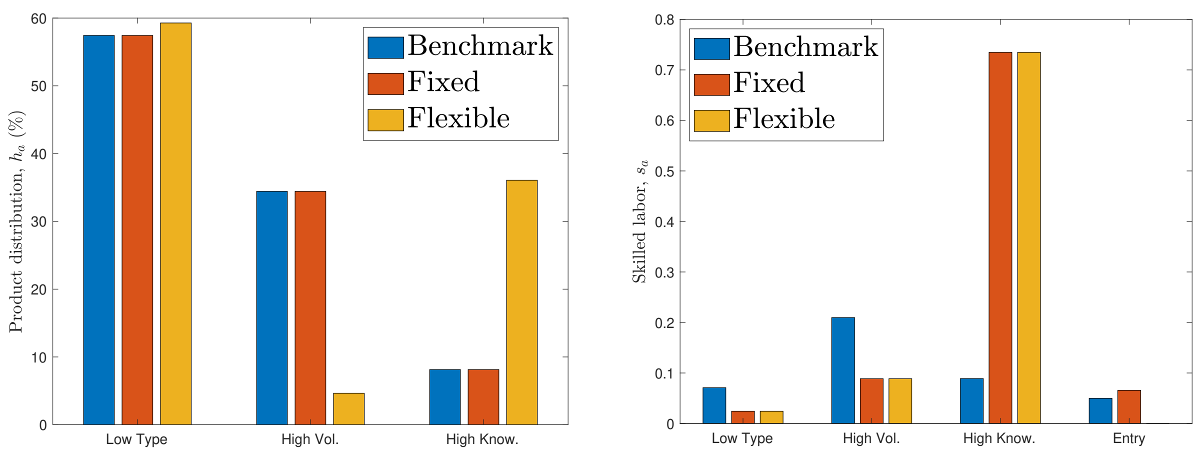
Notes: Benchmark refer to the calibrated equilibrium discussed in Section 4. In the counterfactual equilibrium the private returns to R&D are set such that $\bar{\pi}_a^{CF} = \ln \eta_a$. Output misallocation and the growth rate are given by (6) and (9). Welfare is calculated as described in Appendix A. The welfare calculation in fixed equilibria assume that the distribution of firm types is fixed in all future periods at the benchmark equilibrium level.

In both experiments, R&D misallocation causes a substantial loss in welfare and growth. In the fixed counterfactual economy, the growth rate increases by 18% (or 28 basis points) to

1.91%. This is slightly offset by output declining by around 0.4% relative to the benchmark economy. The net impact is that welfare increases by 15% relative to the benchmark economy. Figure 4 shows that the increase in growth is almost entirely driven by a reallocation from low type and high volume firms to high knowledge firms.

Allowing the firm and product distributions to adjust to the new R&D allocations increases the growth rate to just under double the benchmark economy. The increase in growth is driven by an accumulation of product lines operated by high knowledge firms from under 10% in the benchmark economy to over 30% in the flexible economy. Product accumulation magnifies the gains of R&D misallocation because it allows high research productivity firms to increase their R&D scale. The mechanism is similar to Bento and Restuccia (2017) and Ayerst (2020) where misallocation causes productive firms to invest less in productivity improvements amplifying the costs of misallocation.

Figure 4: Product Allocation and Skilled Labor by R&D Activity



(a) Product Allocation h_a (b) Skilled Labor by R&D s_a

A striking feature in the benchmark and both counterfactual economies is that low types operate the majority of product lines (Figure 4). This is mechanically driven by the transition of high volume and high knowledge types to low types over time. The transition probability also limits the potential gains from reallocation compared with a hypothetical economy where high knowledge firms operate all product lines. Figure 4 also shows that the flexible experiment is a corner solution that allocates zero resources to entry.

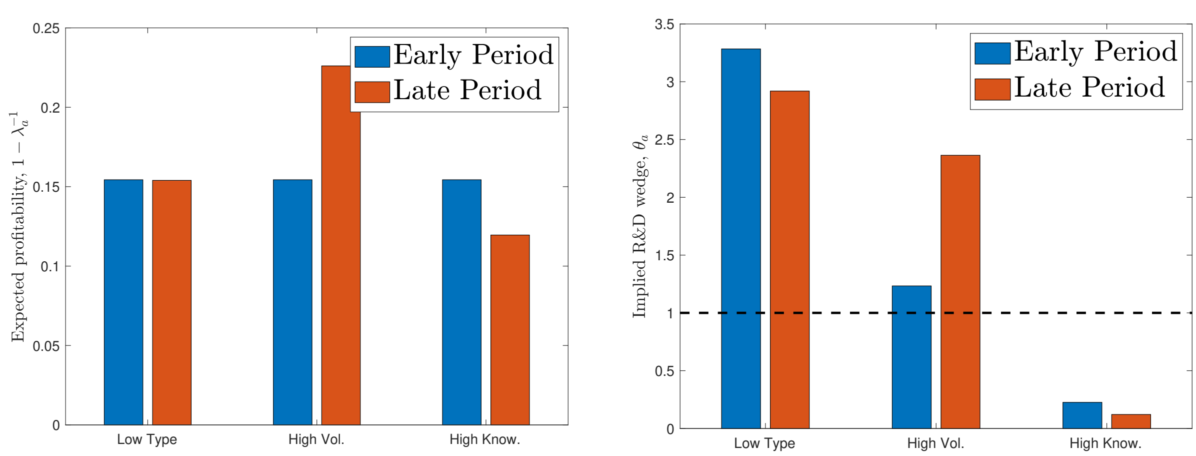
The analysis shows substantial increases to the growth rate from reallocating R&D across firms. A common concern in the output misallocation literature is that the theoretical gains from reallocation (to an undistorted economy) may overestimate the practical gains from reallocation since all economies have some degree of measured firm-level distortions. For example, measured distortions may arise from measurement error or uncertainty inherent in production. A solution to this problem is to compare misallocation across economies since these allocations are feasible. The next section takes this approach by comparing R&D misallocation in the late period with the less-distorted early period. In this regard, the early period acts as a lower bound on the achievable gains from R&D reallocation.

5.2 R&D Misallocation Over Time

I now examine the cost of rising R&D misallocation driven by increasing profit dispersion and its relation to the secular declines in aggregate productivity growth and research productivity. I measure the cost of rising R&D misallocation through a simple experiment that compares growth in the benchmark economy with an economy where the expected profitability of innovations $\bar{\lambda}_a$ is set according to the early period (1976-1990). While I do not take a stance on the drivers of these changes, rising profitability could be related to rising market power or increasing patent protection.

Figure 5 reports the difference in the private returns to innovation and the implied R&D wedges in the two periods. The other parameters are kept at the benchmark values, including expected knowledge spillovers, $\ln \bar{\eta}_a$, consistent with Table 2. Consequently, growth rate differences between the two periods are explained by the R&D allocations and not differences in innovation technologies. Appendix D.1 shows that comparing the benchmark economy to an economy re-calibrated to match early period moments implies a similar decline in aggregate productivity growth due to R&D misallocation.

Figure 5: Private Returns and R&D Wedges by Period



(a) Expected Profitability

(b) R&D Wedge θ_a

Table 8 reports the main results. The model generates a stark growth slowdown between the early and late periods driven by the concentrated increase in the profitability of high volume firms and the decline in profitability of high knowledge firms. R&D misallocation accounts for a decline in productivity growth of around 11%, explaining around one third of the growth slowdown reported by Aghion et al. (2019a). Additionally, output misallocation increases over this period leading to a loss in output of around 0.13%.

Table 8: Rising R&D Misallocation

		Early Period 1976-1990	Late Period 1991-2005
Output Misallocation	χ	1.0000	0.9987
Growth Rate (%)	g	1.83	1.63
Chg. in Growth (%)		-	-11.04

The experiment also helps explain the falling research productivity found by Bloom et al. (2020). Following Proposition 2, measured research productivity is the product of two components: (1) the maximum growth rate g^*, which depends on R&D costs (ψ_a, ψ_e) and expected knowledge spillovers $\ln \bar{\eta}_a$; and (2) the allocative efficiency of R&D. The results indicate that declining allocative efficiency leads to a substantial decline in growth over this period, without requiring an increase in the cost of innovation or a decline in knowledge spillovers. A back-of-the-envelope calculation shows that after accounting for the rise in researchers, R&D misallocation explains around one sixth of the decline in research productivity (see Appendix A.1 for details). A related takeaway of the experiment is that the mechanism is consistent with rising R&D resources over this period since rising profits justify increased investment, even as research productivity declines.

The upshot of the analysis is that R&D misallocation explains a significant share of the decline in aggregate productivity growth and research productivity. Given this, policies that shift cross-firm incentives to innovate could support a recovery in research productivity and aggregate productivity growth, without increasing R&D expenditures.

5.3 Robustness and Other Results

Early period calibration (Appendix D.1). I re-calibrate the model to the early period (1976-1990) and calculate the loss from R&D misallocation. Moving the late-period economy to the early-period economy's allocative efficiency would increase growth by a similar amount as found in the baseline experiment.

R&D intensity (Appendix D.2). I extend the model to include R&D cost shifters and recalibrate the model to match firm-type differences in R&D intensity. The growth gains from R&D reallocation are similar to the baseline experiment.

Social planner's problem (Appendix D.3). I calculate the social planner's R&D allocation that accounts for the dynamic effect of R&D on product accumulation and the cost of output misallocation. The social planner reallocates more R&D resources to high knowledge firms than in the baseline experiment (Table 7) because product accumulation further increases long-run growth. The gap between the social planner's and the growth maximizing allocations is relatively small.

Transition dynamics (Appendix D.4). I calculate the transition between the early and late periods in the two economies in Section 5.2. The transition dynamics are relatively short-lived and that the economy quickly transitions from the early period to the late period balanced growth path.

6 Conclusion

R&D misallocation lowers research productivity and aggregate productivity growth. In this paper, I develop a tractable model of heterogeneous innovative firms that extends the standard model (e.g., Klette and Kortum, 2004) to allow for mismatches in the expected private and public returns to innovation. R&D misallocation arises because firm investment depends on private returns while growth depends on public returns. The empirical analysis highlights persistent wedges between the private and public returns to innovation that widen over time. Using a calibrated version of the model, I find that reallocating R&D to match public returns increases aggregate productivity growth by 18% on impact and doubles productivity growth in the long run. Further, rising R&D misallocation since the 1970s explains over one third of the observed slowdown in aggregate productivity growth.

The results suggest a number of paths for future research. The framework could be used to examine direct channels of R&D misallocation and the rise in R&D misallocation over time. The literature offers many candidates for the latter, such as rising market power (e.g., Cavenaile et al., 2020), patent protection (e.g., Akcigit and Ates, 2019) and intangible investment (e.g., De Ridder, 2020). The model also provides a starting point for other types of R&D misallocation, such as within-firm or cross-sector R&D misallocation. The early-period calibration (Appendix D.1) reveals that high volume firms were both less costly and profitable innovators, potentially related to a change in innovation strategy over time in response to the changing institutional environment (as documented by Kortum and Lerner, 1999). Finally, addressing the gap between private and public returns is a key policy challenge. The comparatively low R&D misallocation in the early period indicates that the R&D allocation could be improved through policy.

References

Abrams, D. S., Akcigit, U., and Grennan, J. (2018). Creative destruction or strategic disruption. Technical report, Working Paper.

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., and Kerr, W. R. (2018). Innovation, reallocation and growth. *American Economic Review*.

Acemoglu, D., Akcigit, U., and Celik, M. A. (2016a). Young, restless and creative: Openness to disruption and creative innovations. Technical report, Working Paper.

Acemoglu, D., Akcigit, U., Hanley, D., and Kerr, W. R. (2016b). Transition to clean technology. *Journal of Political Economy*, pages 52–104.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P., and Li, H. (2019a). Missing growth from creative destruction. *American Economic Review*, 109:2795–2822.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P., and Li, H. (2019b). A theory of falling growth and rising rents. Technical report, Working Paper.

Akcigit, U. and Ates, S. T. (2019). What happened to us business dynamism? Technical report, Working Paper.

Akcigit, U., Hanley, D., and Stantcheva, S. (2019). Optimal taxation and r&d policies. Technical report, Working Paper.

Akcigit, U. and Kerr, W. (2018). Growth through heterogeneous innovations. *Journal of Political Economy*, 126(4):1374–1443.

Asker, J., Farre-Mensa, J., and Ljungqvist, A. (2015). Corporate investment and stock market listing: A puzzle? *Review of Financial Studies*, pages 342–390.

Atkeson, A. and Burstein, A. (2018). Aggregate implications of innovation policy. *Journal of Political Economy*.

Ayerst, S. (2020). Distorted technology adoption. Technical report.

Ayerst, S. (2021). The adoption and diffusion of new technologies. Technical report.

Bento, P. and Restuccia, D. (2017). Misallocation, establishment size, and productivity. *American Economic Journal: Macroeconomics*, 9(3):267–303.

Berkes, E., Manysheva, K., and Mestieri, M. (2022). Global innovation spillovers and productivity: Evidence from 100 years of world patent data. Technical report, Working Paper.

Bloom, N., Jones, C., Van Reenen, J., and Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*, 110(4):1104–44.

Cavenaile, L., Celik, M. A., and Tian, X. (2020). Are markups too high? competition, strategic innovation, and industry dynamics. Technical report, Working Paper.

Chen, Z., Liu, Z., Suarez Serrato, J. C., and Xu, D. Y. (2018). Notching r&d investment with corporate income tax cuts in china. Technical report, Working Paper.

Cornelli, F. and Schankerman, M. (1999). Patent renewals and r&d incentives. *Rand Journal of Economics*, 30(2):197–213.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *Quarterly Journal of Economics*.

De Ridder, M. (2020). Market power and innovation in the intangible economy. Technical report, Working Paper.

Edmond, C., Midrigan, V., and Xu, D. Y. (2015). Competition, markups, and the gains from international trade. *American Economic Review*, 105(10):3183–3221.

Garcia-Macia, D., Hsieh, C.-T., and Klenow, P. J. (2019). How destructive is innovation? *Econometrica*, 87(5):1507–1541.

Gordon, R. J. (2016). Is us economic growth over? faltering innovation confronts the six headwinds. *NBER working Paper*.

Hall, B., Jaffe, A., and Trajtenberg, M. (2001). The nber patent citation data file: Lessons insight and methodological tool. Technical report.

Hall, B., Jaffe, A., and Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *Quarterly Journal of Economics*, 124(4):1403–1448.

Jaffe, A. B. and Lerner, J. (2004). Innovation and its discontents: How our broken patent system is endangering innovation and progress, and what to do about it. *Princeton University Press*.

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

Kogan, L., Papanikolaou, D., Seru, A., and Stoffman, N. (2015). Technological innovation, resource allocation and growth. *Quarterly Journal of Economics*, 132(2):665–712.

Konig, M., Storesletten, K., Song, Z., and Zilibotti, F. (2018). From imitation to innovation: Where is all that chinese r&d going? Technical report.

Kortum, S. and Lerner, J. (1999). What is behind the recent surge in patenting? *Research Policy*, 28:1–22.

Lentz, R. and Mortensen, D. T. (2008). An empirical model of growth through product innovation. *Econometrica*, 76(6):1317–1373.

MacKenzie, G. (2019). Trade and market power in product and labor markets. Technical report.

Peters, M. (2019). Heterogeneous markups, growth and endogenous misallocation. Technical report.

Peters, M. and Walsh, C. (2021). Population growth and firm dynamics. Technical report.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4):707–720.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5):S71–S102.

Scotchmer, S. (1999). On the optimality of the patent renewal system. *The Rand Journal of Economics*, 30(2):181–196.

Online Appendix (Not for Publication)

A Model Appendix

A.1 Other Model Results

Welfare. Welfare on the balanced growth path can be written as

$$W(C_0, g) = \frac{\log(C_0)}{\rho} + \frac{g}{\rho^2},$$

where C_0 is output in period 0 and g is the growth rate. Consider a counterfactual economy with initial consumption $\tilde{C}_0$ and growth $\tilde{g}$. The consumption equivalent between that economy and the baseline economy is equal to

$$1 + \xi = \exp\left\{\log(\tilde{C}_0/C_0) + \frac{\tilde{g} - g}{\rho}\right\}, \tag{21}$$

where ξ is interpreted as the percentage increase in consumption the household would need each period in order to be as well off as in the counterfactual economy.

Firm distribution. The mass of type a firms with size n are described by

$$\dot{M}_a(n) = \begin{cases} x_{e,a}(1) + 2\delta M_a(2) - (x_a + \delta)M_a(1) + \sigma_a(1) & \text{if } n = 1 \\ x_{e,a}(n) + (n-1)x_a M_a(n-1) + (n+1)\delta M_a(n+1) - n(x_a + \delta)M_a(n) + \sigma_a(n) & \text{if } n > 1 \end{cases}$$

where $\sigma_a(n) = \sum_{a'} \gamma_{a',a} M_{a'}(n) - M_a(n)$ describes the relative change of size n firms between types. The mass of firms depends on the entry rate of type a firms $x_e \tilde{h}_a$; the rate that the firm adds products x_a relative to the creative destruction rate δ; and the relative transition between types described by $\sigma_a(n)$. The mass of type a firms is given by $\mathcal{M}_a = \sum_{n=1}^{\infty} M_a(n)$ and the total mass of firms in the economy as $\mathcal{M} = \sum_a \mathcal{M}_a$. The share of type a firms is defined as $m_a = \mathcal{M}_a/\mathcal{M}$.

Aggregate research productivity. From Proposition 2, aggregate research productivity can be written as

$$\frac{g}{S} = \frac{\ln \bar{\eta}_e}{\psi_e}\left[1 + \sum_{a\in\mathcal{A}} \frac{h_a s_a^*}{S}\left(\zeta\theta_a^{1/\zeta} - \theta_a\right)\right], \tag{22}$$

where $\ln \bar{\eta}_e = \sum_{a \in \mathcal{A}} \tilde{h}_a \ln \bar{\eta}_a$. It follows that research productivity can decline either due to a drop in the economy's ability to generate growth—either a decrease in $\bar{\eta}_a$ or an increase in (ψ_e, ψ_a)—or an increase in R&D misallocation, increasing $|\theta_a - 1|$.

Aghion et al. (2019a) report a decline in growth of 30% (≈1.63/2.32-1) while Bloom et al. (2020) report a rise in effective researchers of around two. Together, this implies a decrease in the left hand side of 65% ($\approx 0.7/2.0 - 1$) of its initial value. The results in Section 5.2 show that R&D misallocation accounts for a decline in growth of 11%, or around one sixth (≈ 11%/65%) of the observed decline. There is still a large role for other channels to explain the decline in research productivity and that this remains the more challenging question.

A.2 Proof of Main Results

Lemma 1. *If the value of a firm can be written as $V([\pi_i]_{i=1}^n) = \sum_{i=1}^n V(\pi_i, 1)$ then the entry problem can be written as:*

$$V_e = \max_{x_e} x_e \sum_{a \in \mathcal{A}} \left[\tilde{h}_a \mathbb{E}_a \left[V(\tilde{\pi}, 1)\right] - V_e\right] - \psi_e x_e.$$

Proof The entry problem is given by

$$V_e = \max_{x_e} x_e \sum_{a \in \mathcal{A}} \left[\frac{\tilde{h}_a}{\bar{n}_a} \mathbb{E}_a \left[V([\tilde{\pi}_i]_{i=1}^{\bar{n}_a}, \bar{n}_a)\right] - V_e\right] - \psi_e x_e$$

Given $V([\pi_i]_{i=1}^n) = \sum_{i=1}^{\bar{n}_a} V(\pi_i, 1)$, then $\mathbb{E}_a \left[V([\tilde{\pi}_i]_{i=1}^{\bar{n}_a}, \bar{n}_a)\right] = \mathbb{E}_a \sum_{i=1}^n V(\pi_i) = \bar{n}_a \mathbb{E}_a V(\pi_i, 1)$ implying the result.

Proof of Proposition 1. The firm's problem is given by:

$$\begin{aligned} rV_a([\pi_i]_{i=1}^n, n) - \dot{V}_a([\pi_i]_{i=1}^n, n) = \\ \max_x \sum_{i'=1}^n \left[\begin{matrix} \pi_{i'} Y - w_s B_a + \delta \left[V_a([\pi_i]_{i=1}^n / \{\pi_{i'}\}, n-1) - V_a([\pi_i]_{i=1}^n, n)\right] \\ + x \left[\mathbb{E}_a V_a([\pi_i]_{i=1}^n \cup \{\tilde{\pi}\}, n+1) - V_a([\pi_i]_{i=1}^n, n)\right] - w_s s_a(x) \end{matrix}\right]. \\ + \sum_{a' \in \mathcal{A}} \gamma_{a,a'} \left[V_{a'}([\pi_i]_{i=1}^n, n) - V_a([\pi_i]_{i=1}^n, n)\right] \end{aligned}$$

I solve the value function using the guess and verify method, where I guess that

$$V_a([\pi]_{i=1}^n, n) = \sum_{i=1}^n \frac{\pi_i}{\rho + \delta} Y(t).$$

The guess implies that $\dot{V}_a = gV_a$. Substituting the guess into the value function:

$$(r-g)V_a([\pi_i]_{i=1}^n, n) = \max_x \sum_{i'=1}^n \left[\pi_{i'} - \omega B_a - \delta\left[\frac{\pi_{i'}}{\rho+\delta}\right] + x\mathbb{E}_a\left[\frac{\tilde{\pi}}{\rho+\delta}\right] - \omega\psi_a x^\zeta\right] Y(t)$$

Solving the maximization problem implies that

$$x_a = \left[\frac{\mathbb{E}_a[\tilde{\pi}]/(\rho+\delta)}{\omega\psi_a\zeta}\right]^{\frac{1}{\zeta-1}} = \left[\frac{\bar{\pi}/\psi_a}{\zeta\sum_{a'}\bar{\pi}_{a'}\tilde{h}_a/\psi_e}\right]^{\frac{1}{\zeta-1}}$$

where the second expression follow from substituting $\mathbb{E}_a[\tilde{\pi}] = \bar{\pi}_a$ and the wage rate in (10). It follows that $x\bar{\pi}_a/(\rho+\delta) - \omega\psi_a x^\zeta = \psi_a(\zeta-1)x_a^\zeta$, which is equal to ωB_a by assumption (A). Substituting the interest rate r $(= g+\rho)$ and simplifying the value function confirms the guess:

$$\begin{aligned}
(r-g)V_a([\pi_i]_{i=1}^n, n) &= \sum_{i'=1}^n \left[\pi_{i'} - \omega B_a + \psi_a(\zeta-1)x_a^\zeta\right] Y(t) - \delta V_a([\pi_i]_{i=1}^n, n) \\
(\rho+\delta)V_a([\pi_i]_{i=1}^n, n) &= \sum_{i'=1}^n \pi_{i'}.
\end{aligned}$$

Proof of Proposition 2. The maximum obtainable growth solves

$$g^* = \max_{x_a, x_e} \sum_a \left[h_a x_a + x_e\tilde{h}_a\right] \ln\bar{\eta}_a$$

subject to the resource constraint $S = \psi_e x_e + \sum_a h_a\psi_a x_a^\zeta$. Solving this problem implies that

$$x_a^* = \left[\frac{\psi_e \ln\bar{\eta}_a}{\zeta\psi_a\left[\sum_{a'}\ln\bar{\eta}_{a'}\tilde{h}_{a'}\right]}\right]^{\frac{1}{\zeta-1}}$$

where the definition of θ_a follows. Note that both x_a in the main text and x_a^* above only depend on the relative values of $\bar{\pi}_a$ and $\bar{\eta}_a$. The fourth part of the proposition follows immediately. This first part of the proposition can be derived as:

$$g^* = \sum_a\left[h_a x_a^*\ln\bar{\eta}_a\right] + x_e^*\left[\sum_a\tilde{h}_a\ln\bar{\eta}_a\right] \quad = \sum_a\left[h_a x_a^*\ln\bar{\eta}_a\right] + \left[\frac{S}{\psi_e} - \sum_a\frac{h_a\psi_a(x_a^*)^\zeta}{\psi_e}\right]\left[\sum_a\tilde{h}_a\ln\bar{\eta}_a\right],$$

where the second expression follows from the resource constraint. From the definition of x_a^*:

$$h_a x_a^*\ln\bar{\eta}_a = h_a x_a^*\ln\bar{\eta}_a \times x_a^{*\zeta-1}\left[\frac{\zeta\psi_a\left[\sum_{a'}\ln\bar{\eta}_{a'}\tilde{h}_{a'}\right]}{\psi_e\ln\bar{\eta}_a}\right] = \frac{h_a\psi_a\zeta(x_a^*)^\zeta}{\psi_e}\left[\sum_a\tilde{h}_a\ln\bar{\eta}_a\right].$$

Substituting back into the expression for growth and using $s_a^* = \psi_a x_a^{*\zeta}$ gives

$$g^* = \left[\frac{S}{\psi_e} + \sum_a \frac{h_a s_a^*}{\psi_e}(\zeta - 1)\right]\left[\sum_a \tilde{h}_a \ln \bar{\eta}_a\right]$$

The expression for the equilibrium growth rate follows similar steps:

$$\begin{aligned}
g &= \sum_a \left[h_a x_a \ln \bar{\eta}_a\right] + x_e \left[\sum_a \tilde{h}_a \ln \bar{\eta}_a\right] \\
&= \sum_a \left[h_a x_a \ln \bar{\eta}_a\right] + \left[\frac{S}{\psi_e} - \sum_a \frac{h_a \psi_a x_a^\zeta}{\psi_e}\right]\left[\sum_a \tilde{h}_a \ln \bar{\eta}_a\right] \\
&= \sum_a \left[h_a \theta_a^{\frac{1}{\zeta}} x_a^* \ln \bar{\eta}_a\right] + \left[\frac{S}{\psi_e} - \sum_a \frac{h_a \psi_a \theta_a (x_a^*)^\zeta}{\psi_e}\right]\left[\sum_a \tilde{h}_a \ln \bar{\eta}_a\right] \\
&= \left[\frac{S}{\psi_e} + \sum_a \frac{h_a s_a^*}{\psi_e}(\zeta \theta_a^{\frac{1}{\zeta}} - \theta_a)\right]\left[\sum_a \tilde{h}_a \ln \bar{\eta}_a\right].
\end{aligned}$$

The second part of the proposition then follow from comparing the above expression with g^*. The third part follow from taking the derivative of g with respect to θ_a:

$$\frac{\partial g}{\partial \theta} = \frac{1}{1 + \sum_a h_a \frac{s_a^*}{S}(\zeta - 1)} \left(\theta_a^{1/\zeta - 1} - 1\right),$$

where it follow that the derivative is positive if $\theta_a < 1$ and negative if $\theta_a > 1$.

A.3 CES Production

The baseline model assumes that the final good is produced using a Cobb-Douglas aggregate of intermediate goods j. Cobb-Douglas production implies that the profits of a good j do not depend on its average knowledge spillovers $\bar{k}(t)$ and, consequently, on the innovating firms' knowledge spillovers η. This limits the incentives for firm's with higher expected knowledge spillovers $\bar{\eta}_a$ to innovate since they do not benefit from this innovation component.

To examine the consequences of this incentive, I consider an extension of the baseline model to CES production and where production is $y_j = k_j \ell_j$. CES production creates a link between firm productivity and profits since more productive firms are charge lower prices and capture higher market shares. In this regard, CES production allow for both components of innovation—private returns λ and public returns η—to affect firm innovation decisions. From a calibration perspective, it is unclear that this extension should impact the results since the benchmark calibration does well at matching both firm profitability and other moments related to market share, such as relative employment and sales (Tables 3 and 5). Given this, I focus on the model differences rather than the quantitative differences.

The final good production technology is now given by

$$Y(t) = \left(\int_0^1 \left(\sum_{f \in \mathcal{F}_j} \frac{q_{j_f}}{\bar{q}(t)} y_{j_f}(t) \right)^{\frac{\sigma-1}{\sigma}} dj \right)^{\frac{\sigma}{\sigma-1}}.$$

where $\bar{q}(t) = \left(\int_0^1 q_j^{\sigma-1} dj \right)^{\frac{1}{\sigma-1}}$. I also assume that the distribution of η and λ are independently drawn from type-specific distributions to simplify expressions. The model is otherwise the same as in the main text. The production structure implies that demand for intermediate good j is given by $y_j = Y(t)(q_j/\bar{q}(t))p_j^{-\sigma}$. The main consequence of changing the production structure is that now expenditure on good j depends on its price p_j.

Firms. Solving the firm's problem gives the pricing rule $p = [w(t)/k_j]\,\mu_j$ and $\mu_j = \min\left\{\lambda, \frac{\sigma}{\sigma-1}\right\}$ is the markup set by the firm. The markup must be (weakly) less than the monopolist's profit maximizing markup $\frac{\sigma}{\sigma-1}$. Given the pricing rule, sales for good j are equal to $p_j y_j = Y(t)(\bar{q}(t)w(t))^{1-\sigma}\,(q_j k_j/\mu_j)^{\sigma-1}$. Sales depend on the ability of the firm to charge markups μ_j but also on the relative quality q_j and productivity k_j of good j. This creates a benefit to firms for high public return innovations that is not present in the baseline model. Additionally, sales-per-employee is the same in both models implying that a similar calibration strategy can be used with CES model.[31]

Firm profits are given by

$$\Pi_j = p_j y_j - \frac{w(t)}{k_j}\ell_j = \kappa(t) \left(\frac{\check{q}_j \check{k}_j}{\mu_j} \right)^{\sigma-1} (1 - \mu_j^{-1}) = \kappa(t)\pi_j.$$

where $\kappa(t)$ is a term that is common to all firms and π_j describes the relative profitability of good j, as in the baseline model. The values $\check{q}_j = \frac{q_j}{\bar{q}(t)}$ and $\check{k}_j = \frac{k_j}{k(t)}$ are normalized values of quality and embodied knowledge. Expected profits of an innovating firm a are given by

$$\mathbb{E}_a[\pi_j(t)] = Y(t)\mathbb{E}_a \left[\left(\frac{\lambda_j \eta_j}{\mu_j} \right)^{\sigma-1} (1 - \mu_j^{-1}) \right].$$

For the case where firms draw from a degenerate distribution with $\bar{\lambda}_a < \sigma/\sigma - 1$ then it follows that $\mu_j = \bar{\lambda}_a$ and so $\bar{\pi}_a = \bar{\eta}_a^{\sigma-1}(1 - \bar{\lambda}_a^{-1})$. The main difference in profitability between the baseline model and the CES model is the inclusion of $\eta_j^{\sigma-1}$, which accounts for the

[31] The main differences between markups in the two models is that the monopolist markup acts as a ceiling in the CES case. This is not an issue for the quantitative analysis because average sales-per-employee identifies average markups implicitly accounting for this ceiling.

increase in size of higher knowledge products. The value function and optimal innovation rates are the same as in the baseline model with the appropriate modification of profits.

Aggregate outcomes. Output is given by

$$Y(t) = \frac{\left[\int_0^1 \left(\frac{q_j k_j}{\mu_j}\right)^{\sigma-1} dj\right]^{\frac{\sigma}{\sigma-1}}}{\bar{q}(t)\left[\int_0^1 \frac{(q_j k_j)^{\sigma-1}}{\mu_j^{\sigma}} dj\right]} = \chi \bar{k}(t)$$

where $\chi = \left[\int_0^1 \left(\frac{\check{q}_j \check{k}_j}{\mu_j}\right)^{\sigma-1} dj\right]^{\frac{\sigma}{\sigma-1}} / \left[\int_0^1 \frac{(\check{q}_j \check{k}_j)^{\sigma-1}}{\mu_j^{\sigma}} dj\right]$ and where $\check{q}_j = \frac{q_j}{\bar{q}(t)}$ and $\check{k}_j = \frac{k_j}{\bar{k}(t)}$. Unlike the baseline model, output misallocation in the CES model depends on the joint distribution of knowledge, quality, and markups. However, there are a few similarities between the expressions. First, a proportional increase in the level of markups does not affect output. Second, if the markup is constant for all firms then the markup term μ_j drops out of the expression for output. In this regard, only the dispersion in markups matters for output. Third, a proportional increase in qualities also does not affect output.

It follows that the growth rate is equal to

$$g = \sum_a [x_a h_a + x_e \tilde{h}_a] \frac{\bar{\eta}_a^{\sigma-1} - 1}{\sigma - 1}.$$

Intuitively, this is because increasing the average value of quality does not affect aggregate output and innovation is undirected across goods j.

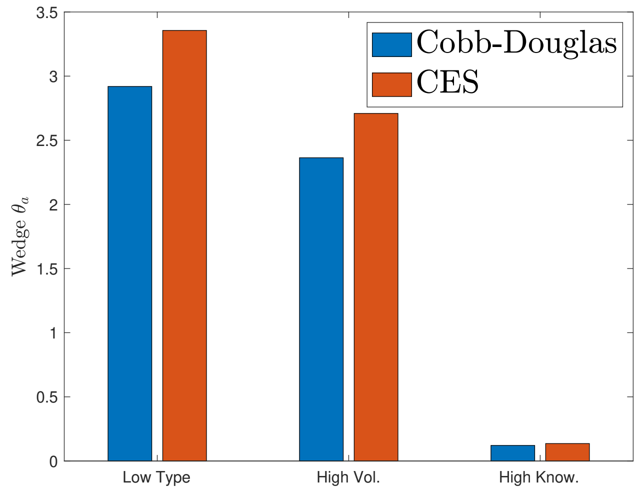
Comparison of wedges. The wedge between the market and growth maximizing R&D allocations in the CES model is given by

$$\theta_a = \left[\frac{\bar{\pi}_a / \sum_{a'} \bar{\pi}_{a'} \tilde{h}_{a'}}{(\bar{\eta}_a^{\sigma-1} - 1)/\sum_{a'}(\bar{\eta}_{a'}^{\sigma-1} - 1)\tilde{h}_{a'}}\right]^{\frac{\zeta}{\zeta-1}}$$

where the differences with the baseline model expression are that the knowledge growth is adjusted by the CES term $\sigma - 1$ and profitability $\bar{\pi}_a$ depends on embodied knowledge.

Despite the change in the model setting and expression, the expressions for sales-per-employee remains the same as in the baseline model implying that the same identification of the private return $\bar{\lambda}_a$. To examine the quantitative impact of CES production, I set $\sigma = 2$ and the remaining parameters as in the baseline calibration. Figure 6 compares the R&D wedges.

Figure 6: Comparison of Cobb-Douglas and CES Wedges



The R&D wedges in the Cobb-Douglas and CES models are almost identical implying similar scope and magnitude of R&D misallocation. The main difference is that in the CES model, the R&D wedges for both high volume and high knowledge firms are slightly larger because of the effect of knowledge spillovers η on market size. However, this is quantitatively negligible compared with distortions caused by differences in private returns. Given that the main effect of CES production is on the profitability of firms, the similarity of wedges implies that the remaining quantitative results are also quantitatively similar.

A.4 Output R&D Costs

In this section, I show that the model produces the same results if R&D costs are paid in output rather than skilled labor. The model is the same as in the baseline text except that there is no skilled labor and R&D costs are paid in output.

Consumption in the output cost model is given by $C = Y - R - \sum_a B_a Y$ where R is the total expenditure on R&D. The second, and main, difference is that now R&D costs for an n product firm are given by

$$n_f R_a(x) = n_f \psi_a x^\zeta Y(t),$$

where $Y(t)$ is now included as a scaling term to prevent explosive growth. Similarly, the cost of entry R&D is now $\psi_e Y(t) x_e$ and the fixed cost paid by firms is $B_a Y(t)$. Total expenditure on R&D as a share of output is then

$$\frac{R}{Y} = \psi_e x_e + \sum_a \psi_a x_a^\zeta h_a. \tag{23}$$

where the right hand side equals total skilled labor employment in the benchmark economy.

The solution to the model follows the same main steps as in the baseline model. Assuming that entry is positive and that the analogue of (A) holds, the model solutions are the same with two notable exceptions. First, the entry condition can now be rearranged to solve for the creative destruction rate

$$\delta = \frac{\sum_{a \in \mathcal{A}} \bar{\pi}_a \tilde{h}_a}{\psi_e} - \rho.$$

Second, the entry rate is equal to $x_e = \delta - \mathbf{x}\left[\mathrm{diag}(1 + \delta - x_a) - \Gamma'\right]^{-1} \tilde{\mathbf{h}}$. All other equilibrium values are the same as in the main text. The upshot being that the reduction in variables allows for the model to be characterized analytically.

A consequence of this setting is that the creative destruction rate δ is determined by the entry condition. The value of δ dictates the aggregate flow of innovations, which along with the relative R&D intensity determines the growth rate. In this regard it is difficult to examine counterfactual economies in equilibrium as the value of δ is pre-determined by the entry condition, which has a strong implication for both the growth rate and total R&D expenditure. The closest comparison with the baseline experiments is to hold the share of R&D expenditure R/Y fixed and examine R&D allocations across firm allowing the creative destruction rate δ to be flexible. It follows from (23) that the counterfactual gains to productivity growth are equal to the baseline experiments. This result is because the total resources used for R&D on the right-hand side of (23) is equal to the skilled labor supply in the benchmark economy. Given that this condition replaces the entry condition, the resulting system of equations is the same as the growth maximizing problem considered in the baseline economy. The consequence is that the output R&D cost model yields the same results as the skilled labor R&D cost model.

The wedges derived in the baseline model can still be derived and capture the degree of R&D misallocation. However, the interpretation is now different as there is an additional inefficiency when R&D is paid in output. Firms now under- or over-invest in R&D because they do not realize the full benefits of new products.[32]

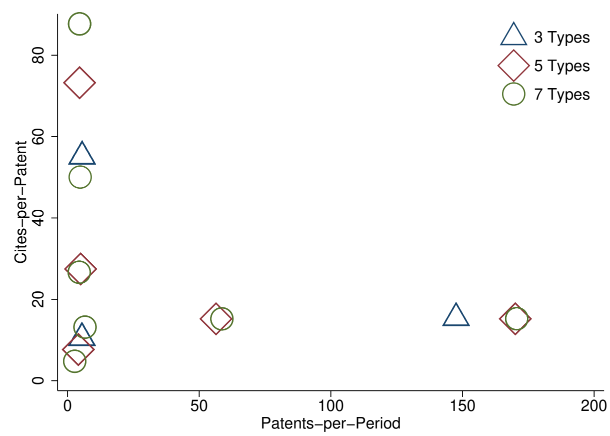
[32] Over-investment is possible because of the separation of public and private returns to R&D. For example, a firm with $\bar{\eta}_a = 0$ and $\bar{\lambda}_a > 0$ will invest in R&D because it is privately optimal but does not generate public value from positive knowledge spillovers. Instead, the added risk from losing a product to this firm will result in other firms further lowering investment.

B Empirical Appendix

B.1 Granularity of Firm Types

The baseline classification of firm types sets the k means clustering algorithm to three groups. Figure 7 compares alternative constructions where five and seven types are used.

Figure 7: Granularity of Firm Types



The main takeaway from adding more firm types is that the placement of the new types is between the low types and high volume or high knowledge firms. Importantly, adding new types does not appear lead to a new combination of innovation characteristics, such as a true high (high frequency, high knowledge) type.[33] Instead, the increased granularity leads to the addition of medium ability types that specialize in one of the two dimensions.

B.2 Comparison with Single Dimension Types

This section presents a comparison between the baseline classification of firms into different types and a classification of firms based on the total citations-per-period. This produces three firm types that can be thought of as low, mid, and high citations-per-period types. On average, the major difference between the three types is the number of patents-per-period. In terms of citations-per-patent, the mid and high types are relatively similar (26 and 31 citations-per-patent) and both outperform the low types (14). Table 9 compares the assignment of types between the unidimensional characteristic and the baseline multidimensional characteristics.

[33] This is not driven by the choice of odd numbers. Moving to four types leads to the addition of a true medium type, but this type disappears under the six type specification.

Table 9: Comparison with Single Dimension Types

Unidimensional Groups	Multidimensional Group			
	Low Type	High Vol.	High Know.	Total
Low Type	72,290	27	6,361	78,678
Mid Type	1,877	693	1,131	3,701
High Type	34	1,203	213	1,450
Total	74,201	1,923	7,705	83,829

The results highlight that the two methods produce different clusters of firms. The unidimensional clustering is most closely related to the high volume firms in the baseline types. Most high volume firms tend to be classified as high types in the unidimensional groups. Only around 1% of high volume firms are classified as low type firms. The majority of low types are common across the two classifications with some of the baseline low types becoming mid types in the unidimensional classification. High knowledge firms are a bit more mixed with around 20% being classified as either mid or high types while the majority of high knowledge firms are classified as low types as a consequence of relatively low patents-per-period.

Both the model and the empirics highlight the importance of the heterogeneity between high volume and high knowledge firms. In this regard, the unidimensional classification misses out on this important aspect of heterogeneity. The difference in the constructed types is most evident when considering the implications for R&D misallocation. The unidimensional types are ordered in terms of their relative innovative capabilities, which coincides with their relative profitability. This would understate the degree of misallocation because from this perspective, high types are relatively more innovative than mid types and also better rewarded.

B.3 Robustness and Additional Results

The baseline results focus on the private and public returns measured by sales-per-employee and citations-per-patent given that these outcomes directly relate to the model outcomes. In this Appendix Section, I briefly outline the robustness to alternative specifications and outcomes. I also explore alternative outcomes that support the model characteristics and the differences between firm types.

B.3.1 Other Innovation Outcomes

Table 10 reports results for patents-per-employee, citations-per-employee, and R&D intensity (R&D as a share of sales). As with the baseline results, the results are divided between the early and late periods.

Table 10: Other Innovation Outcomes

	Early Period (1976-1990)			Late Period (1991-2005)		
	(1) pat/emp	(2) cites/emp	(3) R&D/sales	(4) pat/emp	(5) cites/emp	(6) R&D/sales
High Vol.	0.309	0.717**	-0.201	1.449***	1.264***	-0.310**
	(0.324)	(0.341)	(0.204)	(0.389)	(0.302)	(0.129)
High Know.	0.206	1.314***	0.305***	0.180	1.500***	0.240***
	(0.185)	(0.173)	(0.0967)	(0.121)	(0.141)	(0.0613)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6883	6883	6287	7459	7438	6867

*Notes: * p<0.10, ** p<0.05, *** p<0.01. Robust standard errors clustered at the firm level are reported in parentheses. All regressions include year and sector (four-digit SIC) fixed effects. All columns are estimated using PPML. pat/emp measures total patents applied in period t divided by total employment in period t. cites/emp measures total citations received on all patents applied for in period t divided by total employment in period t. R&D/sales measures the log of one plus reported R&D in period t divided by total sales in period t. Unreported R&D values are assumed to be zero.*

The results show that high knowledge firms invest more in R&D as a share of total sales than high volume firms, in contrast with the predictions of the model. However, it is important to note that while R&D in the model captures all innovative activities, R&D in the data is narrower and only reflects what is reported. This is evident by the number of firms that patent frequently and report zero or near-zero R&D expenditure.[34] Additionally, while the baseline results show higher R&D intensity for high knowledge firms, the results are sensitive to the specification and excluding very high R&D intensity firms (e.g., reported R&D greater than 75% of sales). In Appendix D.2, I show that the main results hold if relative R&D intensity in Table 10 is taken at face value.

B.3.2 Other Market Outcomes

Table 11 reports other market outcomes related to size and growth by firm type. In addition to being more profitable than other firm types, high volume firms also tend to be substantially larger. High knowledge firms appear to be slightly smaller than low types, although the coefficient is not precisely estimated.

[34] This also appears to differ by type with over half of observations by low type and high volume firms reporting R&D intensity below 5% compared with only one-sixth for high knowledge firms. Similarly, high knowledge types are much more likely to report R&D intensity above 75% than the other types.

Table 11: Firm Size and Growth Outcomes

	(1) ln(sales)	(2) ln(emp)	(3) g(sales)	(4) g(emp)	(5) g(sales/emp)
High Vol.	2.429***	2.060***	0.136***	0.0551***	0.0194
	(0.267)	(0.235)	(0.0256)	(0.0185)	(0.0194)
High Know.	-0.151	-0.0872	0.0138	0.0597***	-0.0421**
	(0.136)	(0.102)	(0.0230)	(0.0152)	(0.0168)
Year FE	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes
Observations	7459	7459	6503	6503	6503

*Notes: * p<0.10, ** p<0.05, *** p<0.010. Robust standard errors clustered at the firm level are reported in parentheses. All regressions include year and sector (four-digit SIC) fixed effects. Growth of variable x is measured as* $g(x) = (x_{t+1} - x_t)/0.5(x_{t+1} + x_t)$. *Columns (3) to (5) include the log of one plus the R&D expenditure as an additional control. Columns (3) and (4) include measures of firm size, log of sales and employment, as additional controls.*

sSales and employment growth are faster for both high volume and high knowledge firms, which is consistent with the baseline calibration (Table 5). High volume firms do not tend to grow in terms of sales per employee, whereas high knowledge firms appear to shrink in this measure of profitability over time. This is important in the context of the model assumptions because it shows that the relative ranking of the firm types in terms of sales-per-employee does not change over time. In this regard, the results shows that high knowledge firms are not receiving higher future profits for innovations.

B.3.3 Alternative Measures of Private Returns

Table 12 reports the baseline results using alternative measures of profitability. The three additional measures examined are (i) sales per cost of goods sold, (ii) sales minus costs of goods sold divided by sales, and (iii) operating income before depreciation divided by sales. The results show the same trends in profitability as in the baseline results.

B.3.4 Public and Private Returns Using Patent Values

Table 13 constructs alternative measures of public and private returns using the patent values constructed by Kogan et al. (2015). The alternative private return (PV for patent value) is taken directly as the patent value from Kogan et al. (2015), which measures the change in stock price of a firm around the date of the patents announcement. The alternative public return (FPV for future patents' value) measures the value of patents that cite the firm's

Table 12: Alternative Profit Measures

	Early Period (1976-1990)			Late Period (1991-2005)		
	(1) ln(sales/cogs)	(2) PM1	(3) PM2	(4) ln(sales/cogs)	(5) PM1	(6) PM2
High Vol.	0.00788	0.00248	0.0206**	0.340***	0.0732***	0.0568***
	(0.0429)	(0.0183)	(0.00894)	(0.0965)	(0.0249)	(0.0115)
High Know.	-0.0432	-0.0158	-0.00233	-0.0302	0.0132	-0.00718
	(0.0363)	(0.0151)	(0.00664)	(0.0645)	(0.0174)	(0.00591)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6876	6883	6864	7442	7459	7407

*Notes: * p<0.10, ** p<0.05, *** p<0.010. Robust standard errors clustered at the firm level are reported in parentheses. All regressions include year and sector (four-digit SIC) fixed effects. PM1 measures sales minus cost of goods sold divided by sales. PM2 measures operating income before depreciation divided by sales. Negative values of PM1 and PM2 are set to zero.*

patents issued in the current period. This measure captures the market determined value of future innovations that build on a firm's patent. I exclude the value of patents by the same firm because they may directly relate to the value of the original patent.[35]

Table 13 shows a similar pattern to the baseline results. High volume firms are the best at generating private returns and somewhere between high knowledge and low type firms at generating public returns. Additionally, the ability of high volume firms to generate private returns to innovations has increased over time. Similar to the baseline results, high knowledge firms generate the largest public returns.

Kogan et al. (2015) find that patent value is strongly correlated with citations-per-patent. This is somewhat at odds with the results in columns (2) and (5) since high knowledge firms, which tend to have higher citations, have lower patent value than high volume firms. Columns (3) and (6) shows that citations-per-patent remain positively related to patent value even after accounting for firm types. That is, within firm types more cited patents tend to be more valuable.

B.4 Alternative R&D Wedges

I consider alternative measures of the firm-level R&D wedge constructed in (18). The first alternative measure (Cost Approach) constructs the R&D wedge following more closely to

[35] I include patents that are applied for by multiple firms even if they include the same firm.

Table 13: Alternative Measures of Public and Private Returns

	Early Period (1976-1990)			Late Period (1991-2005)		
	(1)	(2)	(3)	(4)	(5)	(6)
	ln(FPV)	ln(PV)	ln(PV)	ln(FPV)	ln(PV)	ln(PV)
High Vol.	0.808***	1.248***	1.193***	0.667***	1.432***	1.407***
	(0.110)	(0.176)	(0.173)	(0.114)	(0.204)	(0.206)
High Know.	1.332***	0.255**	0.0682	1.289***	0.444***	0.370***
	(0.127)	(0.127)	(0.135)	(0.104)	(0.112)	(0.120)
ln(cites/pat)			0.171***			0.0828**
			(0.0347)			(0.0360)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3885	3702	3668	3711	4414	3883

*Notes: * p<0.10, ** p<0.05, *** p<0.01. Robust standard errors clustered at the firm level are reported in parentheses. All regressions include year and sector (four-digit SIC) fixed effects. All columns are estimated using PPML.* ln *PV (patent value) measures the log of average patent value (from Kogan et al., 2015) of patents applied for by firm f in period t.* ln *FPV (future patents' value) measures the average patent value (from Kogan et al., 2015) of all other firm $f' \neq f$ patents that cite patents applied for by firm f in period t.*

the approach developed by Hsieh and Klenow (2009). Specifically, I infer the R&D wedge as gap between the innovation output and innovation costs, measured by total citations divided by total R&D expenditure reported by firms. It is follows that the firm-level R&D wedge in the model is also equal to:

$$\tau_{f,t} = \left[\frac{\text{Total Citations}_{f,t}}{\text{R\&D Expenditure}_{f,t}}\right]^{-\frac{\zeta}{\zeta-1}}, \tag{24}$$

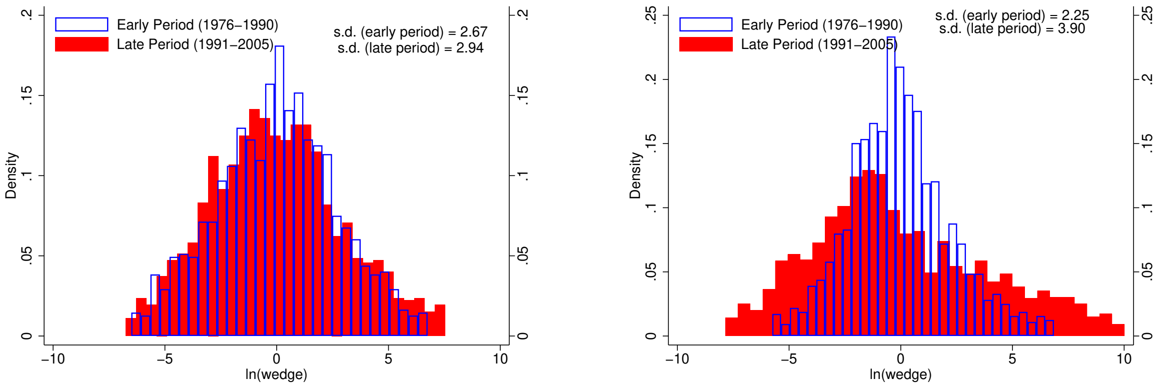
The wedge $\tau_{f,t}$ is the same wedge discussed in Section 3.2.[36] That is, (18) and (24) should produce the same measured wedges, up to a constant, if: (1) the only source of R&D misallocation is the gap in private and public returns; (2) there is no measurement error. In terms of measurement, (24) implicitly infers private returns from R&D expenditure and innovative output but assumes that R&D is accurately recorded in the data.

The second alternative measure (Alt. Profitability) uses operating income before depreciation divided by sales as the measure of profitability π rather than log of sales-per-employee

[36] The exponent makes the value of the wedge directly comparable to the value of (18). Without the exponent, the expression is the analogue of using the average product of a factor (e.g., capital) to examine output misallocation. Since the exponent scales the log wedge, it does not affect the pattern of misallocation.

$\ln \lambda \approx \pi$. Figure 8 plots the alternative R&D wedge distribution for the early and late periods. As in the baseline, wedges are constructed at the firm level for consecutive five-year periods and variables are demeaned each year prior to constructing the wedges.

Figure 8: R&D Wedges by Firm Type with R&D Targets



(a) Cost Approach

(b) Alt. Profitability

Notes: Figure (a) plots the histogram estimate of log of the firm-level wedge in (24) over the early (1976-1990) and late (1991-2005) periods. Figure (b) plots the histogram estimate of the log for the firm-level wedge in (18) using operating income before depreciation instead of log of average sales-per-employee as the measure of profitability. In both figures, the top and bottom two percent of observations, based on firm-level wedges, are dropped in each period. The mean of log wedges is normalized to zero in each period.

The figures highlights a similar increase in the dispersion of R&D wedges over time. The standard deviation of wedges increases by around 10% in the cost approach and by almost double for the alternative profitability approach. This result is reassuring despite the clear measurement issues associated with R&D expenditure (e.g., unreported values). It is also reassuring to note that the wedges are highly correlated with the baseline measure. The coefficient from regressing baseline wedge on cost approach wedge is 0.21 (0.016) and on the alternative profitability wedge is 0.48 (0.011), where the standard error is reported in parentheses, suggesting that the same changes in relative profitability across firms are showing up in R&D expenditure.

C Calibration Appendix

C.1 Identification

Table 14 reports the change in the calibration moments given a 10% increase in parameter values. The table formalizes the intuition for the relationship between model moments and parameter identification discussed in Section 4.

Table 14: Change in Moments from a 10% Change in Parameter Values

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
ψ_L	−1.2	3.3	−0.4	1.2	0.2	−8.1	2.0	−3.5	2.5	−0.3	−2.7	0.0	1.6
ψ_V	−20.3	−0.2	0.0	1.6	−0.5	2.3	−16.1	2.2	6.2	−1.0	10.8	0.1	29.5
ψ_K	−0.9	−1.1	−0.4	−1.0	1.1	−1.7	1.0	−9.1	1.6	−0.7	−0.6	0.0	0.0
λ_L	−7.5	−2.2	0.0	−0.1	0.5	4.2	−7.9	−2.3	−13.0	34.9	11.9	6.9	14.4
λ_V	1.5	2.0	−0.1	−0.3	0.6	−5.1	9.8	−4.2	24.6	1.1	−10.4	0.5	−29.8
λ_K	−3.1	2.1	0.2	1.5	−1.1	0.6	−2.8	8.2	0.8	−28.5	4.0	0.9	7.6
μ_L	0.0	22.3	12.0	1.6	−1.4	1.1	0.0	−8.1	0.0	−16.8	0.0	0.0	0.0
μ_V	0.0	3.5	0.8	24.6	−0.4	−0.5	0.0	4.4	0.4	−16.3	0.0	0.0	0.0
μ_K	0.0	27.3	−1.2	2.5	14.6	−1.6	0.0	12.1	−1.3	16.9	0.0	0.0	0.0
$\tilde{h}_V$	10.4	0.6	0.2	0.6	0.3	0.5	−2.2	−1.1	1.6	0.7	1.2	0.1	0.1
$\tilde{h}_K$	1.2	3.3	0.4	−0.3	0.1	0.8	0.5	1.1	−1.1	2.2	−1.0	−0.2	−3.2
ψ_e	−52.4	45.0	6.4	0.2	3.9	−21.5	−11.0	−9.7	10.3	14.9	−18.4	0.4	−88.8
Λ	6.9	−0.9	0.6	−0.1	−0.8	6.3	6.5	8.2	−2.7	1.9	0.0	0.0	0.0

Notes: The values report the percent change in the indicated moment for a 10% change in the indicated parameter, excepted for λ_a which is calculated as $1.1 \times (\lambda_a - 1)$. The moments are: M1 share of high volume firms; M2 share of high knowledge firms; M3 average citations-per-patent of low type firms; M4 average citations-per-patent of high volume firms; M5 average citations-per-patent of high knowledge firms; M6 average patent count of low type firms; M7 average patent count of high volume firms; M8 average patent count of high knowledge firms; M9 sales-per-employee of high volume firms; M10 sales-per-employee of high knowledge firms; M11 R&D Intensity; M12 average profitability; M13 entry rate. Moments and calculations are described in Section 4.

A striking feature of the matrix is that some areas are relatively sparse. This is driven by some of the parameters being related to the patenting outcomes in the extended model but not the equilibrium outcomes in the main model. For example, the citation distribution does not directly affect the equilibrium entry rate. In the joint calibration, this is more complex since the citation distribution affects the recorded types $\hat{a}$ in the simulated data, which may in turn have an impact on the choice of parameters that affect the entry rate.

C.2 Comparison of Data and Model Transition Matrices

In the baseline calibration, the transition matrix Γ is set to match the implied transition matrix from the data. Table 15 compares the empirical transition matrix with the transition matrix constructed using the simulated model. The empirical transition matrix is calculated as the share of firms in both periods that transition from type $\hat{a}$ to type $\hat{a}'$. The simulated model transition is constructed as in the data. The initial five-year period is constructed as described in Section 4.2, which is then extended for an additional five-year period. As in the baseline calibration, I construct the recorded types $\hat{a}$ for each firm-period based on innovation outcomes, rather than using the true types a. This allows for misclassification of firm types and creates the same type of bias as in the empirical data. The simulated transition matrix is constructed as the probability that a firm with recorded type $\hat{a}$ in the first five-year period will be recorded type $\hat{a}'$ in the second five year period, conditional on remaining the data.

Table 15: Comparison of Empirical and Model Implied Transition Matrix

	Empirical Transition			Simulated Model Transition		
	Low Type	High Vol	High Know	Low Type	High Vol	High Know
Low Type	93.7	1.5	4.8	96.6	0.2	3.2
High Vol	20.0	78.5	1.5	22.1	77.4	0.5
High Know	58.6	2.6	38.8	68.7	0.0	31.3

Notes: Model transition based on recorded firm types of a simulation of 30,000 firms of each type, as described in Section 4.2.

The transition matrices are remarkably similar given that the model transition matrix is taken without modification from the data. Relative to the empirical transition, the model transition displays less own-type persistence for high knowledge firms and less transition between the high types. The former is likely to dampen the gains from R&D reallocation since it implies a lower product share of high knowledge firms. The latter is unlikely to have a major impact on the results because the transition probabilities are small, but would also tend to dampen the gains from R&D reallocation

C.3 Ex-Ante Versus Ex-Post Heterogeneity

A concern is that the construction of the firm types reflect ex-post heterogeneity in patenting outcomes rather than ex-ante differences in firm innovative capabilities. The calibration addresses this by allowing for the same ex-post heterogeneity in patenting outcomes and type misclassification in the model moments. In this section, I examine whether a model with a single firm type could replicate the data moments.

For the single type-model, I set the private returns λ_a and average citations μ_a to match the average values from the baseline calibration, such that the new values are $\lambda'_a = \sum_a m_a \lambda_a$ and $\mu'_a = \sum_a m_a \mu_a$. I also set the public returns η_a to the values implied by μ'_a. Table 16 reports the resulting model moments from this experiment.

The single-type model is unable to replicate the gaps in relative sales-per-employee since all firms draw from the same distribution of private returns λ. In this regard, it is clear that the single-type model is unable to replicate a key feature of high volume firms in the data. The single-type model explains only 1% ($\approx (0.05 \times 11.0 \times 100.3)/(2.3 \times 15.4 \times 147.6)$)

Table 16: Model Moment based on Single Type

Moment	Data	Single-Type Model
Firm Share	(88.5 , 2.30 , 9.20)	(94.4 , 0.05 , 5.54)
Citations-per-Patent	(10.5 , 15.4 , 55.2)	(11.1 , 11.0 , 48.6)
Patent Count	(5.5 , 147.6 , 5.5)	(5.8 , 100.3 , 3.4)
Rel log(Sales-per-Emp)	(0.00 , 0.36 , -0.11)	(0.00 , -0.00 , -0.00)

Notes: Single-Type Model moments based on recorded types of a simulation of 30,000 firms, as described in Section 4.2.

of total knowledge spillovers produced by high volume firms and less than one-third ($\approx (5.5 \times 48.6 \times 3.4)/(9.2 \times 55.2 \times 5.5)$) of total knowledge spillovers produced by high knowledge firms. The single-type model does reasonably at matching citations-per-patent but is unable to replicate either the frequency of firms that fit the criteria or the patent rate of these firms.

The experiment should not be taken as an assessment of how much of the results are explained by ex-post heterogeneity. First, the model incorporates the same source of ex-post heterogeneity implying that the baseline results account for ex-post heterogeneity. Second, Table 16 overstates the role of ex-post heterogeneity in patenting outcomes. This is because the single-type model pools ex-ante heterogeneity across firm types leading to a higher likelihood that firms are recorded as high volume or high knowledge firms and improving outcomes of low type firms. This is clearly evident in Table 16 by the fact that low type firms over-perform the data across all dimensions. In this regard, the experiment is a test of whether ex-post heterogeneity can explain the data rather than how much of the data can be explained.

D Additional Quantitative Results and Extensions

D.1 Early Period Calibration

The baseline results find that growth declined because of rising R&D misallocation. The decline in growth is measured as growth in the benchmark economy compared with growth in a counterfactual economy where profitability of firm types is adjusted to match the early period of the data. An alternative approach is to recalibrate the economy to the early period and compare the R&D misallocation in the early and late period economies.

I follow the same approach as in the baseline calibration. I reduce the stock of skilled labor to half its initial value, such that $S = 0.085$, to be consistent with the increase in effective researchers documented by Bloom et al. (2020). I also target an R&D intensity of

6.6% to be consistent with the late period value and increase in R&D to GDP reported by the National Science Foundation. Table 17 summarizes the fit of the model.

Table 17: Early Period Calibration Moments

Moment	Data	Model
Firm Share (%)	(89.3 , 2.30 , 8.40)	(89.7 , 2.28 , 8.04)
Citations-per-Patent	(11.1 , 13.1 , 54.6)	(10.8 , 13.0 , 53.9)
Patent Count	(5.2 , 150.7 , 4.1)	(5.5 , 140.2 , 4.5)
Rel log Sales-per-Emp	(0.00 , 0.00 , 0.00)	(-0.01 , -0.01 , -0.00)
R&D Intensity (%)	6.6	6.3
Avg Profit Margin (%)	12.6	13.0
Entry Rate (%)	10.2	10.3
Growth Rate (%)	2.32	2.32

Notes: Moments are ordered corresponding to low type, high volume, and high knowledge firm types. Type-specific moments are calculated using recorded types from data generated from a simulation of 30,000 firms (described in Section 4.2).

The table highlights several differences between the early and late period calibrations. The most notable difference is the sharp rise over time in the sales-per-employee by high volume firms. In terms of innovation outcomes, the performance of the three firm types is relatively similar to the early period, as also noted in the empirical analysis. The aggregate moments highlight the decline in entry rate and growth as well as the rise in average profitability. Table 18 reports the parameter values for the early period.

Table 18: Comparison of Early and Late Period Parameters

Parameter		Early Period	Late Period
Entry Cost	ψ_e	1.11	1.91
Innovation Cost	ψ_a	(8.67 , 4.16 , 5.72)	(11.44 , 8.34 , 5.50)
Entry Probability	$\tilde{h}_a$	(0.57 , 0.26 , 0.16)	(0.52 , 0.24 , 0.23)
Avg Private Return	$\bar{\lambda}_a$	(1.15 , 1.15 , 1.15)	(1.08 , 1.14 , 1.39)
Avg Citation	μ_a	(2.08 , 2.31 , 3.22)	(2.01 , 2.50 , 3.35)
Avg Public Return	$\bar{\eta}_a$	(1.16 , 1.20 , 1.57)	(1.08 , 1.13 , 1.32)
Patent-Innovation Ratio	Λ	5.80	4.50
Know Spillover-Cite Ratio	ν	0.0135	0.0073

Notes: Parameters are ordered corresponding to low type, high volume, and high knowledge firm types. Externally calibrated parameters are kept the same as in the baseline calibration with the exception of the total stock of skill labor, which is set to $S = 0.085$.

The parameter differences reflect the differences in moments between the early and late period calibrations. R&D costs (ψ_a, ψ_e) are substantially lower than in the late period calibration, reflecting both the lower stock of researchers S and higher growth rate g in the data. This reflects the secular decline in research productivity found by Bloom et al. (2020). Relative R&D costs also display a different pattern, with high volume firms having the lowest R&D costs and the biggest change relative to the late period calibration. The relative knowledge spillovers also narrow slightly relative to the late period, reflecting a narrowing in citations-per-patent in the data, and each citation reflects a larger contribution to aggregate productivity growth.

Table 19 compares the potential gains in growth from reallocating R&D in both the early and late period. As in the baseline experiment, I compare the growth rates when R&D is reallocated both under the assumption that firms and products are held fixed (Fixed Distribution) and able to vary endogenously (Flexible Distribution).

Table 19: R&D Misallocation in the Early and Late Periods

	Early Period	Late Period
Benchmark Growth Rate (%)	2.32	1.63
Counterfactual Growth Rate (%)		
Fixed Distribution	2.47	1.91
Flexible Distribution	2.83	3.06

The two experiments reveals slightly different takeaways about growth over this period. The fixed distribution experiment shows almost the same result as the baseline results. The decline in the maximum obtainable growth rate was only around 23% ($\approx 1.91/2.47 - 1$), around two thirds of the decline in the growth rate. Put differently, if R&D allocative efficiency was held fixed, the growth rate in the late period would be around 1.79% ($\approx 2.32/2.47 \times 1.91$) or around 10% higher. This value is similar to the baseline experiment.

The flexible distribution experiment shows that the long-term potential of the economy increased relative to the early period. The late-period BGP growth rate from reallocating R&D increases relative to the early period. The increase in the growth potential is driven by the higher share of high knowledge firms in that the economy can support in the BGP equilibrium, in part due to higher $\tilde{h}_K$. An important caveat with this result is that the doubling of R&D resources S implies that the increase in growth would still constitute a decline in aggregate research productivity.

D.2 Adding R&D Expenditure as a Calibration Target

The baseline calibration identifies R&D wedges using relative profitability and citations-per-patent across firm types. An important moment that is not directly targeted in the calibration is the relative research intensity. Instead, relative R&D intensity is inferred using the firm's optimality conditions. Appendix Table 10 shows that research intensity is higher for high knowledge firms. In general, there are reasons to suspect that reported R&D expenditures do not accurately reflect the entirety of innovation-related expenditures. For example, this could be the consequence of R&D being misreported (e.g., many firms report zero R&D expenditure) or R&D expenditure excluding activities not legally classifiable as R&D expenditure (e.g., consulting fees).[37] Given these concerns, I avoid directly targeting relative R&D expenditures in the baseline calibration. In this section, I explore the quantitative robustness of the results when R&D expenditures are also targeted.

To this end, I extend the model to include an additional wedge on R&D expenditures. Specifically, I assume that R&D expenditure is equal to

$$\text{Actual R\&D Expenditure} = \varphi_a \times \underbrace{w_s(t)\psi_a x_a^{\zeta} n}_{=\text{Reported R\&D}}$$

where φ_a is a cost shifter on R&D expenditure that is unreported in the data. An interpretation is that the cost shifter represents true costs to R&D that are not reported, or unobserved (as in, for example, Akcigit et al., 2019). In this case, the baseline experiments are correct and φ_a should be considered to be the same as ψ_a, i.e., there is no loss in treating the two as a composite. This is my preferred interpretation. An alternative interpretation is that φ_a represents an additional distortion to the market and could be removed through resolving institutional frictions. In this interpretation, the distortions φ_a are analogous to the taxes in the misallocation literature (see Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009), where $\varphi_a = 1$ indicates no distortion. I explore the quantitative implications of this latter interpretation in this section since it provides an alternative view of R&D misallocation.

Given the modification to R&D expenditure, the R&D wedge is now given by

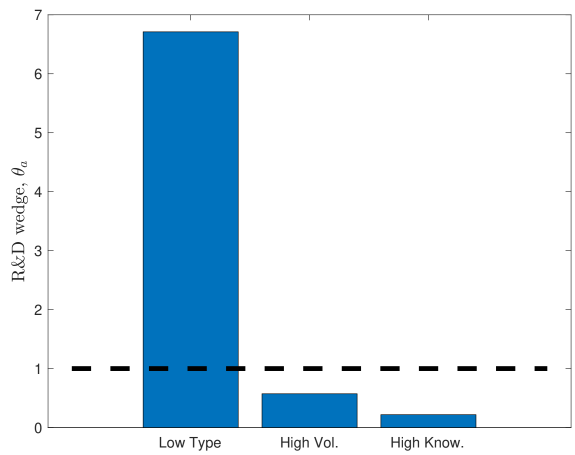
$$\theta_a = \left[\frac{\bar{\pi}_a / \sum_{a'} \bar{\pi}_{a'} \tilde{h}_{a'}}{\varphi_a \ln \bar{\eta}_a / \sum_{a'} \varphi_{a'} \ln \bar{\eta}_{a'} \tilde{h}_{a'}}\right]^{\frac{\zeta}{\zeta-1}}.$$

The interpretation is the same as in the baseline model. The impact of the cost shifter φ_a

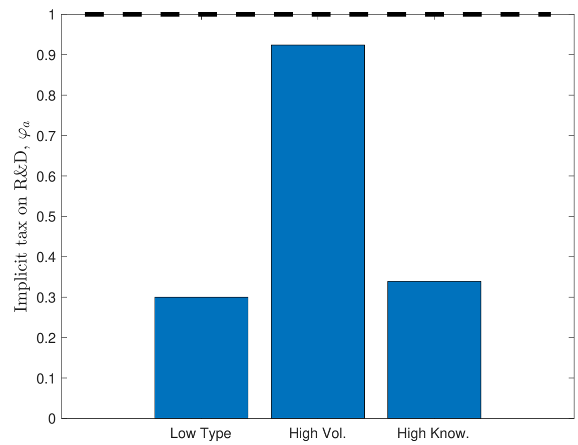
[37] Additionally, these issue appear to bias towards higher R&D spending by high knowledge firms, which less frequently reports zeros and more frequently report high R&D shares (>75% sales), and lower R&D spending by high volume firms, which more frequently report zeros.

is that now the cost of R&D affects the gap between the market and growth maximizing allocations. I set the cost shifter φ_a to match the relative R&D intensity between types in Table 10 and scale the average value of φ_a such that the average tax across product lines is one. I also set the product of R&D costs ψ_a and the cost shifter φ_a to be equal to R&D costs in the baseline model. This implies that all other model moments will match the baseline calibration when parameters are held at their benchmark values. Figure 9 reports wedges θ_a and cost shifters φ_a for each firm type.

Figure 9: R&D Wedges by Firm Type with R&D Targets



(a) R&D Wedge θ_a



(b) R&D Cost Shifter φ_a

The model interprets the low R&D intensity of high volume firms with a high R&D cost shifter. Consequently, the R&D wedge adjusts such that high volume firms are allocated close to the growth maximizing value while low type firms are even more over-allocated R&D relative to the baseline calibration. High knowledge firms still perform relatively too little R&D compared to the growth maximizing R&D allocation. The cost shifter changes the dynamics of R&D misallocation to be a reallocation from low type firms to high knowledge firms. Table 20 reports the gains from reallocating R&D to maximize growth by setting the R&D cost shifters equal to one, $\varphi_a = 1$, and profitability proportional to knowledge spillovers, $\bar{\pi}_a \propto \ln \bar{\eta}_a$. As in the baseline experiments, I consider the gains when the distribution is held fixed at the benchmark distribution and endogenously adjusts to the new R&D allocations.

For the fixed distribution experiment, the inclusion of R&D wedges slightly increases the gains from reallocation relative to the benchmark experiment. This is because there is now an additional channel of reallocation compared to the baseline experiment. For the flexible distribution experiment, the inclusion of R&D wedges decreases the potential growth gains from R&D reallocation by around half. The decline in the growth gains is due to the implied investment cost ψ_K of high knowledge firms declining relative to the baseline

Table 20: Gains from Reallocation with R&D Expenditure as a Target

		Benchmark	Counterfactual	
			Fixed	Flexible
Output Misallocation	χ	0.9987	0.9956	0.9974
Growth Rate (%)	g	1.63	1.55	1.49
Chg in Growth (%)	$g^{CF}/g - 1$	-	-4.5	-8.2
Chg in Welfare (%)	ξ	-	-3.9	-6.6

parameterization, leading to less scope for high knowledge product accumulation.

The baseline experiments provide support for policies that reallocate R&D resources from high volume to high knowledge firms. The above experiment highlights a note of caution to policy by showing that if R&D expenditure values are taken at face value then high volume firms may also be under allocated R&D resources. Consequently, disincentivizing R&D by high volume firms can be potentially damaging to aggregate productivity growth.[38] Given this, the more prudent policy advice is to focus on encouraging R&D and innovation by high knowledge firms that tend to invest too little in R&D in both calibrations.

D.3 Dynamic Social Planner's Problem

I define R&D misallocation relative to the growth maximizing R&D allocation to draws a clear parallel with the output misallocation literature where the primary interest is examining how misallocation affects aggregate productivity, not welfare. This comparison is also more relevant to the question of slowing productivity growth and relies less on the model structure. An alternative approach is to define R&D misallocation relative to the welfare maximizing R&D allocation that would account for the effect of markup dispersion on output misallocation as well as the effect of the R&D allocation on the product distribution. I compare the dynamic social planner's allocation with the growth maximizing allocation.

The constrained dynamic Social Planner's Problem is given by

$$\max_{x_e, [x_a, h_a]_{a \in \mathcal{A}}} W(C_0, g) \tag{25}$$

subject to (6), (9), (13), and the resource constraints and where welfare $W(c_0, g)$ is defined in Appendix A.1. I assume that the social planner is unable to reallocate product lines across firms. If this were the case then the social planner would reallocate all product lines to high

[38]There are other potential reasons that disincentivizing R&D by one group of firms may lower aggregate productivity growth. A simple example is that if R&D resources are perfectly elastically supplied then lowering R&D by a firm type lowers aggregate R&D and consequently growth.

knowledge firms because they have the highest research productivity.

Table 21 compares output, growth and welfare under the benchmark allocation, the knowledge spillover allocation (Know Spill), and the dynamic social planner's allocation (DSPP). The knowledge spillover allocation is the same as the flexible distribution experiment in Section 5. Figure 10 compares the allocations in the three economies.

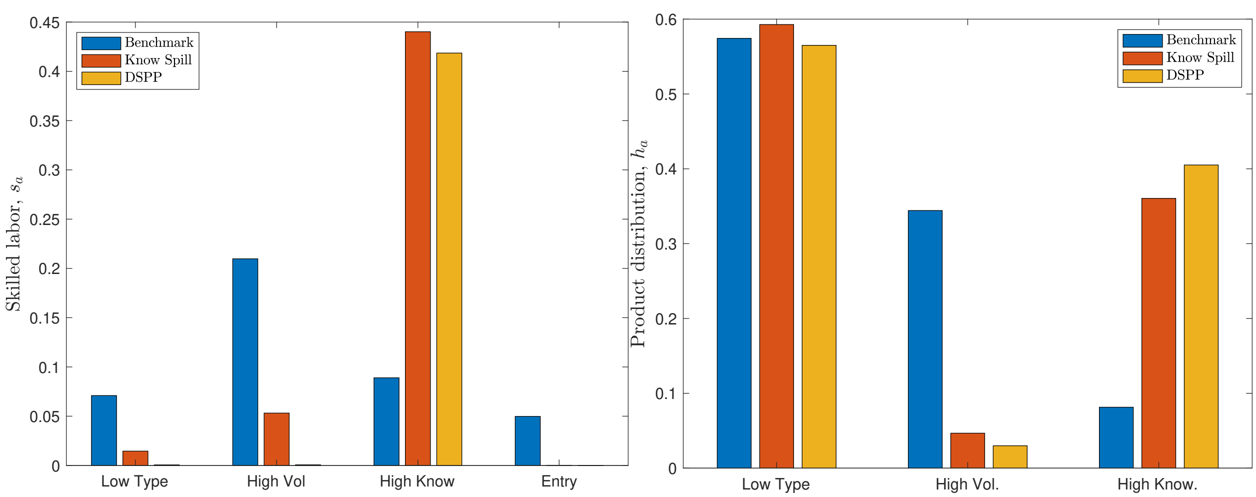
Table 21: Output, Growth, and Welfare with Dynamic SPP Allocation

		Benchmark	Know Spill	DSPP
Output	χ	0.9987	0.9950	0.9988
Growth (%)	g	1.63	3.06	3.17
Chg. in Welfare (%)	ξ	-	104.3	116.2

Notes: Benchmark refers to Benchmark Calibration described in Section 4. Know. Spill. refers to the equilibrium where firm profits are set equal to knowledge spillovers, $\bar{\pi}_a = \ln \bar{\eta}$, as in the main quantitative experiment in Section 5. DSPP refers to the dynamic social planner's problem in (25).

The social planner's allocation increases investment by high knowledge firms further than in the benchmark experiment in order to increase the share of high knowledge products in the steady state.[39] Given the relatively small incremental gain in growth and welfare, setting profits to match knowledge spillovers leads to a reasonable approximation of the socially optimal R&D allocation.

Figure 10: Dynamic SPP Allocations



(a) Labor Allocation s_a (b) Product Distribution h_a

[39] s_K declines relative to the knowledge spillover allocation because s_K is reported per product line. The increase in high knowledge products h_K in the social planner's allocation increases aggregate R&D spending causing s_K to decrease per product line.

The social planner's allocation highlights an important caveat with the R&D wedges θ_a. The wedges highlight the R&D allocation that would maximize aggregate productivity growth for a given distribution of products. However, this is not necessarily the R&D allocation that maximizes growth after accounting for the effects of R&D allocations on the product distribution nor is it the R&D allocation that maximizes welfare. Consequently, short run growth could be increased above the social planner's allocation growth rate by reallocating resources to match knowledge spillovers. However, this would be short lived as the distribution of products would adjust to the new R&D allocations and the economy would converge to the knowledge spillover allocation. The social planner's allocation also implies that small wedges that favor high research productivity innovators may be desirable as it improves the distribution of innovators in the long-run. This is an important caveat as it implies that some degree of R&D misallocation may be desirable.

D.4 Transition Dynamics

Transition dynamics can have important implications for aggregate welfare and the relevant time horizon of policies. I examine the transition dynamics of the model between the early and late period economies examined in Section 5.2. I assume that the profitabilities immediately adjust in period 0 to the late period values. Given that the changes in profitabilities were likely gradual and involved other parameters simultaneously changing, this experiment is meant for illustrative purposes rather than to take a stance on the path of growth over this time period. Figure 11 summarizes the transition path for the growth rate and the product distribution $h_a(t)$.

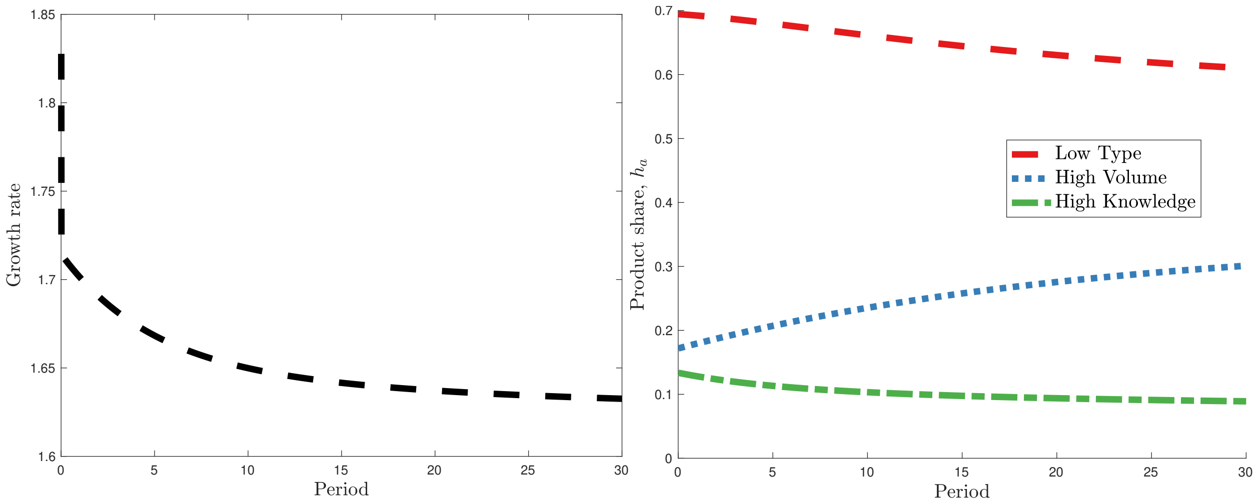
The figures show that growth drops immediately due to the increase in R&D misallocation and then gradually declines as the product distribution adjusts to the new BGP. The majority of the decline occurs within the first 10 years of the transition supporting that changes in profitability over this time period would be observed in the growth statistics.

Algorithm. I calculate the transition path over T periods with length $dt = 0.01$. I start the economy in the initial benchmark BGP equilibrium and assume that the shock occurs in period $t = 0$ and the economy reaches the new BGP after T periods. The algorithm modifies the algorithm in Acemoglu et al. (2016b). The transition path is calculated as:

1. Guess the value function $v_a(t)$ for each t, where $v_a(t)\pi = V_a(\pi, t)$;
2. Solve $\{x_a(t), x_e(t)\}$ sequentially where $x_a(t) = \left[e^{-\rho dt}\psi_e\bar{\pi}_a/\psi_a\zeta\left[\sum_{a'}\bar{\pi}_{a'}\tilde{h}_{a'}\right]\right]^{1/(\zeta-1)}$ and $x_e(t)$ is solved using the resource constraint;[40]

[40] The entry condition is used in the expression for $x_a(t)$ ensuring that it holds along the transition path.

Figure 11: Transition Dynamics



(a) Growth Rate (b) Product Distribution

3. Update the firm-type distribution, $h_a(t+dt) = h_a(x_a(t)-\delta)dt + x_e(t)\tilde{h}_a dt + \sum_{a'} \gamma_{a',a} h_{a'}$;

4. Solve the new value function $\tilde{v}_a(t)$ backwards (from period T to 0) using $v_a(t) = 1 + e^{-\rho dt}\left[v_a(t+dt) - \delta(t)v_a(t+dt)\right]$;[41]

5. Check if the value function has converged, $|\tilde{v}_a(t) - v_a(t)| < 10^{-6}$. If not then set $v_a(t) = \tilde{v}_a(t)$ and repeat from step 2.

[41]The form of $v_a(t)$ assumes that $B_a(t)$ is set to cancel out the option value of innovation. This is a reasonable approximation given that the value of $B_a(t)$ is quantitatively small in the baseline calibration and that $B_a(t)$ would not vary much over the transition path.