Inventors are Rival, Inventions are Not

Jason Hall*

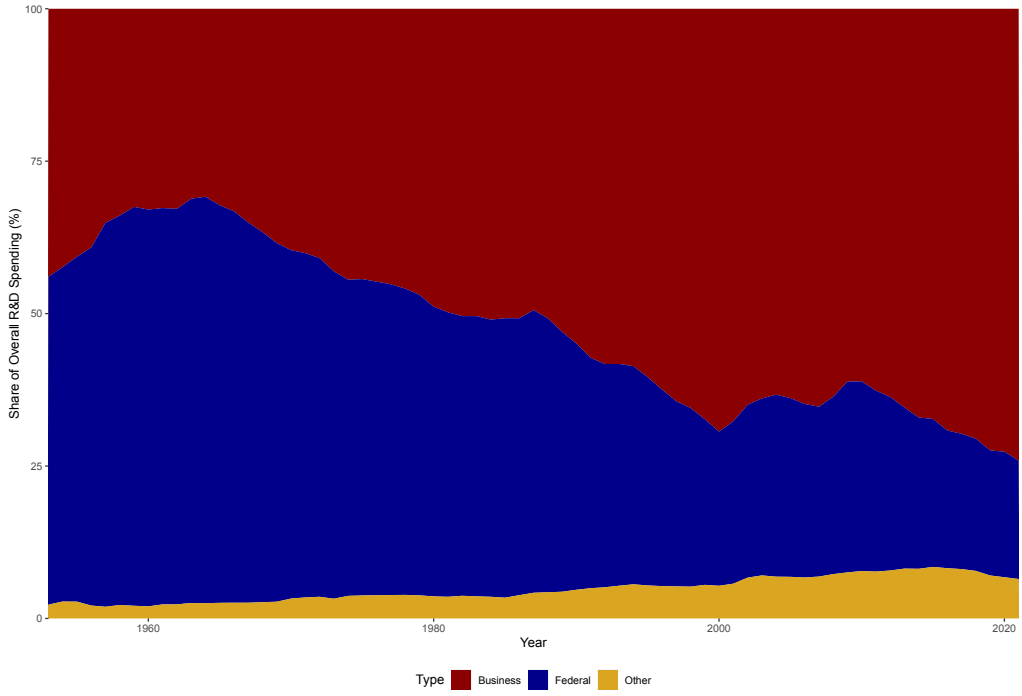
July 1st, 2025

Abstract

I develop a general equilibrium model of dynamic competition where oligopolistic firms compete in product and researcher labor markets and thus internalize their ability to affect competitors' cost of innovation. I calibrate the model using micro evidence on discrepancies between inventive output and wages at small and large firms. I then use the model to compare welfare effects of differential research subsidies for leading and laggard firms. At the calibrated steady state, I find that laggard firms overhire researchers relative to the social optimum. Consequently, tilting R&D subsidies towards leading firms results in welfare gains of 0.25%.

*I thank Marco Bassetto, Honey Batra, Anmol Bhandari, Kyle Herkenhoff, Hannes Malmberg, Simon Mongey, Todd Schoellman, and Kjetil Storesletten for many discussions about this project. I also benefited from discussions with Nicolas Crouzet, Martin Garcia-Vazquez, Fatih Guvenen, Thomas Hasenzagl, William Jungerman, Loukas Karabarbounis, Jane Olmstead-Rumsey, Luis Perez, and Marta Prato, as well as the many students in the Growth and Development and Macro-Labor workshops at the University of Minnesota.

Figure 1: Composition of R&D Spending Over Time



1 Introduction

The average growth rate in the United States from 1947 to 1999 was almost 1.5 percentage points higher than it has been since 2000, whereas R&D expenditure as a percentage of GDP was roughly half a percentage point lower. Popular explanations for this feature of the data include ideas getting harder to find (Bloom et al., 2020) and the rise of intangibles (Crouzet et al., 2022,De Ridder, 2024). This change in levels, however, belies an enormous change in the composition of R&D. Government spending has declined and been replaced by business spending, as shown in Figure 1.

Why might this shift matter? The fundamental reason is that creative destruction comes at the expense of current market leaders. Leading firms have an incentive to reduce the rate of creative destruction even if doing so generates aggregate misallocation of innovative activity. How might they do this? One way is by affecting the allocation of scarce scientific talent to their competitors. Researchers are specialized, generating labor market segmentation. To the extent that a relatively small segment of firms dominate the R&D spend distribution, such labor market segmentation provides market leaders an opportunity. They can over-hire researchers, raising the market wage,

increasing the cost of innovation at competing firms, and thereby prolonging the leader's market power.

In this paper, I develop a general equilibrium model of dynamic competition between leading and laggard firms that incorporates this mechanism. In the model, a continuum of duopolistic industries face upwards sloping researcher labor curves that depend on both their own and their industry competitor's choices of researcher labor. The model is capable of matching micro-level evidence on dispersion in inventor output and wages across firms, but does not impose that this dispersion arises because of anti-competitive actions by leading firms. Rather, the model nests cases where the accumulation of inventors by these firms is strongly anticompetitive and cases where such accumulation is completely benign. The data then determine where we lie on that spectrum.

I discipline the model using micro-level evidence on the correlation between leader R&D expense shares and leader industry sales shares among publicly listed U.S. firms and recent micro-level evidence detailing the discrepancy between innovative output and wages as matched inventors switch to established and new firms. Armed with the calibrated model, I then ask how we should optimally structure R&D subsidies and find that a constrained planner would want to tilt research subsidies *towards* leading firms. The welfare gains of this budget neutral policy are 0.25% in consumption equivalent terms.

I review the literature and proceed as follows. Section II reviews empirical evidence relating to inventors and their output as they move across firms. Section III introduces a novel model of endogenous growth and calibrates the model to match relevant moments in the data. Section IV describes how I estimate the parameters of this model from the data and evaluates the fit. Section V characterizes the calibrated steady state and then considers a set of comparative statics. Section VI contains the main policy exercises of the paper. Section VII concludes.

1.1 Literature

The core methodological innovation of this paper is a model that is able to simultaneously think about dynamic competition in product markets, imperfections in inventor labor markets, and complementarities between the two. Models that incorporate either of these mechanisms on their own are abundant. For instance, the literature on dynamic competition and growth, beginning with Aghion et al. (1997) and Aghion et al. (2001), is typically characterized by models that have dynamic duopoly in product markets but competitive labor markets for research workers. Akcigit and Ates (2021) provide a link between the theoretical predictions of step-by-step innovation models and the observed slowdown in productivity growth, and in follow-up work in Akcigit and Ates (2023) argue that what drives much of the observed changes in business dynamism since 1970 in the United States is a decline in knowledge diffusion between firms. Akcigit et al. (2023) document an empirical relationship between the political connections of incumbent firms and innovation in Italy and build a quantitative growth model that incorporates the incentives of innovating firms to build political connections to affect the cost of regulation.

In contrast to the literature on dynamic oligopoly and growth, another strand of the growth literature examines the importance of inventors themselves and their allocation across firms for innovation and aggregate output growth. These papers abstract from dynamic oligopoly in product markets to more carefully model the choice problems of inventors. Akcigit et al. (2025) augment a relatively standard model of intermediate goods monopolists who invest to improve their productivity with a rich model of the training of innovators featuring innovative teams, time to build, financial frictions and R&D capital and calibrate the model with Danish data on PhD recipients. Prato (2025), similarly, augments a standard model of intermediate monopolists with a search and matching problem for inventors and international trade to understand the role of international migration in shaping the global innovation frontier. Babalievsky (2022) studies the extent of misallocation in the market for inventors by augmenting a firm dynamics model in the vein of Klette and Kortum (2004) with a search and matching model for inventors. Although these papers carefully model the dynamics of inventor labor markets, the structure of product market competition in these

papers severely limits their capacity to understand dynamic strategic considerations embedded in competition between large firms.

My model tractably combines the dynamic considerations of oligopolists from the former strand with the imperfect labor markets of the latter. This merging of the two allows me to make contact with the empirical literature examining inventor output. Most directly, Akcigit and Goldschlag (2023) show that matched inventors who switch to a large, old firm are more productive than their counterparts who switch to young firms. Despite this discrepancy in innovative output, inventors working at incumbents are paid more. My model is capable of reconciling these facts.

The closest paper to mine is contemporaneous work by Fernández-Villaverde et al. (2024). While both papers focus on the extent to which firms can use researcher labor markets to maintain product market power, we differ methodologically and in applications we consider. Where they model competition between incumbent monopolists and potential entrants, I model competition as a dynamic game between incumbent firms. This choice is motivated by Garcia-Macia et al. (2019), who argue that most growth comes from the actions of incumbents. On the labor market side, we have very different models of labor supply with different equilibrium properties; where they have a degenerate researcher wage distribution, my model generates researcher wage and markdown dispersion both within and across industries. My choice is motivated by Akcigit and Goldschlag (2023), who document that inventors who switch to large, old firms are better paid and less productive than otherwise similar inventors who switch to young firms. We also differ in application: while they focus on model-based counterfactuals involving higher switching costs, lower dispersion in entrant R&D efficiency, and various types of research subsidies, my main application uses the calibrated model to study the optimal design of research subsidies.

The core mechanism in my model is essentially raising rivals' cost, and thus bears close similarity to work in industrial organization examining vertical conduct and market power both theoretically (as in e.g., Salop and Scheffman, 1983, Salop and Scheffman, 1987 and Hart et al., 1990), and empirically (as in e.g., Crawford et al., 2018). These papers are focused on individual industries and are thus not concerned with aggregate growth. A more recent literature in macroe-

conomics has focused on on labor market power (e.g., Yeh et al., 2022, Azar et al., 2020, Azar et al., 2022). Berger et al. (2022) derive a welfare relevant metric of labor market concentration and construct a quantitative model to examine the welfare cost of oligopsony in the United States. They find that the welfare effect is large (around 6% in consumption equivalent terms), but that declining concentration has boosted the labor share over time. My contribution is to link theories of labor market power with dynamic considerations in product market competition. To the best of my knowledge, my paper is the first model of dynamic oligopoly in either partial or general equilibrium where firms are strategic in both product markets and the markets for innovation inputs. In the data, these inputs are largely labor. In my model, researchers are the only input to innovation.

2 Empirics

In this section, I construct a matched employee-employer dataset using Patent data and analyze relative changes in inventive output in employees who switch to old as opposed to young firms.

2.1 Data

To evaluate the differential impact of inventors moving to young or old firms, one needs some way of matching inventors to their employer. To do this, I rely on information about assignments of intellectual property rights by individual inventors to firms. In practice, most labor market contracts in the United States include an intellectual property assignment clause. These typically require employees of a firm to assign intellectual property rights developed during their employment and related to their job to their employer. This feature of labor markets facilitates matching inventors to their employers, and allows me to construct matched employee-employer data from patent records and thus examine how an inventor's productivity changes when they move across firms.

I use disambiguated patent data from PatentsView, which is a product of the United States Patent and Trademark Office (USPTO). The vintage of PatentsView that I use consists of the collection of U.S. patents granted since January 1st, 1976 up until September 2023. The key advan-

tage of PatentsView over the Bulk XML files is that PatentsView uses frontier machine learning techniques to disambiguate inventors and firms. Broadly, the algorithms work by first generating grouping of likely interrelated authors (or firms) called *canopies*. Within a canopy, clustering algorithms are then used to identify similar individuals. Individuals (or firms) who are sufficiently similar are identified as one individual (or firm). For a more detailed description of this algorithm, see here. This innovation on the part of PatentsView facilitates inventor-level - as opposed to more traditional firm-level - analyses of innovative output. This is crucial to my mechanism, which emphasizes how the productivity of individual inventors changes when they move between leading and laggard firms.

I augment PatentsView with data from Kogan et al. (2017), which the authors have updated to include patents granted up to 2022. Importantly, this paper included an effort to match patents assigned to subsidiaries of public firms with their parent company. This enables me to better match inventors with their true employer, as opposed to firms created merely as receptacles for intellectual property holdings. To recover firm age, I use data from https://foundingpatents.com. This file contains founding years for roughly 85% of patents and 73% of assignees.

To construct my analysis sample, I begin by restricting the PatentsView data to utility patents, and then further restrict myself to patents associated with inventors who are listed on the patent as present in the United States and employed by a private firm.[1] This set of inventors associated with those patents provides my sample frame. To construct this inventor-level sample, I must map individual inventors to their employers via their assignee. The PatentsView disambiguation provides one set of matched firms, but is unfortunately incomplete. I thus augment this matching using data from Kogan et al. (2017), which matches subsidiaries to their parent firms, and an additional extensive manual search based on the name of the firm.

I assign an inventor to an employer by aggregating the data to the yearly level at the file year and taking the most common assignee within that file year.[2] Note, however, that all observations are at

[1] I do not include patents assigned to universities and thus my sample excludes researchers who only assign their intellectual property rights to a university.

[2] For example, suppose that inventor Jane Doe invents two patents that are ultimately assigned to firm Widgets Incorporated and one to Sprockets LLC in the year 2004. In this case, I would assign Jane Doe's employer as Widgets

the inventor level, so while I assign one firm as an inventor's employer, their innovative output is measured in terms of citations to all of their patents, irrespective of to whom those property rights are assigned.

2.2 Empirical Exercise

Having described the construction of the dataset, I now turn to a description of my empirical exercise. The object of interest is an average treatment effect on the treated (ATT) where the treatment is the effect of switching to a young firm relative to an old firm. The ideal experiment would be one in which inventors of identical ability were randomly assigned to firms of different ages. In this case, the difference in mean inventive output (which I will take to be citations) after treatment between individuals randomly assigned to different firms would identify the average treatment effect. As is immediately apparent, this ideal environment is very different from the one that prevails in reality. Most importantly, one would expect that inventors are not randomized across young and old firms: inventors who expect they have a good idea would be prone to start their own firm. This would imply that the result of the prior exercise would be to bias the estimated treatment effect downward, so that the estimated impact of working at a large firm is more negative than it truly is. To address this concern even partially, one must proxy for the unobserved ability of inventors at different firm types. I do this using the synthetic difference in differences technique outlined in Arkhangelsky et al. (2021). Synthetic difference-in-differences works in essentially two steps. In the first step, one finds a set of weights $\hat{\omega}_i$ that align pre-exposure trends in untreated units with those of treated units, and time weights $\hat{\lambda_t}$ to balance pre- and post-exposure time periods. Then, taking as given those weights, one estimates a treatment effect (denoted here by τ) by finding a vector of parameters $(\hat{\tau}, \hat{\mu}, \hat{\alpha}, \hat{\beta})$, where μ is an intercept, α is a vector of unit fixed effects, and β is a vector of time fixed effects, which solves

$$\arg \min_{\tau,\mu,\alpha,\beta} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{it} - \mu - \alpha_i - \beta_t - \mathbf{1}_{it}[\text{Treated}]\tau)^2 \hat{\omega}_i \hat{\lambda_t} \right\} \tag{1}$$

Incorporated.

The advantage of synthetic difference-in-differences techniques in this context is that it allows me to control for potential confounding due to unobserved and time invariant heterogeneity in inventor skill in the period around the switch. It does not, however, allow me to eliminate bias in the estimated average treatment effect due to, for example, private knowledge about the value of an idea on the part of inventors, and so these results should not be interepreted as causal. Rather, they should viewed merely as differences in innovative output after trying to net out some of the underlying differences between the two groups.

To estimate these effects, I construct a balanced panel of individuals who switch to old and young firms from the dataset described above by stacking inventor switches relative to the year of switch. This means that the output of an inventor who switches in 2000 is directly compared to the output of an inventor who switches in 2010. Following Hall et al. (2005) I use the number of citations as a proxy of patent quality, and define the innovative output of an inventor as the number of citations by patents that are ultimately assigned entities other than the assignee of the original patent within 5 years. I will refer to these as external forward citations. This quantity will mechanically be biased downwards because older firms will - all else equal - have a larger share of the existing patent stock, so I also consider raw 5-year citations (i.e. including patents assigned to the same firm). The results of this exercise are displayed in Figure 2, which demonstrates that individuals who switch to younger firms have systematically higher deviations in their trend for innovative output than their counterparts who switch to older firms.[3]

One potential concern is that the nature of both patents and citations have changed over time, so that the value of a citation in 2010 is different from that of one in 2000. To address this concern, I residualize the outcome variables at the year level and then estimate the synthetic difference-in-differences again. Figure 3 plots the result of the exercise. Consistent with the raw synthetic difference-in-differences approach, the effect is estimated to be 2 citations.

[3]Note that synthetic difference-in-differences allows for heterogeneous levels among treated and non-treated units. This would correspond in my context to systematic differences in average inventive ability for individuals who switch to young firms vis-à-vis old firms. Because these unit specific fixed effects are time invariant, the SDID will not use them to estimate the reduced form effect, and thus for ease of comparison I have constructed the plots so that the overlay reflects only the variation that identifies the reduced form effect.

Figure 2: Inventive Output at For Different Switches

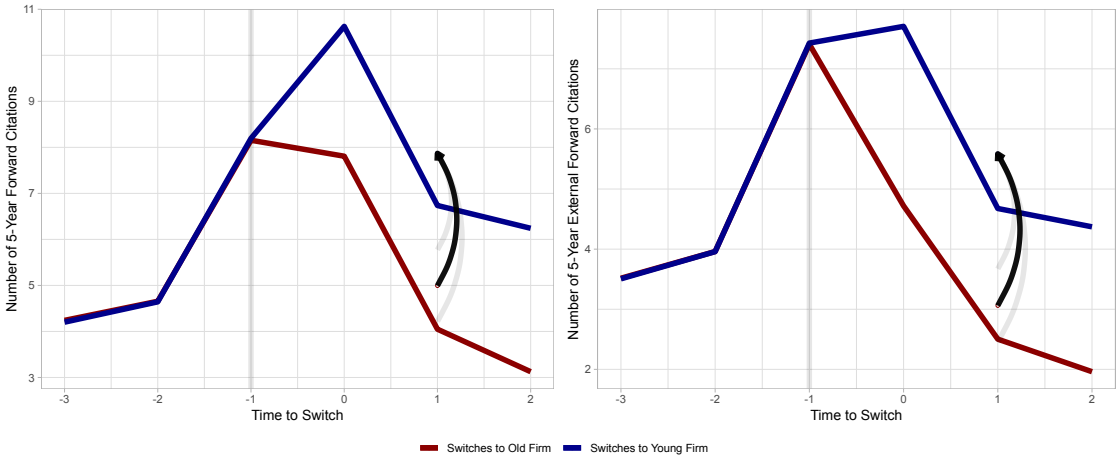


Figure 3: Inventive Output: Year Fixed Effects

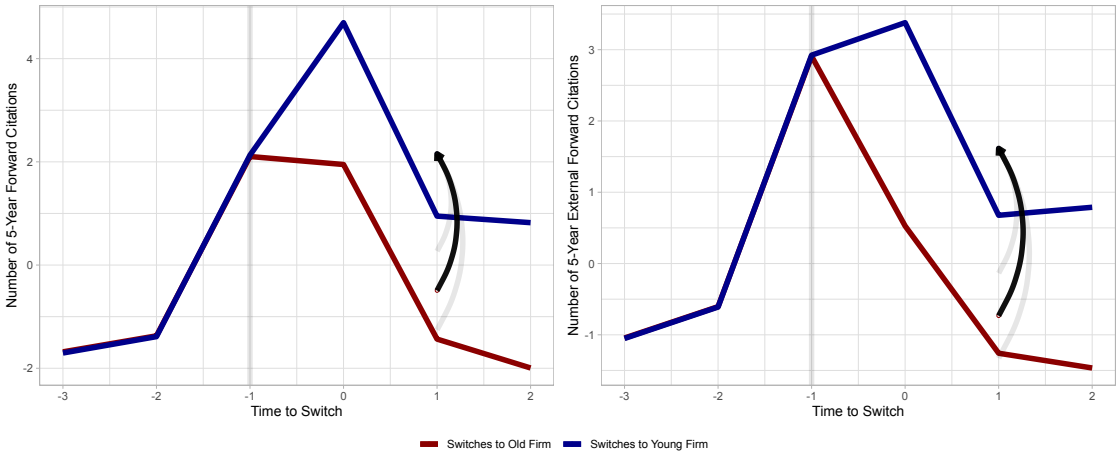
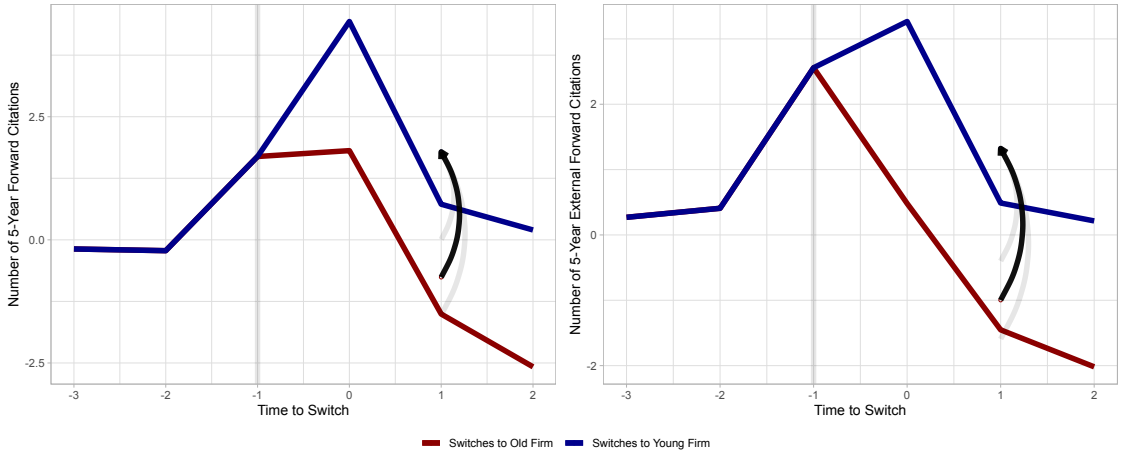


Figure 4: Inventive Output: Age Fixed Effects



Dependent Variable:	5-Year Forward Citations			5-Year Outside Forward Citations		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
τ	2.875***	2.544***	2.649***	2.524***	2.320***	2.347***
	(0.778)	(0.766)	(0.763)	(0.611)	(0.597)	(0.596)
Fixed-effects						
Proxied Age	No	Yes	No	No	Yes	No
Calendar Year	No	No	Yes	No	No	Yes

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 1: Synthetic Difference-in-Differences result. All standard errors are calculated using the jackknife approach.

One might worry that younger inventors are more likely to move to start-ups. If the individual innovation production function exhibits decreasing returns to scale in individual human capital, then the estimated impact of switching to an older firm would again be biased downward, as we would attribute the higher incremental return from human capital accumulation to the effect of working at a young firm. To address this concern, I residualize the outcome variables against the number of years since the first patent filed by that inventor as a proxy for age, and repeat the exercise. The results, which are qualitatively unchanged, are depicted in Figure 4. Table 1 summarizes the results of all exercises.

These results partially replicate the findings of Akcigit and Goldschlag (2023), who also find

that in administrative data covering the majority of the American populace, while individuals at young firms produce relatively more innovative output, they are paid relatively less.

This is a puzzling result from the lens of theory, which suggests that wages are a function of one's marginal product and that larger firms should have relatively more market power vis-à-vis workers. These would together imply that inventors who produce fewer innovations adjusting for quality should be paid less at older firms (which are mechanically larger than younger firms) in the sense that their marginal product is lower and their employer has more labor market power. That we see the opposite pattern in the data suggests that the marginal private value of a worker is higher at an older firm vis-à-vis a younger firm due to reasons other than the inventor's innovative output, at least as measured by citations and various other measures of innovative output. This exercise alone, however, cannot speak to the source of those differences. One potential reason is compensation for market power - incumbent firms pay higher wages because they want to deny laggard firms access to a crucial input for firm growth - researchers. Alternatively, however, these differences in wages could simply reflect persistent differences in the capacity of firms to commercialize these innovations. A model is needed to distinguish between these two competing hypotheses, and I turn to describing such a model now.

3 Model

3.1 Environment

Time is continuous and indexed by $t \in [0, \infty)$. There are a continuum of industries, indexed by j. Within each industry, there are I_j firms indexed by i, so that firm i in industry j is denoted as ij. Each firm i does two things. Firstly, they produce a good that is imperfectly-substitutable with the good produced by other firms in the same industry. The substitutability between these goods is characterized by a parameter η, corresponding to an elasticity of substitution $\frac{1}{1-\eta}$. The production of each of these goods $y_{ij,t}$ is done using production labor, denoted by $l_{ij,t}$, that is paid a wage rate

$\bar{w}_t$. A firm's production function is thus given by:

$$y_{ij,t} = z_{ij,t} l_{ij,t} \tag{2}$$

Firms compete dynamically by hiring researchers to improve their productivity. Hiring a quantity of researchers $x_{ij,t}$ who are paid $w_{ij,t}$, generates innovation arrival rate $a_{ij,t}$ according to the innovation production function:

$$a_{ij,t} = \epsilon x_{ij,t}^{\gamma} \tag{3}$$

The arrivals of an innovation generate different improvements in productivity across firms. The law of motion for the productivity process of the leader, conditional on an innovation is given by:

$$z_{ij,t+\Delta t} = \lambda z_{ij,t} \tag{4}$$

Following Akcigit and Ates (2023), I allow the possibility of quick-catchup: if an innovation arrives, with some probability $1 - \phi$ the laggard's productivity jumps to that of the leader, erasing the producitivty difference. In the other case, the productivity increases only by a step-size λ. This is reflected in the following law of motion:

$$z_{-ij,t+\Delta t} = \begin{cases} \lambda z_{-ij,t} & \text{with probability } \phi \\ z_{ij,t} & \text{with probability } 1 - \phi \end{cases} \tag{5}$$

There is a competitive final goods producer that aggregates intermediates using a Cobb-Douglas with unit elasticity:

$$Y_t = \exp\left\{ \int_0^1 \frac{1}{\eta} \ln\left(\sum_{i \in I_j} y_{ij,t}^{\eta} \right) dj \right\} \tag{6}$$

There is a representative household. This representative household values aggregate consumption C_t, supplies a quantity of production labor $\bar{L}$ inelastically, and then supplies an aggregate

bundle of research labor X_t. The representative household's flow preferences are thus given by:

$$U(C_t, X_t) = \ln(C_t) - \kappa X_t \tag{7}$$

While porduction labor is homgoeneous and supplied inelastically, the research labor at different firms and in different industries is differentiated. Following Berger et al. (2022), the aggregate bundle of research labor is given by:

$$X_t = \left(\int_0^1 x_{jt}^{\frac{\theta+1}{\theta}} dj \right)^{\frac{\theta}{\theta+1}} \tag{8}$$

Where

$$x_{jt} = \left(\sum_{i \in I_j} x_{ij,t}^{\frac{\psi+1}{\psi}} \right)^{\frac{\psi}{\psi+1}} \tag{9}$$

ψ controls the elasticity of substitution within an industry (i.e. between i and i') is given by ψ, while θ controls the elasticity of substitution across industries (i.e. between j and j').

Market clearing in this environment consists of four conditions. The first is an aggregate condition that production of the final good be equal to consumption.

$$C_t = Y_t \tag{10}$$

The second is the labor supply condition for the research labor condition. Letting $x_{ij,t}^S$ be the quantity of labor supplied by the household and $x_{ij,t}^D$ as the derived demand of the firm for research labor, the market clearing condition is given by:

$$\forall i \; \forall j \; x_{ij,t}^S = x_{ij,t}^D \tag{11}$$

The third is the labor supply condition for the aggregate production labor.

$$\int_0^1 \sum_{i \in I_j} l_{ij,t} \, dj = \bar{L} \equiv 1 \tag{12}$$

The final market clearing condition is the asset market clearing condition, which says that the collective value of savings in the economy must be equal to the aggregate market value of the firms. Denoting the value of firm i in industry j by $V_{ij,t}$ this condition can be written as

$$\int_0^1 \sum_{i \in I_j} V_{ij,t} \, dj = A_t \tag{13}$$

This completes the characterization of preferences, endowments, and technology. I will delay the definition of an equilibrium until I characterize the problem of the intermediates firms, who form the centerpiece of this model.

3.2 Household Problem

There is no aggregate uncertainty, and the representative household holds all shares of all firms in the economy. The household problem is thus given by:

$$\max_{C_t, \{\{x_{ij,t}\}_{i \in I_j}\}_{j \in [0,1]}} \int_0^\infty e^{-\rho t} \left[\ln(C_t) - \kappa X_t \right] dt \tag{14}$$

Subject to:

$$\dot{A}_t = r_t A_t + \bar{w}_t \bar{L} + \int_0^1 \sum_{i \in I_j} w_{ij,t} x_{ij,t} dj + T_t - C_t \tag{15}$$

(8)

(9)

In Appendix A.2, I derive the first order conditions that characterize the solution to this problem. One is the standard household Euler condition:

$$r_t = \rho + \frac{\dot{C}_t}{C_t} \tag{16}$$

The second is the residual labor supply condition for research labor at firm i in industry j:

$$w_{ij,t} = \kappa C_t \left(\frac{x_{jt}}{X_t}\right)^{\frac{1}{\theta}} \left(\frac{x_{ij,t}}{x_{j,t}}\right)^{\frac{1}{\psi}} \tag{17}$$

3.3 The Government

In anticipation of my policy exercises, I introduce a government to this problem along the lines of Akcigit and Ates (2023). This government has the capacity to tax gross profits of firms at a tax rate τ_π and can subsidize the research expenditures of leader and laggard firms at rates τ_{sl} and τ_{sf}. To balance the government budget at every instant, they levy a lump sum tax T_t on the household. As a consequence of this, the government budget constraint at every t is given by:

$$\begin{aligned} T_t = &\int\int 2\tau_\pi(\pi(0)) - \tau_s(w(x_l, x_f; X)x_l + w(x_f, x_l; X)x_f)\mu(0, x_l, x_f)dx_l dx_f + \\ &+ \sum_{m=1}^{\bar{m}} \int\int \tau_\pi(\pi(m) + \pi(-m)) - \tau_s(w(x_l, x_f; X)x_l + w(x_f, x_l; X)x_f)\mu(m, x_l, x_f)dx_l dx_f \end{aligned} \tag{18}$$

3.4 Firms

The problem of the final goods aggregator is given by:

$$\max_{y_{ij,t} k \in I_j, j \in [0,1]} P_t \exp\left\{\int_0^1 \frac{1}{\eta} \ln\left(\sum_{i \in I_j} y_{ij,t}^\eta\right) dj\right\} - \int_0^1 \sum_{i \in I_j} p_{kj,t} y_{kj,t} dj$$

This gives rise to the standard iso-elastic demand:

$$y_{kj,t} = Y_t \left(\frac{p_{kj,t}^{\frac{1}{\eta-1}}}{\sum_{i \in I_j} p_{ij,t}^{\frac{\eta}{\eta-1}}} \right) \tag{19}$$

Since productivity increases by a discrete step-size λ when an innovation occurs, we can characterize the relative difference in productivity between firms in terms of the difference in the number of quality levels that they have progressed, i.e. we can write $m = N_{ij,t} - N_{-ij,t}$, where $N_{ij,t}$ is a random variable denoting the number of quality levels that a firm has progressed through up to time t. I show in the appendix that this is a sufficient statistic to characterize the flow profits accruing to each firm, and that the equilibrium static profits of each intermediate firm are given by

$$\pi(m) = Y_t \left(\frac{(1-\eta) s_{ij,t}}{1 - \eta s_{ij,t}} \right) \tag{20}$$

Where $s_{kj,t}$ is the share of firm k in sector j, so that

$$s_{kj,t} = \frac{p_{kj,t} y_{kj,t}}{Y_t} = \frac{p_{kj,t}^{\frac{\eta}{\eta-1}}}{\sum_{i \in I_j} p_{ij,t}^{\frac{\eta}{\eta-1}}}$$

This result enables me to separate an intermediate's firm static payoff from production from the dynamic payoff due to investment, as in Ericson and Pakes (1995) and the work that followed. Denote the equilibrium choice of production labor by the leading and laggard firm by $l^*(m)$ and $l^*(-m)$. As in this literature, I restrict consideration to symmetric and anonymous Markov Perfect Equilibria.[4] In addition, I require two assumptions that together render the equilibrium computable.

Assumption 1: There exists an upper bound $\bar{m}$ on the productivity difference between the two firms. In the case when a firm at the upper bound innovates, the productivity difference does not change.

[4]Symmetry in this context is the restriction that each firm, if their states were switched, would follow the strategy that their rival had ex-ante. Anonymity, on the other hand, refers to the restriction that firms would not prefer to change their strategy in the event that states of their rivals were permuted. In games with two players, as in my setting, the latter restriction has no bite, as the only admissible permutation is the identity.

Assumption 2: For all industries j, $I_j = 2$.

These assumptions are needed to render the equilibrium computationally tractable. I make one additional assumption about the structure of the oligopoly game.

Assumption 3: Employment quantities $x_{ij,t}$ are "sticky" in the sense that a firm is committed to a quantity of labor $x_{ij,t}$ until receiving an adjustment shock, which is distributed Poisson with rate parameter ν.

This assumption merits discussion. The observed difference in log wages across firms suggest that strategic considerations on the part of firms as to their competing firms' choices are potentially an important determinant of inventor wages. Conventional simultaneous move games obviate this this possibility in the sense that firms engaging in simultaneous move games take each others' best responses as given. The natural alternative to this is a hierarchical game (as in e.g. Von Stackelberg, 1934), but in the dynamic context this suffers from a pronounced concern that market behavior is determined by ad-hoc assumptions as to the order of moves. To avoid such ad-hoc assumptions, I draw on insights from Maskin and Tirole (1988), who demonstrate that leader follower-dynamics are generated endogenously in dynamic games when players have sufficient commitment power. Stochastic adjustment shocks provide such commitment power. By hiring large quantities of researchers today without certainty about their ability to fire them tomorrow, leading firms can signal to laggard firms that if the latter hires large quantities today, they will face large costs for a potentially long period of time. This has the potential to induce laggards to reduce their spending on researchers. Put in the language of Jun and Vives (2004), these frictions provide firms the capacity to influence future market outcomes.

The value function of firm i in industry j is characterized by four states. The first is whether or not a firm is the leader or the laggard. The second is the productivity difference between the two firms m.[5] The third is the quantity of research input that the leader is currently committed to. The

[5]For notational convenience I will follow the literature and summarize both the productivity difference and the

fourth is the quantity of research input that the laggard is committed to.[6]

Industries are also exposed to death shocks at a rate δ. These exogenous shocks are distributed Poisson with rate parameter δ. When a death shock arrives, the laggard firm dies and is replaced with an exact replica. In the case when the industry is leveled, the identity of the firm that dies when the death shock arrives is decided by coin flip. Given this environment, the value function of a leading firm when committed is given by:

$$
\begin{aligned}
r_t V_t(m, x_l, x_f) - \dot{V}_t = & \left[(1-\tau_\pi)\pi(m) - (1-\tau_{sl})w(x_l, x_f; X)x_l\right] + \\
& \underbrace{\epsilon x_l^\gamma \left[V_t(m+1, x_l, x_f) - V_t(m, x_l, x_f)\right]}_{\text{Leader improves their productivity}} + \\
& \underbrace{\epsilon x_f^\gamma \left[\phi V_t(m-1, x_l, x_f) + (1-\phi)V_t(0, x_l, x_f) - V_t(m, x_l, x_f)\right]}_{\text{Laggard improves their productivity}} + \\
& \underbrace{\nu \left[W_t(m, x_f) - V_t(m, x_l, x_f)\right]}_{\text{Leader can change their quantity of researchers}} + \\
& \underbrace{\nu \left[V_t(m, x_l, g(-m, x_l)) - V_t(m, x_l, x_f)\right]}_{\text{Laggard can change their quantity of researchers}}
\end{aligned}
$$

Where W is the value function of the firm when its research quantity is freely adjustable (i.e. when it receives an adjustment shock) and the productivity gap is m. This function is given by:

identity of the leader in one term. A positive value of m indicates the firm is leading by m steps, while a value of $-m$ indicates that a firm is behind the leader by m steps.

[6]Note that here I refer to the firm with the higher productivity as the "leader" while the firm with the lower productivity is the "laggard." In the case where the firms are tied, the names given to each firm are arbitrary because of the symmetry assumption.

$$
\begin{aligned}
r_t W_t(m, x_f) - \dot{W}_t = \max_x \left[(1-\tau_\pi)\pi(m) - (1-\tau_{sl})w(x, x_f; X)x\right] + \\
\epsilon x^\gamma \left[V_t(m+1, x, x_f) - W_t(m, x_f)\right] + \\
\epsilon x_f^\gamma \left[\phi V_t(m-1, x, x_f) + (1-\phi)V_t(0, x, x_f) - W_t(m, x_f)\right] + \\
\nu \left[V_t(m, x, g(-m, x)) - W_t(m, x_f)\right]
\end{aligned}
$$

The first line in each of the values corresponds to the flow value of a firm due to its gross profits and research expenditure. The second line corresponds with the change in value to the leader that is due to the arrival of an innovation. Because this firm is the leader, every innovation is incremental, so the gap increases only by one quality point. The third line corresponds to the change in value due to the arrival of an innovation by the laggard. With probability ϕ, such an innovation is incremental, so that the productivity gap decreases by a single point, while with probability $1-\phi$ the innovation is drastic, and the industry becomes leveled. The third and fourth lines correspond with adjustments. The third line corresponds to the change in value that occurs when the leader receives an adjustment shock, while the fourth line corresponds with an adjustment by the laggard.[7] Note also that when the input of the firm is freely adjustable, the only adjustment that appears is that of the rival. This is because, conditional on the industry state (bar the firm's choice of research labor) not changing, the firm would not want to adjust its quantity of researchers again.

The problem of a laggard differs from the leaders' in two ways. Firstly, unlike the leader, every innovation by the laggard has the possibility to induce quick catch-up, which reduces the productivity gap to 0. This asymmetry increases the value of research for laggard firms relative to leading firms. Secondly, laggards face a constant risk of dying and being replaced by an identical firm. These exogenous death shocks are distributed Poisson with rate parameter δ.[8] Given these

[7] I relegate the value functions for the special cases when the industry productivity difference is $\bar{m}$ and 0 to the appendix.

[8] In cases where the industry is leveled, only one firm (chosen at random with equal probability) dies when the death shock arrives.

differences, the laggard's value function when committed is given by:

$$\begin{aligned} r_t V_t(-m, x_l, x_f) - \dot{V}_t = & \left[(1-\tau_\pi)\pi(-m) - (1-\tau_{sf})w(x_f, x_l; X)x_f\right] + \\ & \epsilon x_l^\gamma \left[V_t(-m-1, x_l, x_f) - V_t(-m, x_l, x_f)\right] + \\ & \epsilon x_f^\gamma \left[\phi V_t(-m+1, x_l, x_f) + (1-\phi)V_t(0, x_l, x_f) - V_t(-m, x_l, x_f)\right] + \\ & \nu \left[V_t(-m, g(m, x_f), x_f) - V_t(-m, x_l, x_f)\right] + \\ & \nu \left[W_t(-m, x_l) - V_t(-m, x_l, x_f)\right] + \\ & \delta \left[0 - V_t(-m, x_l, x_f)\right] \end{aligned}$$

Their value function, conditional on adjustment shock is given by:

$$\begin{aligned} r_t W_t(-m, x_l) - \dot{W}_t = & \max_x \left[(1-\tau_\pi)\pi(-m) - (1-\tau_{sf})w(x, x_l; X)x\right] + \\ & \epsilon x_l^\gamma \left[V_t(-m-1, x_l, x) - W_t(-m, x_l)\right] + \\ & \epsilon x^\gamma \left[\phi V_t(-m+1, x_l, x) + (1-\phi)V_t(0, x_l, x) - W_t(-m, x_l)\right] + \\ & \nu \left[V_t(-m, g(m, x), x) - W_t(-m, x_l)\right] + \\ & \delta \left[0 - W_t(-m, x_l)\right] \end{aligned}$$

In steady state, the time derivative of the value functions is zero, and the leader's optimality condition is given by:

$$(1-\tau_{sl})w(x,x_f;X)\left[\frac{\partial \ln(w(x,x_f;X))}{\partial \ln(x)}+1\right]=\underbrace{\nu\left[\frac{\partial V_t(m,x,g(-m,x))}{\partial x}+\frac{\partial V_t(m,x,g(-m,x)}{\partial g(-m,x)}\frac{\partial g(-m,x)}{\partial x}\right]}_{\text{Response term}}$$
$$+\underbrace{\gamma\epsilon x^{\gamma-1}\left[V_t(m+1,x,x_f)-W_t(m,x_f)\right]}_{\text{Research return}}$$
$$+\underbrace{\epsilon x^{\gamma}\frac{\partial V_t(m+1,x,x_f)}{\partial x}}_{\text{Avoidance term}}$$
$$+\underbrace{\epsilon(x_f)^{\gamma}\left[\phi\frac{\partial V_t(m-1,x,x_f)}{\partial x}+(1-\phi)\frac{\partial V_t(0,x,x_f)}{\partial x}\right]}_{\text{Hoarding term}}$$

The laggard's optimality condition is given by:

$$(1-\tau_{sf})w(x,x_l;X)\left[\frac{\partial \ln(w(x,x_l;X))}{\partial \ln(x)}+1\right]=\underbrace{\nu\left[\frac{\partial V_t(-m,g(m,x),x)}{\partial g(m,x)}\frac{\partial g(m,x)}{\partial x}+\frac{\partial V_t(-m,g(m,x),x)}{\partial x}\right]}_{\text{Response term}}$$
$$+\underbrace{\gamma\epsilon x^{\gamma-1}\left[\phi V_t(-m+1,x_l,x)+(1-\phi)V_t(0,x_l,x)-W_t(-m,x_l)\right]}_{\text{Research return}}$$
$$+\underbrace{\epsilon x^{\gamma}\left\{\phi\frac{\partial V_t(-m+1,x_l,x)}{\partial x}+(1-\phi)\frac{\partial V_t(0,x_l,x)}{\partial x}\right\}}_{\text{Avoidance Term}}$$
$$+\underbrace{\epsilon x_l^{\gamma}\left[\frac{\partial V_t(-m-1,x_l,x)}{\partial x}\right]}_{\text{Hoarding Term}}$$

The left hand side of both conditions is the standard oligopsony condition, as in Berger et al. (2022), bar the fact that this elasticity is calculated at the currently committed value of the other's choice of quantity of researchers. The right-hand side is divided into four separate pieces. The first corresponds to changes in the value of the firm due to adjustment by its competitor. This has a direct component, which corresponds to wage changes due to the other firm's choice, and

an indirect component, which reflects the firm's capacity to shape its competitor's choice. The second term is the marginal increase in arrival rate times the value of that increase in arrival rate, increasing the productivity gap by one step (or, in the case of the laggard, reducing it either 0 or by one step). The third and fourth terms correspond to the fact that the firm is not able to readjust its' quantity immediately when the productivity difference m changes. This provides either an additional incentive to hire researchers today, if for instance the leading firm would like to accumulate more researchers as it builds up its lead, or an additional disincentive to avoid hiring researchers, if the value of a researcher when the gap changes is relatively smaller.

3.5 The Balanced Growth Path

In order to characterize the balanced growth path and the aggregate growth rate g, I must first detrend the value functions. In the appendix, I show that any function satisfying the system of Hamilton-Jacobi-Bellman equations in levels must also satisfy the normalized problems given by $\mathcal{V}_t(m, x_l, x_f) = V_t(m, x_l, x_f)/Y_t$ and analogously for W. For notational convenience, I define $\hat{\pi} = \pi/Y_t$ and $\hat{w} = w/Y_t$. These terms are constant along the balanced growth path.

$$\begin{aligned}
\rho\mathcal{V}(m, x_l, x_f) = &[(1-\tau_\pi)\hat{\pi}(m) - (1-\tau_{sl})\hat{w}(x_l, x_f; X)x_l] + \\
&\epsilon x_l^\gamma \left[\mathcal{V}(m+1, x_l, x_f) - \mathcal{V}(m, x_l, x_f)\right] + \\
&\epsilon x_f^\gamma \left[\phi\mathcal{V}(m-1, x_l, x_f) + (1-\phi)\mathcal{V}(0, x_l, x_f) - \mathcal{V}(m, x_l, x_f)\right] + \\
&\nu \left[\mathcal{W}(m, x_f) - \mathcal{V}(m, x_l, x_f)\right] + \\
&\nu \left[\mathcal{V}(m, x_l, g(-m, x_l)) - \mathcal{V}(m, x_l, x_f)\right] +
\end{aligned} \tag{21}$$

$$\begin{aligned}\rho\mathcal{W}(m,x_f) = \max_x \left[(1-\tau_\pi)\hat{\pi}(m) - (1-\tau_{sl})\hat{w}(x,x_f;X)x\right] + \\ \epsilon x^\gamma \left[\mathcal{V}(m+1,x,x_f) - \mathcal{W}(m,x_f)\right] + \\ \epsilon x_f^\gamma \left[\phi\mathcal{V}(m-1,x,x_f) + (1-\phi)\mathcal{V}(0,x,x_f) - \mathcal{W}(m,x_f)\right] + \\ \nu\left[\mathcal{V}(m,x,g(-m,x)) - \mathcal{W}(m,x_f)\right]\end{aligned} \tag{22}$$

$$\begin{aligned}\rho\mathcal{V}(-m,x_l,x_f) = \left[(1-\tau_\pi)\hat{\pi}(-m) - (1-\tau_{sf})\hat{w}(x_f,x_l;X)x_f\right] + \\ \epsilon x_l^\gamma \left[\mathcal{V}(-m-1,x_l,x_f) - \mathcal{V}(-m,x_l,x_f)\right] + \\ \epsilon x_f^\gamma \left[\phi\mathcal{V}(-m+1,x_l,x_f) + (1-\phi)\mathcal{V}(0,x_l,x_f) - \mathcal{V}(-m,x_l,x_f)\right] + \\ \nu\left[\mathcal{V}(-m,g(m,x_f),x_f) - \mathcal{V}(-m,x_l,x_f)\right] + \\ \nu\left[\mathcal{W}(-m,x_l) - \mathcal{V}(-m,x_l,x_f)\right] + \\ \delta\left[0 - \mathcal{V}(-m,x_l,x_f)\right]\end{aligned} \tag{23}$$

$$\begin{aligned}\rho\mathcal{W}(-m,x_l) = \max_x \left[(1-\tau_\pi)\hat{\pi}(-m) - (1-\tau_{sf})\hat{w}(x,x_l;X)x\right] + \\ \epsilon x_l^\gamma \left[\mathcal{V}(-m-1,x_l,x) - \mathcal{W}(-m,x_l)\right] + \\ \epsilon x^\gamma \left[\phi\mathcal{V}(-m+1,x_l,x) + (1-\phi)\mathcal{V}(0,x_l,x) - \mathcal{W}(-m,x_l)\right] + \\ \nu\left[\mathcal{V}(-m,g(m,x),x) - \mathcal{W}(-m,x_l)\right] + \\ \delta\left[0 - \mathcal{W}(-m,x_l)\right]\end{aligned} \tag{24}$$

The last step prior to defining a balanced growth path is to characterize aggregates in this economy.[9] Firstly, aggregate productivity Z_t is given by the following expression:

$$Z_t \equiv \exp\left\{\int_0^1 \ln(z_{ij,t})\, dj\right\} \tag{25}$$

[9]All proofs are available in Appendix A.6.

Where $z_{ij,t} = \max_k\{z_{kj,t}\}$. In addition, define the production labor share $\omega_t = \bar{w}_t/Y_t$. Letting $s(m)$ denote the share of the leading firm in the equilibrium of the static differentiated Bertrand game and $\mu(m)$ the mass of industries with productivity gap m,[10] aggregate output Y_t can be expressed as the following:

$$Y_t = \frac{Z_t}{\omega_t} \exp\left\{\frac{1}{\eta}\sum_{m=0}^{\bar{m}} \ln\left[\left(\frac{\eta s(m)(1-s(m))}{1-\eta s(m)}\right)^{\eta} + \left(\lambda^{-m}\frac{\eta s(m)(1-s(m))}{1-\eta(1-s(m)}\right)^{\eta}\right]\mu(m)\right\} \tag{26}$$

This equation can be divided into three parts. The first term is the frontier productivity term, which is the aggregate TFP term in this economy. The second is the production labor share, while the third is a term that depends on the distribution of industries across productivity gaps. Along the balanced growth path, these latter two terms are constant, and thus we can characterize the aggregate growth rate solely through the growth rate of the frontier productivity term. This growth rate is endogenously determined by the innovation choices of firms, and in particular leading firms. Hence we can characterize the growth rate of Z_t as:

$$\ln(\dot{Z}_t) = \ln(\lambda)\left[\int\int \left(\epsilon x_l^{\gamma} + \epsilon x_f^{\gamma}\right)\ \mu(0, x_l, x_f)dx_l dx_f + \sum_{m=1}^{\bar{m}}\int\int \epsilon x_l^{\gamma}\ \mu(m, x_l, x_f)dx_l dx_f\right] \tag{27}$$

[10] This integrates out the research labor of leader and laggard.

Definition: A balanced-growth path equilibrium is

1. Prices

$$\omega, r, \{\omega_{x_{kjt}}\}_{k\in I_j, j\in[0,1], m\in\mathcal{M}}$$

2. Leader and laggard value functions

$$\mathcal{V}(m, x_l, x_f), \mathcal{W}(m, x_f) \text{ and } \mathcal{V}(-m, x_l, x_f), \mathcal{W}(-m, x_l)$$

3. Leader and laggard policy functions

$$l^*(m), l^*(-m), g^*(m, x_f), g^*(-m, x_l)$$

4. Quantities

$$\{Y_t, C_t\}_{t\geq 0}, L, \{x_{kjt}\}_{k\in I_j, j\in[0,1], m\in\mathcal{M}\}$$

5. Aggregate productivity $\{Z_t\}_{t\geq 0}$

6. Distribution of industry masses over states $\{\mu(m, x_l, x_f)\}_{m\in\mathcal{M}}$

Such that:

1. Given prices, the equilibrium quantities solve the household problem.

2. Given the demand schedule from the final goods producer and residual labor supply curves generated by the household, intermediate firms' policy functions constitute an industry MPE, with equilibrium payoffs given by their value functions.

3. C_t,Y_t and Z_t all grow at the same constant rate, with rate given by (27)

4. The ergodic distribution over industry states $\mu(m, x_l, x_f)$ is consistent with the firm policy functions.

5. Markets clear at every point in time.

4 Quantification

Following Akcigit and Ates (2023), I set τ_π to 0.2 and τ_{sl} and τ_{sf} to 0.2.[11] There are 12 structural parameters to calibrate. Three I calibrate externally. I calibrate ρ to a standard value. I set the arrival rate of the death shocks to match the frequency that firms disappear from Compustat post 2000. I set the frequency of adjustment to match the frequency of adjustment in R&D expenditures greater than or equal to 2.5% in Compustat post 2000. I am free to pick one of either κ or ϵ, so I set κ to 0.1.

The remaining 8 parameters I calibrate with indirect inference. Although the parameters are identified jointly, certain moments are more informative about one parameter versus another. λ, the innovation step-size, strongly influences the aggregate growth rate, as can be seen in Equation (27). η, the within industry elasticity of substitution, overwhelmingly determines the average markup. If η is relatively low, firms within an industry are not differentiated, so small changes in productivity drastically reduce the laggard firms markup, driving the average markup down. As η goes to zero, the goods converge to perfect complements, and in that case the planner chooses to allocate its spending equally across all firms in the economy sending the average markup to its maximal level. ϵ, the productivity term in the innovation production function, is identified from the level of R&D spending relative to GDP. Fixing a scale parameter, γ, the level of ϵ determines how much x is required to generate a certain arrival rate. When ϵ is high, the quantity of researchers required is lower, which means aggregate expenditure is lower (both because firms are further down their residual labor supply curve and because they mechanically purchase less x). ϕ, the probability that the laggard makes the industry leveled conditional on receiving an innovation shock, is identified off of the frequency of leader changes in the data. If ϕ is high, so that most innovations are incremental, the sales leader is more persistent. On the other hand, if ϕ is low, so that most innovations induce catch-up, then their is substantial churn in the identity of the sales leader. This is for two reasons. The first is mechanical. When ϕ is low, most changes are large and restore the industry to being leveled. The second is behavioral. When ϕ is low, the leader internalizes the fact

[11] These are the values for the endpoint of the transition path in that paper - see their Table 4.

that any additional profits it would gain from the incremental innovation it has are relatively short lived. This reduces the leader's incentive to engage in research relative to the laggard, which is transmitted through the industry specific portion of the wage back to the laggard.

There are three elasticity parameters to calibrate. The first is ψ, which governs the elasticity of substitution of labor within sectors. The second elasticity parameter is θ, which governs the elasticity of substitution of labor across sectors. The third elasticity is the technological elasticity of innovations with respect to researchers γ. I identify these three parameters jointly off of three moments. The first moment is the estimated log wage difference between inventors who switch to young/small firms as opposed to old/large firms from Akcigit and Goldschlag (2023). Matching this moment to my model requires care, because my model is one of firms, whereas that moment compares inventors who move across firms. To make the two comparable, I treat that moment as the steady state mean log wage difference between leading firms (i.e. those with $m > 0$) and laggard firms (i.e. those with $m < 0$). The second moment is the variance of residualized log wages in the Survey of Doctoral Recipients. This variance is calculated as follows. I restrict the SDR to individuals employed in the private sector who work 40 hours and then trim the top and bottom 10% of the sample. I then run a fixed effects regression include age, year, occupation, and other demographic controls, as well as interactions between year and occupation. The third moment is the mean difference in successful patent applications between individuals who move to old/large and young/small firms. Analogously with the wage differential, I interpret this as a steady state differential in arrival rate per worker between leading and laggard firms. Specifically, I calculate the model analogue of this moment as $\epsilon x_l^{\bar{\gamma}-1} - \epsilon x_f^{\bar{\gamma}-1}$. Appendix Table 6 describes the sensitivity of each moment to 1% increases in each of the parameters.

These three moments together serve to identify the three elasticity parameters in the model. Increases in θ, holding ψ fixed, serve to make industries more substitutable from the perspective of the household. This serves to increase the capacity of laggard firms to attract outside labor, reducing the capacity of the leading firm to raise its rival's costs. As a consequence, leading firms reduce their purchases of output more, inducing a decline in the estimated innovation productivity

gap. The impact on the variance of log wages, in contrast, is non-monotonic. At low values of θ, an increase in θ serves to make industries more substitutable, which leads the household to reallocate workers to industries that have higher demand. The consequence of this is that the variance of log wages modestly increases. As θ goes to infinity, however, the household views all industries as perfectly substitutable, and industry shadow wages must be equated. A similar effect ensues when ψ is increased. When the household views different firms as relatively more substitutable within the industry, it becomes relatively easier for the leading firm to affect the laggard's costs. This induces them to increase their purchases, pushing up the innovative productivity gap and the log wage difference between leading and laggard firms. In the limit, however, the difference in log wages within an industry declines to 0, as one firm offering an ϵ higher wage would immediately acquire all labor allocated to the industry.[12] Increases in γ serve to decrease the productivity of relatively smaller firms, which induces a reallocation of research labor towards leading firms in the industry. This drives up the variance of log wages by increasing the gap in researchers across leader and laggard firm, and increases the estimated wage gap. The innovative productivity gap, however, declines because of reduced curvature in the innovation production function. Stacking the targeted statistics implied by a vector of structural parameters ϑ and denoting them by $\mathbf{m}(\vartheta)$, the problem of indirect inference can be written as:

$$\min_{\vartheta} \mathbf{D}(\mathbf{m}(\vartheta), \mathbf{m}_{\text{DATA}})' \mathbf{W} \ \mathbf{D}(\mathbf{m}(\vartheta), \mathbf{m}_{\text{DATA}})$$

Where $\mathbf{D}(\cdot, \cdot)$ is a distance function and $\mathbf{W}$ a weighting matrix. I use a distance function $\mathbf{D}$ that returns percentage deviations from data level of the targeted moment. Following Altonji and Segal (1996), I use the identity as my weighting matrix.[13]

[12] This non-monotonicity serves to make calibration challenging. I am working on an alternative calibration approach that identifies ψ and θ via a variance decomposition of researcher wages within and across industries. Specifically, if $\frac{1}{\theta} - \frac{1}{\psi}$ increases, the variance of within industry log wages should decline. Quantitatively, θ itself seems to have only modest effects on the variance of log wages.

[13] No adjustment in the weighting matrix for different levels of the moments is needed by virtue of the choice of distance function.

Table 2: Calibration Parameters

Parameter		Value	Moment	Model	Data
Externally calibrated					
ρ	Discount rate	0.030	Standard Value		
δ	Death rate	0.091	Frequency of Firm Exit in Compustat		
ν	Rate parameter of adjustment	0.590	Frequency of $> 2.5\%$ fluctuations in R&D expenditure in Compustat		
κ	Inventor Disutility	0.100	Normalization		
Internally calibrated					
λ	Step Size	1.052	Growth rate from Fernald TFP series	0.81%	0.77%
η	Within industry product substitutability	0.907	Markup at end of series in Akcigit and Ates (2023)	0.24	0.240
ϕ	Probability of Incremental Innovation	0.847	Frequency of Change in Industry Sales Leader in Compustat	0.072	0.108
ψ	Within industry labor substitutability	0.801	ATT on Log Earnings in Akcigit and Goldschlag (2023)	0.107	0.118
θ	Across industry labor substitutability	0.317	Variance of Trimmed Log Wages in the Survey of Doctoral Recipients	0.043	0.046
ϵ	Innovation TFP term	0.304	Average R&D to GDP Ratio post 2000	0.027	0.024
γ	Innovation returns to scale	0.474	ATT on Patent Filing in Akcigit and Goldschlag (2023)	-0.041	-0.036
Policy Parameters					
τ_π	Tax on gross profits	0.200	Taken from Akcigit and Ates (2023)		
τ_{sf}	laggard research subsidy	0.200	Taken from Akcigit and Ates (2023)		
τ_{sl}	Leader research subsidy	0.200	Taken from Akcigit and Ates (2023)		

I summarize the model parameters and associated identifying moments in Table 2. The model does adequately at matching the targeted moments, with the exception of the frequency of sales leader changes. This mismatch occurs both because the market shares of leading firms in my model are unrealistically large: all leading firms in my model have instantaneous market shares of at least 50%. In the data, a large number of sales leaders have considerably smaller leadership shares, and moreover are much closer to the leader than the minimum 10% difference implied by my calibration. That firms in the data are much closer to each other than in my model, in turn, leads market leadership in the data to change for reasons other than productivity improvements at one firm or the other, principally firm specific demand shocks.[14]

As demonstrated in Table 3, the model also performs well in replicating qualitatively - in an untargeted manner - the within industry correlation in Compustat between the leader sales share and leader share of R&D expenditures. This is computed by simulating a panel of 10000 firms for 25 periods, and then projecting leader share of industry R&D on leader share of industry sales yearly and industry fixed effects.

[14]For instance, NVIDIA's revenue in real terms in 2023 was roughly two and a half times its revenue in 2022. This corresponded to a roughly doubling of its share of industry revenue in SIC code 3674 *Semiconductors and Related Devices*. This increase was driven by increased demand to feed generative AI models, not an underlying change in the productivity of NVIDIA or the quality of its products. It was changes in the productivity of complements of the products NVIDIA makes that drove the changes in revenue.

Table 3: Within Industry Correlation Between Leader Share of Sales and Leader Share of R&D

Dependent Variable Leader Share of Industry R&D	Data	Model
Variable		
Leader Share of Industry Sales	0.8207***	0.7248***
	(0.0584)	(0.0015)
Fixed-effects		
Industry	Yes	Yes
Fit statistics		
Observations	5,681	216,230
R^2	0.70286	0.875
Within R^2	0.19123	0.525

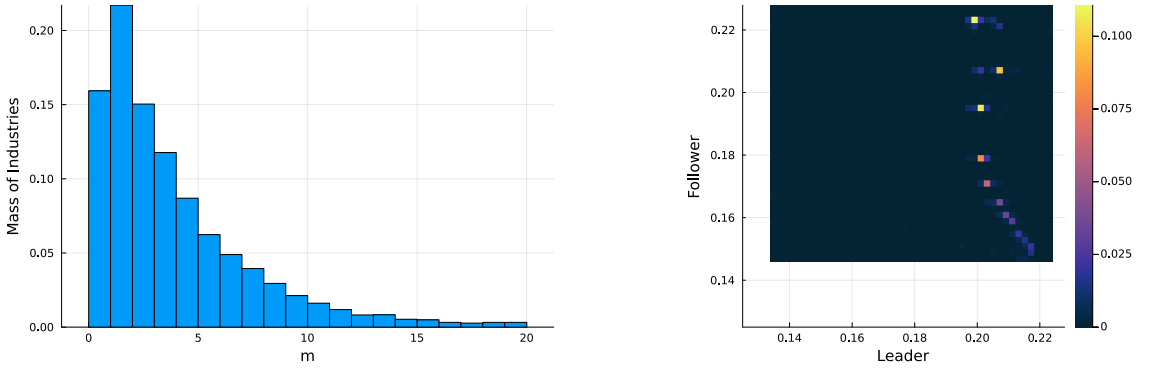
Industry clustered standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

5 The Steady State

In this section, I examine the structure of the balanced growth path equilibrium and engage in a series of comparative statics. The first three of these comparative static exercises are designed to imitate key themes in the literature: ideas getting hard to find (as in Bloom et al., 2020), changes in the returns to scale in the innovation production function, and an increase in "industry toughness" which makes industries winner-take-all (as in Autor et al., 2020). In the final comparative static, I consider an increase in ψ, the within industry elasticity of substitution. To characterize more sharply how changes in these parameters affect the aggregate outcomes, I focus my attention relatively large shocks to the parameters values.

5.1 The Calibrated Balanced Growth Path

Recall that the ergodic distribution in this economy is a joint distribution of masses over productivity gaps, denoted by m, and quantities of leader and laggard research labor, x_l and x_f. Figure 5 provides some information about this joint distribution by plotting the unconditional distribution over industry gaps and unconditional joint distribution over research labor in the simulated ergodic



(a) Unconditional Distribution over m

(b) Unconditional Distribution over (x_l, x_f)

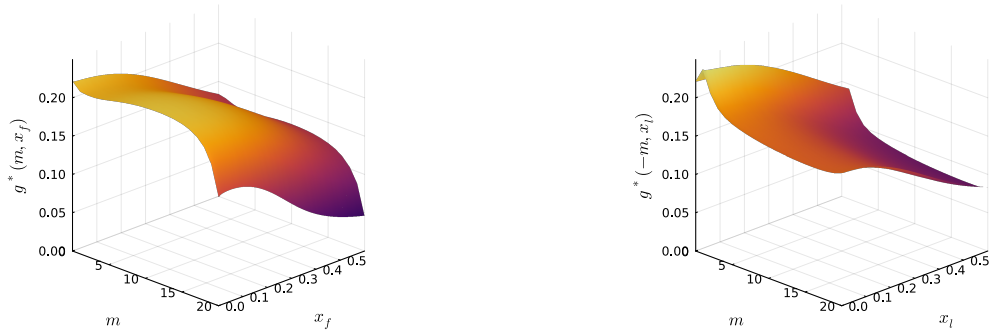
Figure 5: Unconditional distributions of ergodic distribution.

distribution.[15]

Figure 5 conveys two key takeaways. Firstly, most industries have relatively small productivity differences, with roughly 70% of industries having a productivity difference of less than 5 steps. Secondly, in most industries, the leading firm hires relatively more researchers. This second result can also be seen in the policy functions for leading and laggard firms, which are depicted in Figure 6. At $m = 0$ the symmetry assumption implies that the value policies are the same. As m increases, however, the policy functions exhibit a non-monotonic shape, with the leader first choosing fewer researchers, then increasing numbers of researchers before a decline as they move towards $\bar{m}$. The laggard firm, in contrast, aggressively increases the number of researchers they hire before leading to stronger declines.

To gain intuition for the shape of the policy functions, it is useful to consider a firm at $m = 0$ that takes the policy functions of the other as given. Now suppose that the firm receives a productivity shock, moving the industry state to $m = 1$. This has two major consequences for the firm, independent of the resultant increase in flow profits. The first is that it means that they are no longer exposed to the probability of replacement δ. This serves to raise the value in the state $m = 1$ relative to state $m = 0$. Importantly, however, once a firm has positive productivity difference, this persists so long as they maintain that productivity difference. The second consequence is that the

[15] I recover the ergodic distribution by simulation. Specifically, given an initial allocation of research labor, I simulate exactly the evolution of $N = 10000$ industries, stopping at $T = 5000$. The details of my algorithm are in the appendix, as are figures plotting the same unconditional distributions at $T = 10000$.



(a) Leader Policy Functions

(b) Laggard Policy Functions

Figure 6: Policy Functions at Calibrated Balanced Growth Path

other firm will aggressively increase their usage of research labor. This serves to reduce the value of the leading firm. As the leading firm builds up its lead, the laggard firm reduces its innovation expenditures, and this raises the expected duration of firm profits for the leading firm. This raises the value of incremental innovations and the value of "tempering" the laggard by raising their costs, giving rise to the hump shape. The other noticeable feature of the leader's policy function is a pronounced hump in the quantity of research labor demanded conditional on the productivity difference m. This arises because, when the laggard is committed to low quantities of research labor, the leading firm's elasticity is quite close to the elasticity of substitution across industries θ. This relatively low elasticity, in turn, means that wages must rise substantially to hire large quantities of new researchers.

This intuition takes the laggard's choices as given, but they are also endogenous objects. What drives their decisions? As with before, begin at the $m = 0$ step but now suppose that the other firm receives a productivity shock. In that case, they face a relatively higher probability of death instantaneously, which pushes down the value at $m = -1$ relative to $m = 0$, and serves to increase their incentive to invest to avoid death, leading to the spike at $m = 1$. In addition to this somewhat mechanical effect, the fact that ϕ is large means that most innovations are incremental. As a result, the laggard firm internalizes that the further it falls down the quality ladder, the lower its probability of retaking the lead before a death shock arrives. This induces them both to compete

very aggressively when only slightly behind, contributing to the large hump at $m = -1$, but also to lower their innovation expenditures, with the effect growing as they go further and further behind. This is what generates the downward sloping shape in the policy functions.

5.2 Who Wins When Ideas Get Harder to Find?

The core observation of Bloom et al. (2020) is that despite vastly increasing quantities of researchers, the rate of growth has remained roughly constant over time. This immediately implies that for the same quantity of researchers, the quantity of new ideas that are generated is declining over time. In my model, such a change is straightforward to implement by simply reducing the innovation TFP term, ϵ. I consider a 10% reduction in this parameter, which takes ϵ from approximately 0.304 to approximately 0.274.

The relatively lower innovation productivity generates slower arrival rates. This mechanically increases the persistence of leadership in industries, as it takes longer for firms to generate potential innovation shocks. As a consequence, leading firms have a stronger incentive to invest in productivity improvements, with the exception of laggard firms who are just one step behind. These firms take up a larger share of industry R&D because they internalize that, given the approximately step-by-step behavior in the model that occurs as a result of the calibrated ϕ, falling behind more than one step in a world with lower ϵ implies a much more persistent status as the laggard firm.

R&D to GDP falls as a result of the decline in research productivity, and the gap in between average patents per worker and log wage differences across leader and laggard firms opens up over time. The raw variance of researcher log wages also increases because of the asymmetry in leader and laggard responses to the decline in productivity.

5.3 Changes in R&D Scale

Increases in the R&D scale parameter behave very similarly to the ideas getting harder to find equilibrium. This balanced growth path is characterized by a moderately slower growth rate, but increased research expenditures as a percentage of GDP. Leading firms benefit relatively more

from the decline in research efficacy per worker, as it implies longer periods of leadership, and thus higher returns to leading firm innovation. The variance of log researcher wages increases for much the same reason.

As with the prior comparative static, leaders benefit disproportionately from the change in γ. This can be seen in the fact that the mean log wage differential between the two firms becomes substantially larger, increasing by nearly 30% from the calibrated baseline. In contrast to the prior comparative static, however, the gap in inventive productivity between leading and laggard firms lies relatively unchanged despite differences in the number of workers. This is because decreases in the curvature of the innovation production function make the average innovation rate per worker less sensitive to the number of workers.[16]

5.4 More Cutthroat Competition

Next consider cases where the elasticity of substitution within an industry, as parameterized by η, varies. It is useful to keep in mind two limit cases. The first is when η limits to 0. In that case, the goods are perfect complements, and competition within the industry is as weak as possible. In contrast to this, in the case when η limits to 1, the static price competition game is undifferentiated Bertrand. In this case, the optimal strategy of the more productive firm is to limit price the other firm, meaning that any firm that has a productivity advantage will immediately acquire the entirety of the market. In line with this scenario, I consider a roughly 5% increase in η, going from roughly 0.91 to roughly 0.97.

This change results in a substantially different balanced growth path from the prior two comparative statics. Growth and R&D to GDP decline modestly because improvements after gaining the initial productivity lead are relatively smaller, reducing the incentive of leading firms to invest in pushing the productivity frontier forward. In contrast, however, the average markup tilts declines roughly 27% from the baseline. This occurs because the markups of the laggard firm decline

[16] One way to see this is to take the limit case where the innovation production function has constant returns to scale. In that case, the average arrival rate in per worker terms is just ϵ, the innovation TFP term.

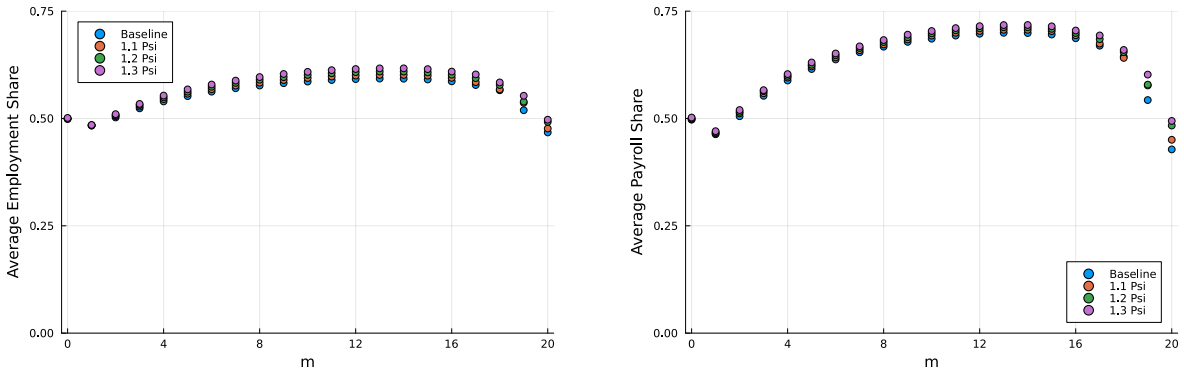
precipitously. Less differentiation between firms immediately implies that they will charge a price closer to their marginal cost, leading to a lower average markup in equilibrium.

Laggard firms experience substantially reduced incentives to innovate relative to leading firms, as the roughly step-by-step nature of the dynamic game implied by the calibrated ϕ means that firms who are several steps behind regain much smaller market shares in expectation upon innovating. This drives them to reduce their innovation expenditures. This substantially increases the gap in log wages between leading and laggard firms, which is reflected in all three of the micro-moments.

5.5 An Increase in ψ

What differentiates firms within the same industry in the eyes of the household? One possible reason is non-compete clauses on the part of firms. Motivated by recent bans on non-competes by the Federal Trade Commission, I model the effects of changes in non-compete policy as a 10% increase in the elasticity of substitution across firms within an industry ψ. Somewhat surprisingly, an increase in ψ increases the average log wage difference between leader and laggard firms. This is despite the fact that in the $\psi \to \infty$ limit, within-industry determination of wages is essentially undifferentiated Cournot, wherein both firms will pay the same wage, meaning that this statistic necessarily limits to zero. In fact, this result is driven by the ability of the leading firm to affect the laggard's choice. Consider a limiting case where θ goes to zero, but ψ remains strictly greater than zero. In this limiting case, the household chooses to assign an equivalent mass of workers x_j to each industry. ψ will modulate the extent to which the leading firm can affect the laggard's quantity of researchers: as ψ limits to zero, the household will assign an equivalent mass of workers to every firm in the economy. If, on the other hand, ψ is strictly greater than zero, increases in the quantity of researchers hired by the leading firm reduce the mass available to the laggard, at the cost of the incremental wage increase. To better make this point, Figure 7 plots both the leader's average payroll share and average employment share at each productivity difference.

This provides some insight into a potentially interesting tradeoff to be explored in future work:



(a) Employment Share for Various Perturbations in ψ (b) Payroll Share for Various Perturbations in ψ

Figure 7: Leading Firm Share of Employment and Payroll at Different Productivity Gaps. At the case of $m = 0$, the leading firm is defined to be the firm that last had a productivity lead.

Moment	Baseline	Ideas Getting Harder to Find	Increases in R&D Scale	Cutthroat Competition	Increased Psi
Growth Rate	0.81%	0.69%	0.77%	0.79%	0.82%
Average Markup	0.240	0.242	0.242	0.175	0.240
R&D to GDP	0.027	0.024	0.029	0.025	0.028
Sales Leader Change	0.072	0.065	0.069	0.0598	0.073
Variance of Log Researcher Wages	0.043	0.050	0.048	0.129	0.042
ATT on Patent Filing	-0.041	-0.045	-0.040	-0.184	-0.046
ATT on Log Wages	0.107	0.129	0.133	0.484	0.112

Table 4: New values of moments after perturbing a subset of parameters. "Ideas Getting Harder to Find" corresponds to a 5% reduction in the innovation TFP term. "Changes in R&D Scale" corresponds an increase in the elasticity of innovation with respect with to the number of researchers of 10%. "Cutthroat Competition" corresponds with an increase in the within industry substitutability of 5%. "Increased ψ" corresponds with a 10% increase in the level of ψ.

banning non-competes might tilt the allocation of high skill workers towards relatively larger firms. If this arises because of labor hoarding actions by leading firms and not because of differences in marginal productivity, then this might serve to lessen aggregate efficiency, leading to welfare losses.

To conclude this section, I summarize the results of all four comparative statics in Table 4.

6 Optimal Policy

In this model, neither the structure nor the level of optimal research subsidies is obvious. Increasing the level of research subsidies across the board induces more demand for researchers on the part of firms, which increases the disutility associated with the provision of research labor. On the

one hand, the fact that firms internalize their impact on their residual labor supply curve would, ignoring the other anticompetitive and hoarding value of researchers, lead them to restrain their purchase of labor below the optimum. In the same direction, the leading firm would not internalize the societal benefits of their additional innovation, namely that a laggard firm generates a larger productivity improvement in levels when they recieve a quick catchup shock when they are further behind. These two forces would push a planner to increase the level of R&D subsidies. On the other hand, firms in this economy have an incentive to accumulate research labor to respond to potential changes in their demand for said labor as the productivity difference between the two firms might change. This precautionary hoarding incentive pushes in the direction of too much R&D suggesting a decline in the level of subsidies might be warranted. Which force dominates is an empirical matter.

In the same vein, changes in the structure of R&D subsidies can have subtle impacts on the equilibrium. Consider, for instance, a decline in the level of τ_{sl} relative to τ_{sf} raises the cost of innovation of the leader relative to the laggard, tilting the balance of research expenditures towards the latter. This reduces the average gap between the leader and laggard, reducing the aggregate markup and thereby increasing the level of output. The cost of such a policy is a slower growth rate of the productivity frontier.

In this section, I consider two policy experiments. The first simply asks what constitutes the optimal policy of a government who is forced to pick between steady states. The second policy experiment I consider is one in which the government faces a political constraint that dictates that the set of feasible steady states is only those where the total expense of the research subsidy as a percentage of output is less than or equal to the current steady state expenditure as a percentage of output.

To make the problems that the government solves more precise, I must first derive an expression for welfare in this model. In the appendix, I show that steady state welfare at time 0 can be given by:

$$\mathbf{W} = \frac{1}{\rho}\left[\ln(Z_0) - \ln(\bar{\omega}) + \frac{1}{\eta}\int_0^1 \ln(\mathcal{S}(j))dj - \kappa\bar{X}\right] + \frac{\bar{g}}{\rho^2} \tag{28}$$

Where $Z_t = \exp\left\{\int_0^1 \ln(z_{ijt}dj\right\}$, z_{ijt} is the productivity level of the leading firm in industry j, and $\mathcal{S}(j)$ is a constant that depends on the productivity gap of industry j. To compare across steady states, I fix the initial level of productivity, Z_0, in every steady state to 1. Note that this does not imply that the level of time 0 *output* is constant across steady-states. As can be seen from Equation (26), this itself is a function of the equilibrium distribution of industries across markup gaps. Instead, this makes stark the tradeoff of a planner - increase the growth rate by subsidizing the leader more aggressively or decrease the level distortion induced by markups by subsidizing the laggard more aggressively - without introducing arbitrary bias due to differences in ex-ante productivity across steady-states.

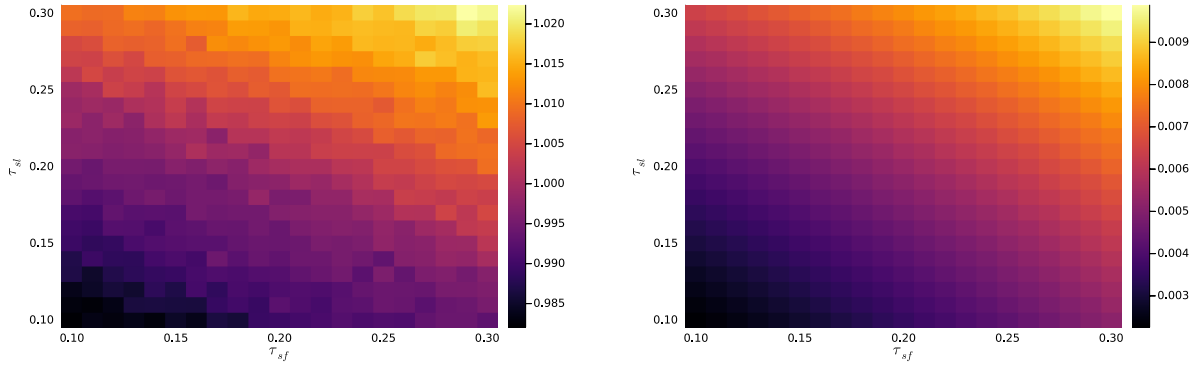
The difference in welfare across steady-states is merely $\mathbf{W}_2 - \mathbf{W}_1$. To render this object interpretable, I express it in terms of compensating variation. I calculate this as the percentage change in the level of consumption at every point in time that makes the household indifferent between the two steady states. Analogously, this is how much one would have to increase initial productivity in the new steady state to make the household indifferent between the new and old steady states. Algebra in the appendix shows that this quantity can be expressed as:

$$\mathbf{CV}(\mathbf{W}_2; \mathbf{W}_1) = \exp\{\rho\left[\mathbf{W}_2 - \mathbf{W}_1\right]\} \tag{29}$$

In all policy exercises, I calculate the compensating variation with respect to the calibrated steady-state. The problem that I consider is thus the following:

$$\max_{\tau_{sl},\tau_{sf}\in[0.1,0.3]^2} \mathbf{CV}(\mathbf{W}(\tau_{sf},\tau_{sl}); \mathbf{W}_{\text{Baseline}})$$

Where $\mathbf{W}(\tau_{sf},\tau_{sl})$ is equal to the steady state welfare in the economy with the policy parame-



(a) Welfare Gains in terms of Compensating Variation (b) Subsidy Cost

Figure 8: All variables used to calculate welfare or cost are taken as average of five separate simulations to reduce effect of simulation error.

ters fixed at that point.[17] Figure 8 displays the results of this exercise. Consistent with the literature, I find that large increases in the size of R&D subsidies generate large welfare gains. Increasing the subsidy rate from 20% to 30% uniformly generates welfare gains equivalent to increasing output by 2.1% at every instant. The fiscal cost of the increased subsidies is quite substantial however, almost doubling government expenditures from 0.53% of GDP in the calibrated steady state to nearly 1%.[18] The total welfare gains and the fiscal cost are presented below in Figure 8.

Inspection of the left panel of Figure 8 also provides a somewhat surprising result. Steady state welfare is relatively more sensitive to the *leader's* subsidy than the laggard's. The subsidy cost, on the other hand, is almost symmetric. This foreshadows the result of my next policy exercise, which I now turn to. In this exercise, I fix the maximum value of the R&D subsidies as a percentage of output to that of the calibrated balanced growth path. This is approximately 0.54% of output. I then ask how much can one improve upon the calibrated equilibrium without increasing the fiscal cost of the subsidy relative to steady state output. Figure 9 plots the set of feasible balanced growth paths and the ensuing welfare gains.

The takeaway from Figure 9 is that subsidies to leading firms are relatively more efficient use

[17] To elucidate the role of research subsidies in this economy, I keep τ_π fixed at the calibrated level.

[18] In the model economy, this is easily fiscally sustainable. The gross profit share amounts to roughly 20% of output, 20% of which accrues to the government as tax revenue. This is an artifact of both the fact that the profit share in this model is unrealistically large and the fact that the corporate income tax in this setup falls on *gross* profits, not *net* profits, as in reality. By way of comparison, the corporate income tax generates revenues of between 1 to 2 percent of United States GDP.

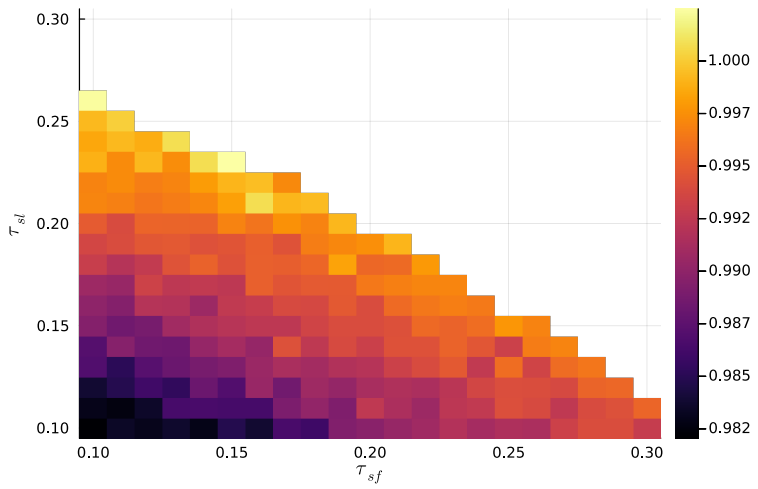


Figure 9: Set of tax policies that satisfy the imposed fiscal condition. Brighter colors indicate higher levels of compensating variation.

of funds. Indeed, the choice of tax policy that solves the aforementioned problem sets the leading firms subsidy to τ_{sl} to 23% and the laggard's subsidy to 15%. Given the initial facts that workers seem to be less productive at leading firms, this may seem a counterintuitive result. To recover some intuition, it is useful to recall Equation 27, which I reproduce below:

$$\ln(\dot{Z}_t) = \ln(\lambda)\left[\int\int \left(\epsilon x_l^{\gamma} + \epsilon x_f^{\gamma}\right)\ \mu(0, x_l, x_f)dx_l dx_f + \sum_{m=1}^{\bar{m}} \int\int \epsilon x_l^{\gamma}\ \mu(m, x_l, x_f)dx_l dx_f\right]$$

The growth rate of TFP - which recall is the same as the growth rate of output - depends the leader's choice of x_l directly, but only the laggard's choice of x_f indirectly, through $\mu(m, x_l, x_f)$.[19] Changes in the leader's cost through changes in the subsidy rate directly affect the leader's choice of x_l and through that the aggregate TFP term. Put slightly differently, an increase in τ_{sl} generates a second order effect on the aggregate growth rate by reducing the cost of a given quantity of x_l, inducing the leading firm to purchase more. In contrast, increases in τ_{sf} have third order effects on the aggregate TFP term by changing the invariant distribution in steady state. There is also a slightly more subtle effect that comes from the government's choice of the relative levels of the

[19]Note that in the case of neck and neck industries, there is no leading firm, I denote the leader as the last firm to have a positive m, but provide both firms with the τ_{sl} subsidy rate.

subsidy. When the government sets a lower subsidy rate for laggard firms, they make competition in leveled industries more aggressive. This is because falling behind has become relatively more costly for firms, so they respond by increasing their choice of equilibrium arrival rate. Finally, there is a direct cost effect of the follower's subsidy on the leader's costs. When τ_{sf} is relatively large, laggard firms hire more researchers. This increases the wages that the leading firm must pay to hire any given quantity of researchers. Reducing τ_{sf} reduces the quantity that the laggard firm purchases, reducing the wages that the leading firm must pay and inducing them to hire more researchers. Quantitatively, this third force seems to be quite powerful, as experiments where θ and ψ are closer together generate much smaller welfare gains and can indeed even flip the direction of the subsidies, so that a planner would choose to subsidize laggards relatively more. These three forces lead the growth rate to be relatively more responsive to leader subsidies.

The other portion of the planner's trade off is the level of output. Larger subsidies for the leader, in theory, might push the distribution of industries further out, increasing the mean productivity difference between the two firms, and thus the average markup. This force, however, is relatively weak at the calibrated parameters. The distribution of productivity gaps is not very responsive to changes in the subsidy rate. This is for three reasons. The first is that quick catch-up is relatively uncommon, in the sense that ϕ is quite high. This means that even along the calibrated balanced growth path, leading firms do a relatively large proportion of the R&D in the economy. A side effect of this is that the invariant distribution over m is quite stable. The second reason is that despite differences in research spending, there are relatively small discrepancies between the effective arrival rates of leading and laggard firms, with leading firms receiving a shock on average roughly once every 7.1 years and laggard firms receiving shocks on average once every 7.3 years in the calibrated steady state.[20] Because leader innovation is incremental by construction, leading firms are limited in their capacity to build up large leads. The third reason is that by offering a higher subsidy rate for firms in leveled industries relative to laggard firms, you increase the value

[20]This is calculated in a rule of thumb manner by taking the average arrival rate of leading and laggard firms in industries where $m > 0$ in the ergodic distribution. The expected time to arrival is then calculated as the inverse of that arrival rate, ignoring the potential for changes in the rate parameter.

of firms in the leveled industry relative to laggard firms. Because laggard firms always have a fixed probability of being able to level the industry, conditional on an innovation shock, this increases the return to research to recover the additional subsidies when the industry is leveled. These effects together serve to make the average markup relatively less sensitive to changes in subsidies than the growth rate, and thus weaken the impact of the subsidies on the level of output.

7 Conclusion

I build a novel model of dynamic oligopoly that features interlaced competition in product markets and researcher labor markets. I calibrate it to match findings in Akcigit and Goldschlag (2023) about the relative productivity of inventors at large and small firms. I use the calibrated model to examine the impact of R&D subsidies, and find that reshuffling the subsidies to target leading firms generates improvements in welfare roughly equal to increasing the level of output in the steady state by a quarter of a percentage point in perpetuity.

My model suggests that naive intuitions about the sources of differences in wages and innovative output across firms may lead one astray if leading firms produce better products than laggard firms. Importantly, however, this results depends on whether or not innovation by laggard firms takes the form of improving their existing products, as Garcia-Macia et al. (2019) argue, or creating new varieties. Further evidence examining differences in the innovative activities of leading and laggard firms and how these have changed over time are crucial to understanding the policy impacts of changes in research subsidies. In addition, more work is required to get at more direct identification of the differing elasticities of substitution across firms and industries on the part of workers, as well the sources thereof. To that end, in ongoing work I examine the movement patterns of individual doctoral researchers to better estimate these elasticities and identify sources of variation.

Dependent Variable:	5-Year Forward Citations			5-Year Outside Forward Citations		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
τ	2.213***	1.742***	1.952***	2.034***	1.737***	1.813***
	(0.719)	(0.696)	(0.702)	(0.579)	(0.535)	(0.557)
Fixed-effects						
Proxied Age	No	Yes	No	No	Yes	No
Calendar Year	No	No	Yes	No	No	Yes

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 5: Synthetic Difference-in-Differences result when considering only first switches. All standard errors are calculated using the jackknife approach.

A Appendix

A.1 Additional Synthetic Difference in Difference Results

In this section, I repeat the synthetic difference in difference estimation but limit consideration to only the first time an individual appears in the dataset of relevant switches, which means that individuals are only "treated" a single time. As can be seen in Table 5, the results are robust to this alternative specification.

Figures 10, 11, and 12 present the analogous synthetic difference in difference plots to those found in the main text.

A.2 Household Problem

The household problem in this environment is:

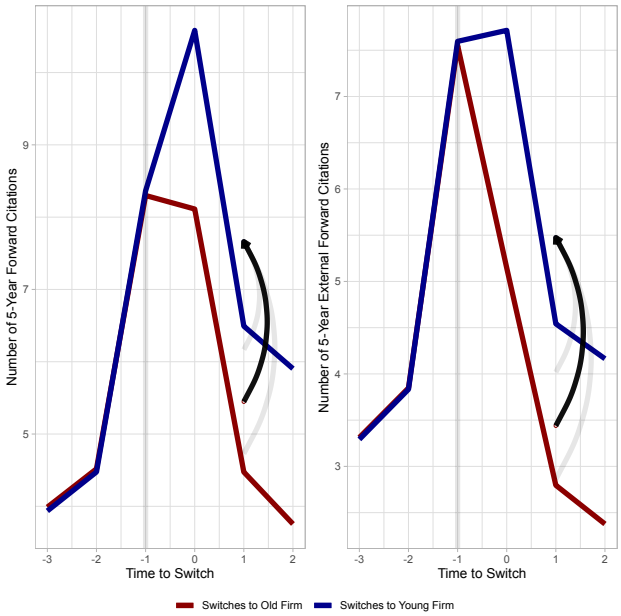


Figure 10: Baseline Synthetic Difference in Differences results considering only first switches.

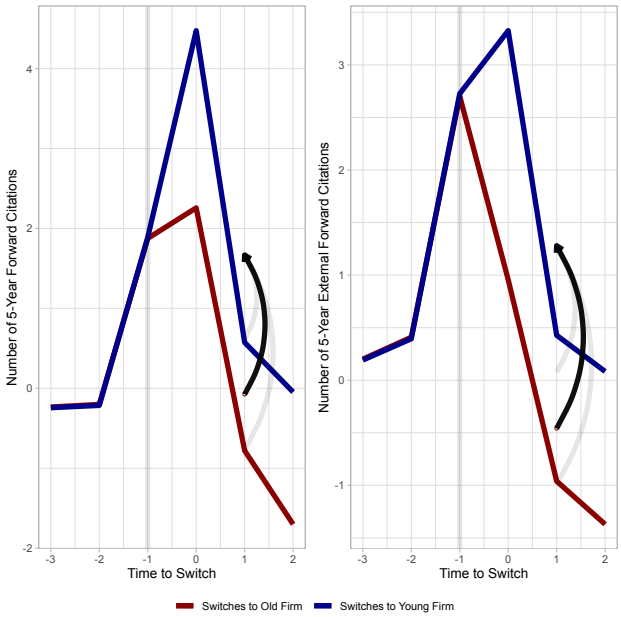


Figure 11: Age Proxy Synthetic Difference in Differences results considering only first switches.

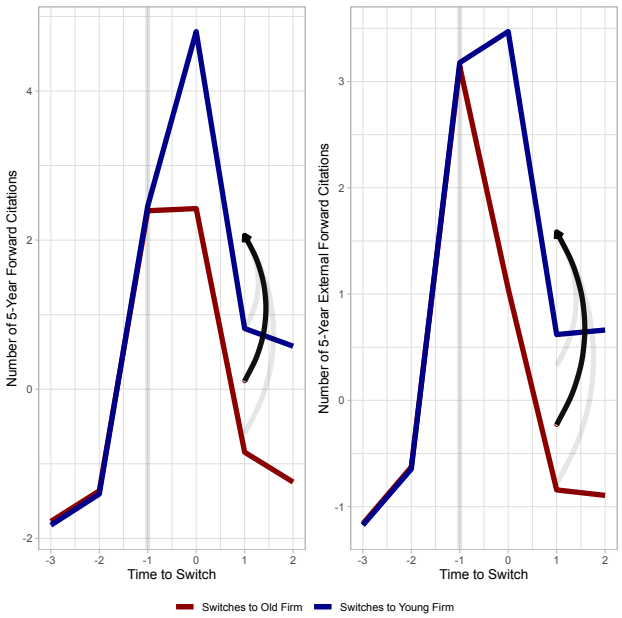


Figure 12: Calendar Year Synthetic Difference in Differences results considering only first switches.

$$\max_{C_t,\{\{x_{ij,t}\}_{i\in I_j}\}_{j\in[0,1]}} \int_0^\infty e^{-\rho t}\left[\ln(C_t) - \kappa X_t\right] dt$$

Subject to:

$$\dot{A}_t = r_t A_t + \bar{w}_t + \int_0^1 \sum_{i\in I_j} w_{ij,t} x_{ij,t} + T_t - C_t$$

Where I have already normalized $\bar{L} = 1$.

The current-value Hamiltonian formulation of this problem is

$$\begin{aligned}\mathcal{H}_t(C_t, \{\{x_{ij,t}\}\}, A_t, \lambda_t) =& e^{-\rho t}\left[\ln(C_t) - \kappa\left(\int_0^1 \left(\sum_{i\in I_j} x_{ij,t}^{\frac{\psi+1}{\psi}}\right)^{\frac{\psi}{\psi+1}\frac{\theta+1}{\theta}} dj\right)^{\frac{\theta}{\theta+1}}\right] \\ &+\lambda_t\left(r_t A_t + \bar{w}_t + \int_0^1 \sum_{i\in I_j} w_{ij,t} x_{ij,t} + \Pi_t - C_t\right)\end{aligned}$$

The maximum principle gives rise to the following set of optimality conditions:

$$e^{-\rho t}\frac{1}{C_t} - \lambda_t = 0$$

$$\lambda_t r_t + \dot{\lambda}_t = 0$$

$$-e^{-\rho t}\kappa \left(\frac{x_{jt}}{X_t}\right)^{\frac{1}{\theta}} \left(\frac{x_{ij,t}}{x_{j,t}}\right)^{\frac{1}{\psi}} + w_{ij,t}\lambda_t = 0$$

Combining the first and third conditions gives rise to same wage condition in Berger et al. (2022), modulo an adjustment for growth. This adjustment amounts to the standard condition that - on a balanced growth path - the marginal utility weighted wage must be constant.

$$\frac{w_{ij,t}}{C_t} = \kappa \left(\frac{x_{jt}}{X_t}\right)^{\frac{1}{\theta}} \left(\frac{x_{ij,t}}{x_{j,t}}\right)^{\frac{1}{\psi}}$$

The first term characterizes the value of the costate multipler λ_t. Differentiating this expression for the costate multiplier with respect to t, one recovers:

$$\dot{\lambda}_t = -\rho e^{-\rho t}\frac{1}{C_t} - e^{-\rho t}\frac{\dot{C}_t}{C_t^2}$$

This term can be re-expressed as

$$\dot{\lambda}_t = -\rho\lambda_t - \lambda_t\frac{\dot{C}_t}{C_t}$$

$$-\frac{\dot{\lambda}_t}{\lambda_t} = \rho + \frac{\dot{C}_t}{C_t}$$

Using the second condition to substitute in for the left hand side, one recovers the standard Household Euler equation.

$$r_t = \rho + \frac{\dot{C}_t}{C_t}$$

A.3 Static Industry Problem

There is a final goods producer that aggregates output from a continuum of duopolistic industries according to:

$$\ln(Y_t) = \int_0^1 \frac{1}{\eta} \ln(y_{i,jt}^{\eta} + y_{-i,jt}^{\eta}) dj \tag{30}$$

This means that the problem of the final goods producer is given by:

$$\max P_t Y_t - \int_0^1 \sum_i p_{i,jt} y_{i,jt} dj$$

s.t. :

$$Y_t = \exp\left\{ \int_0^1 \frac{1}{\eta} \ln(y_{i,jt}^{\eta} + y_{-i,jt}^{\eta}) dj \right\}$$

Substituting in, we have:

$$\max P_t \exp\left\{ \int_0^1 \frac{1}{\eta} \ln(y_{i,jt}^{\eta} + y_{-i,jt}^{\eta}) dj \right\} - \int_0^1 \sum_i p_{i,jt} y_{i,jt} dj$$

First order conditions give us that:

$$P_t Y_t \left[\frac{y_{i,jt}^{\eta-1}}{y_{i,jt}^{\eta} + y_{-i,jt}^{\eta}} \right] = p_{i,jt}$$
$$P_t Y_t \left[\frac{y_{-i,jt}^{\eta-1}}{y_{i,jt}^{\eta} + y_{-i,jt}^{\eta}} \right] = p_{-i,jt}$$

Which yields the familiar equivalence:

$$p_{i,jt}y_{i,jt}^{1-\eta} = p_{-i,jt}y_{-i,jt}^{1-\eta}$$
$$\frac{p_{i,jt}}{p_{-i,jt}} = \left(\frac{y_{-i,jt}}{y_{i,jt}}\right)^{1-\eta}$$

Now we do further algebra:

$$\begin{aligned}
P_tY_t &= p_{i,jt}y_{i,jt} + p_{i,jt}\left[\frac{y_{-i,jt}}{y_{i,jt}}\right]^{\eta} y_{i,jt} \\
&= p_{i,jt}y_{i,jt}\left[1 + \left(\frac{y_{-i,jt}}{y_{i,jt}}\right)^{\eta}\right] \\
&= p_{i,jt}y_{i,jt}\left[1 + \left(\frac{p_{i,jt}}{p_{-i,jt}}\right)^{\frac{\eta}{1-\eta}}\right] \\
&= p_{i,jt}y_{i,jt}\left[1 + \left(\frac{p_{-i,jt}}{p_{i,jt}}\right)^{\frac{\eta}{\eta-1}}\right] \\
&= p_{i,jt}y_{i,jt}\left[\frac{p_{i,jt}^{\frac{\eta}{\eta-1}} + p_{-i,jt}^{\frac{\eta}{\eta-1}}}{p_{i,jt}^{\frac{\eta}{\eta-1}}}\right] \\
&= y_{i,jt}p_{i,jt}^{1-\frac{\eta}{\eta-1}}\left[p_{i,jt}^{\frac{\eta}{\eta-1}} + p_{-i,jt}^{\frac{\eta}{\eta-1}}\right] \\
&= y_{i,jt}p_{i,jt}^{-\frac{1}{\eta-1}}\left[p_{i,jt}^{\frac{\eta}{\eta-1}} + p_{-i,jt}^{\frac{\eta}{\eta-1}}\right] \\
y_{i,jt} &= P_tY_t\left(\frac{p_{i,jt}^{\frac{1}{\eta-1}}}{p_{i,jt}^{\frac{\eta}{\eta-1}} + p_{-i,jt}^{\frac{\eta}{\eta-1}}}\right) \\
y_{i,jt} &= Y_t\left(\frac{p_{i,jt}^{\frac{1}{\eta-1}}}{p_{i,jt}^{\frac{\eta}{\eta-1}} + p_{-i,jt}^{\frac{\eta}{\eta-1}}}\right)
\end{aligned}$$

Where the last line follows from a normalization of the aggregate price level to unity. Each firm's

production technology is given by:

$$y_{i,jt} = z_{i,jt} l_{i,jt}$$

Firms compete à-la-Bertrand. This means that problem of firm i, given firm $-i$'s choice of price $p^*_{-i,jt}$ is to pick $p^*_{i,jt}$ to solve:

$$\max y_{i,jt} \left(p_{i,jt} - \frac{w_t}{z_{i,jt}} \right)$$

s.t.

$$y_{i,jt} = Y_t \left(\frac{p_{i,jt}^{\frac{1}{\eta-1}}}{p_{i,jt}^{\frac{\eta}{\eta-1}} + p_{-i,jt}^{\frac{\eta}{\eta-1}}} \right)$$

Substituting in, this problem becomes:

$$\max Y_t \left(\frac{p_{i,jt}^{\frac{1}{\eta-1}}}{p_{i,jt}^{\frac{\eta}{\eta-1}} + p_{-i,jt}^{\frac{\eta}{\eta-1}}} \right) \left(p_{i,jt} - \frac{w_t}{z_{i,jt}} \right)$$

The first order condition of this problem is given by:

$$0 = Y_t \left(\left[\frac{1}{\eta - 1} \frac{p_{i,jt}^{\frac{1}{\eta-1}-1}}{(p_{i,jt}^{\frac{\eta}{\eta-1}} + p_{-i,jt}^{\frac{\eta}{\eta-1}})} - \frac{\eta}{\eta - 1} \frac{p_{i,jt}^{\frac{1}{\eta-1}} p_{i,jt}^{\frac{\eta}{\eta-1}-1}}{(p_{i,jt}^{\frac{\eta}{\eta-1}} + p_{-i,jt}^{\frac{\eta}{\eta-1}})^2} \right] \left(p_{i,jt} - \frac{w_t}{z_{i,jt}} \right) + \left(\frac{p_{i,jt}^{\frac{1}{\eta-1}}}{p_{i,jt}^{\frac{\eta}{\eta-1}} + p_{-i,jt}^{\frac{\eta}{\eta-1}}} \right) \right)$$

From there it follows that:

$$\left(\frac{w_t}{z_{i,jt}}\right)\left[\frac{1}{\eta-1}\frac{p_{i,jt}^{\frac{1}{\eta-1}-1}}{(p_{i,jt}^{\frac{\eta}{\eta-1}}+p_{-i,jt}^{\frac{\eta}{\eta-1}})}-\frac{\eta}{\eta-1}\frac{p_{i,jt}^{\frac{1}{\eta-1}}p_{i,jt}^{\frac{\eta}{\eta-1}-1}}{(p_{i,jt}^{\frac{\eta}{\eta-1}}+p_{-i,jt}^{\frac{\eta}{\eta-1}})^2}\right]$$

$$=\left(\left[\frac{1}{\eta-1}\frac{p_{i,jt}^{\frac{1}{\eta-1}}}{(p_{i,jt}^{\frac{\eta}{\eta-1}}+p_{-i,jt}^{\frac{\eta}{\eta-1}})}-\frac{\eta}{\eta-1}\frac{p_{i,jt}^{\frac{1}{\eta-1}}p_{i,jt}^{\frac{\eta}{\eta-1}}}{(p_{i,jt}^{\frac{\eta}{\eta-1}}+p_{-i,jt}^{\frac{\eta}{\eta-1}})^2}\right]+\left(\frac{p_{i,jt}^{\frac{1}{\eta-1}}}{p_{i,jt}^{\frac{\eta}{\eta-1}}+p_{-i,jt}^{\frac{\eta}{\eta-1}}}\right)\right)$$

$$=\left[\frac{\eta}{\eta-1}\frac{p_{i,jt}^{\frac{1}{\eta-1}}}{(p_{i,jt}^{\frac{\eta}{\eta-1}}+p_{-i,jt}^{\frac{\eta}{\eta-1}})}-\frac{\eta}{\eta-1}\frac{p_{i,jt}^{\frac{1}{\eta-1}}p_{i,jt}^{\frac{\eta}{\eta-1}}}{(p_{i,jt}^{\frac{\eta}{\eta-1}}+p_{-i,jt}^{\frac{\eta}{\eta-1}})^2}\right]$$

$$=\frac{\eta s_{i,jt}}{\eta-1}\left[\frac{1-s_{i,jt}}{p_{i,jt}}\right]$$

More algebra (specifically doing an analogous step to the one done from the second to the third line to expression premultiplying the marginal cost) from this step will indicate that:

$$p_{i,jt}=\left[\frac{1-\eta s_{i,jt}}{\eta(1-s_{i,jt})}\right]\frac{w_t}{z_{i,jt}} \tag{31}$$

Where $s_{i,jt}=\frac{p_{i,jt}y_{i,jt}}{Y_t}=\frac{p_{i,jt}^{\frac{\eta}{\eta-1}}}{p_{i,jt}^{\frac{\eta}{\eta-1}}+p_{-i,jt}^{\frac{\eta}{\eta-1}}}$ is the share of firm i in industry j.

One also sees that the ratio of relative prices is an increasing function of the ratio of the productivity levels of the two firms, that is that:

$$\frac{p_{i,jt}}{p_{-i,jt}}=\frac{1-\eta s_{i,jt}}{1-\eta(1-s_{i,jt})}\frac{s_{i,jt}}{1-s_{i,jt}}\frac{z_{-i,jt}}{z_{i,jt}}$$

This expression implicitly characterizes the choice of $p_{i,jt}$ given $p_{-i,jt}$, which is to say that it is analogous to the standard reaction function in static competition games. Notice also that the ratio of $z_{i,jt}$ is the only thing that will change the optimal choice of price for each firm. Rewriting this expressly, we have that:

$$\frac{p_{i,jt}}{p_{-i,jt}} = \frac{1-\eta s_{i,jt}}{1-\eta(1-s_{i,jt})} \frac{p_{i,jt}^{\frac{\eta}{\eta-1}}}{p_{-i,jt}^{\frac{\eta}{\eta-1}}} \frac{z_{-i,jt}}{z_{i,jt}}$$

$$p_{i,jt} = \frac{p_{-i,jt}^{\frac{\eta}{\eta-1}} + (1-\eta) p_{i,jt}^{\frac{\eta}{\eta-1}}}{p_{i,jt}^{\frac{\eta}{\eta-1}} + (1-\eta) p_{-i,jt}^{\frac{\eta}{\eta-1}}} \left(\frac{p_{i,jt}^{\frac{\eta}{\eta-1}}}{p_{-i,jt}^{\frac{1}{\eta-1}}} \right) \frac{1}{\lambda^m}$$

$$p_{-i,jt} = \frac{1-\eta s_{-i,jt}}{1-\eta(1-s_{-i,jt})} \frac{p_{-i,jt}^{\frac{\eta}{\eta-1}}}{p_{i,jt}^{\frac{1}{\eta-1}}} \lambda^m$$

The production labor demand of firm i in industry j is given by:

$$l(m_{i,jt}) = \frac{y_{i,jt}}{z_{i,jt}} = \frac{\eta s_{i,jt}(1-s_{i,jt})}{1-\eta s_{i,jt}} \times \frac{Y_t}{w_t} \tag{32}$$

Hence, profits which are simply $\left(p_{i,jt} - \frac{w}{z_{i,jt}}\right) y_{i,jt}$ are given by:

$$\pi(m) = \left(\frac{(1-\eta) s_{i,jt}}{1-\eta s_{i,jt}} \right) Y_t \tag{33}$$

A.4 Value Functions: Special Cases

The value function of both firms when the industry state is 0 are given by:

$$\begin{aligned} r_t V_t(0, x_l, x_f) - \dot{V}_t = & \left[\pi(0) - w(x_l, x_f; X) x_l\right] + \\ & \epsilon x_l^{\gamma} \left[V_t(1, x_l, x_f) - V_t(0, x_l, x_f)\right] + \\ & \epsilon x_f^{\gamma} \left[V_t(-1, x_f, x_l) - V_t(0, x_l, x_f)\right] + \\ & \nu \left[W_t(0, x_f) - V_t(0, x_l, x_f)\right] + \\ & \nu \left[V_t(0, x_l, g(0, x_l)) - V_t(0, x_l, x_f)\right] + \\ & \frac{\delta}{2} \left[0 - V_t(0, x_l, x_f)\right] \end{aligned}$$

And when inputs are freely adjustable, the ensuing adjustment value W is given by:

$$\begin{aligned} r_t W_t(0, x_f) - \dot{W}_t = \max_x \left[\pi(0) - w(x, x_f; X)x\right] + \\ \epsilon x^\gamma \left[V_t(1, x, x_f) - W_t(0, x_f)\right] + \\ \epsilon x_f^\gamma \left[V_t(-1, x_f, x) - W_t(0, x_f)\right] + \\ \nu \left[V_t(0, x, g(0, x)) - W_t(0, x_f)\right] + \\ \frac{\delta}{2}\left[0 - W_t(0, x_f)\right] \end{aligned}$$

Symmetry ensures that relabeling x_l to x_f generates the other firm's value.

The other special case is when $\bar{m} = 0$. In this case, the leader's value is given by:

$$\begin{aligned} r_t V_t(\bar{m}, x_l, x_f) - \dot{V}_t = \left[\pi(\bar{m}) - w(x_l, x_f; X)x_l\right] + \\ \epsilon x_f^\gamma \left[\phi V_t(\bar{m} - 1, x_l, x_f) + (1 - \phi)V_t(0, x_l, x_f) - V_t(\bar{m}, x_l, x_f)\right] + \\ \nu \left[W_t(\bar{m}, x_f) - V_t(\bar{m}, x_l, x_f)\right] + \\ \nu \left[V_t(\bar{m}, x_l, g(-\bar{m}, x_l)) - V_t(\bar{m}, x_l, x_f)\right] \end{aligned}$$

The value of the leading firm at adjustment is thus:

$$\begin{aligned} r_t W_t(\bar{m}, x_f) - \dot{W}_t = \max_x \left[\pi(\bar{m}) - w(x, x_f; X)x\right] + \\ \epsilon x_f^\gamma \left[\phi V(\bar{m} - 1, x, x_f) + (1 - \phi)V_t(0, x, x_f) - W_t(\bar{m}, x_f)\right] + \\ \nu \left[V_t(\bar{m}, x, g(-\bar{m}, x)) - W_t(\bar{m}, x_f)\right] \end{aligned}$$

The laggard values in this scenario are given by:

$$
\begin{aligned}
r_t V_t(-\bar{m}, x_l, x_f) - \dot{V}_t = &[\pi(-\bar{m}) - w(x_l, x_f; X)x_l] + \\
&\epsilon x_f^{\gamma} \left[\phi V_t(-\bar{m}+1, x_l, x_f) + (1-\phi)V_t(0, x_l, x_f) - V_t(-\bar{m}, x_l, x_f)\right] + \\
&\nu \left[W_t(-\bar{m}, x_f) - V_t(-\bar{m}, x_l, x_f)\right] + \\
&\nu \left[V_t(-\bar{m}, g(\bar{m}, x_f), x_f) - V_t(-\bar{m}, x_l, x_f)\right] + \\
&\delta \left[0 - V_t(-\bar{m}, x_l, x_f)\right]
\end{aligned}
$$

The value of the leading firm at adjustment is thus:

$$
\begin{aligned}
r_t W_t(-\bar{m}, x_l) - \dot{W}_t = &\max_x \left[\pi(-\bar{m}) - w(x, x_f; X)x\right] + \\
&\epsilon x^{\gamma} \left[\phi V(-\bar{m}+1, x_l, x) + (1-\phi)V_t(0, x_l, x) - W_t(-\bar{m}, x_l)\right] + \\
&\nu \left[V_t(-\bar{m}, g(\bar{m}, x), x) - W_t(-\bar{m}, x_l)\right] + \\
&\delta \left[0 - \bar{W}_t(-\bar{m}, x_l)\right]
\end{aligned}
$$

A.5 Normalized Value Functions

To prove that the normalization along the BGP is correct, we proceed via a guess and verify approach. Formally, the claim is that any policy solving V and W necessarily solves $\mathcal{V}$ and $\mathcal{W}$ as well. Beginning from the definition of W, we have that:

$$
\begin{aligned}
r_t W(m, x_f) - \dot{W}_(m, x_f) = &\max_x \left[\pi(m) - w(x, x_f; X)x\right] + \\
&\epsilon x^{\gamma} \left[V_t(m+1, x, x_f) - W(_t m, x_f)\right] + \\
&\epsilon x_f^{\gamma} \left[\phi V_t(m-1, x, x_f) + (1-\phi)V_t(0, x, x_f) - W(m, x_f)\right] + \\
&\nu \left[V(m, x, g(-m, x)) - W(m, x_f)\right]
\end{aligned}
$$

Using goods market clearing, we have that $C_t = Y_t$, which implies that (17) can be rewritten as:

$$w(x_{ij}, x_{-ij}; X)x_{ij} = \kappa Y_t X_t^{-\frac{1}{\theta}} x_j^{\frac{1}{\theta}-\frac{1}{\psi}} x_{ij}$$

Dividing both sides by Y_t[21], we can then write:

$$\begin{aligned} r_t \mathcal{W}(m, x_f) - \frac{\dot{W}(m, x_f)}{Y_t} = \max_x \left[\hat{\pi}(m) - \hat{w}(x, x_f; X)x\right] + \\ \epsilon x^\gamma \left[\mathcal{V}(m+1, x, x_f) - \mathcal{W}(m, x_f)\right] + \\ \epsilon x_f^\gamma \left[\phi \mathcal{V}(m-1, x, x_f) + (1-\phi)\mathcal{V}(0, x, x_f) - \mathcal{W}(m, x_f)\right] + \\ \nu \left[\mathcal{V}(m, x, g(-m, x)) - \mathcal{W}(m, x_f)\right] \end{aligned}$$

Substituting into the left hand side using the Household Euler (16), we then recover:

$$\begin{aligned} (\rho + \frac{\dot{Y}_t}{Y_t})\mathcal{W}(m, x_f) - \frac{\dot{W}(m, x_f)}{Y_t} = \max_x \left[\hat{\pi}(m) - \hat{w}(x, x_f; X)x\right] + \\ \epsilon x^\gamma \left[\mathcal{V}(m+1, x, x_f) - \mathcal{W}(m, x_f)\right] + \\ \epsilon x_f^\gamma \left[\phi \mathcal{V}(m-1, x, x_f) + (1-\phi)\mathcal{V}(0, x, x_f) - \mathcal{W}(m, x_f)\right] + \\ \nu \left[\mathcal{V}(m, x, g(-m, x)) - \mathcal{W}(m, x_f)\right] \end{aligned}$$

Now observing that $\dot{\mathcal{W}} = \left(\frac{\dot{W}}{Y}\right) = \frac{\dot{W}_t}{Y_t} - \frac{W_t \dot{Y}_t}{Y_t^2}$, we conclude that

[21] Which is acceptable because Y_t is continuous in t.

$$\rho\mathcal{W}(m,x_f) - \dot{\mathcal{W}}(m,x_f) = \max_x \left[\hat{\pi}(m) - \hat{w}(x,x_f;X)x\right] + \epsilon x^{\gamma}\left[\mathcal{V}(m+1,x,x_f) - \mathcal{W}(m,x_f)\right] + \epsilon x_f^{\gamma}\left[\phi\mathcal{V}(m-1,x,x_f) + (1-\phi)\mathcal{V}(0,x,x_f) - \mathcal{W}(m,x_f)\right] + \nu\left[\mathcal{V}(m,x,g(-m,x)) - \mathcal{W}(m,x_f)\right]$$

Since in steady state, the time derivatives are zero, this expression reduces to:

$$\rho\mathcal{W}(m,x_f) = \max_x \left[\hat{\pi}(m) - \hat{w}(x,x_f;X)x\right] + \epsilon x^{\gamma}\left[\mathcal{V}(m+1,x,x_f) - \mathcal{W}(m,x_f)\right] + \epsilon x_f^{\gamma}\left[\phi\mathcal{V}(m-1,x,x_f) + (1-\phi)\mathcal{V}(0,x,x_f) - \mathcal{W}(m,x_f)\right] + \nu\left[\mathcal{V}(m,x,g(-m,x)) - \mathcal{W}(m,x_f)\right]$$

Analogous algebra for V then shows that the consistency is preserved. These two expressions together then deliver the normalizations that are used in the paper.

A.6 Aggregation Results

In this section, I derive the aggregation results used in the main body of the paper to characterize aggregate output Y_t, the (production) labor share ω_t, and the steady state growth rate g.

I begin by deriving the expression for aggregate output Y_t.

$$
\begin{aligned}
Y_t &= \exp\left\{\int_0^1 \frac{1}{\eta}\ln\left(y_{ij,t}^{\eta}+y_{-ij,t}^{\eta}\right)dj\right\} \\
&= \exp\left\{\int_0^1 \frac{1}{\eta}\ln\left(\left(\left[\frac{z_{ij,t}}{\omega_t}\right]\frac{\eta s_{i,jt}(1-s_{i,jt})}{1-\eta s_{i,jt}}\right)^{\eta}+\left(\left[\frac{z_{-ij,t}}{\omega_t}\right]\frac{\eta s_{i,jt}(1-s_{i,jt})}{1-\eta s_{-i,jt}}\right)^{\eta}\right)dj\right\} \\
&= \exp\left\{\int_0^1 \frac{1}{\eta}\ln\left(\left(\left[\frac{z_{ij,t}}{\omega_t}\right]\frac{\eta s_{i,jt}(1-s_{i,jt})}{1-\eta s_{i,jt}}\right)^{\eta}+\left(\left[\frac{z_{ij,t}\lambda^{-m}}{\omega_t}\right]\frac{\eta s_{i,jt}(1-s_{i,jt})}{1-\eta s_{-i,jt}}\right)^{\eta}\right)dj\right\} \\
&= \exp\left\{\int_0^1 \frac{1}{\eta}\ln\left(\left[\frac{z_{ij,t}}{\omega_t}\right]^{\eta}\underbrace{\left[\left(\frac{\eta s_{i,jt}(1-s_{i,jt})}{1-\eta s_{i,jt}}\right)^{\eta}+\left(\lambda^{-m}\frac{\eta s_{i,jt}(1-s_{i,jt})}{1-\eta(1-s_{i,jt})}\right)^{\eta}\right]}_{\equiv\mathcal{S}(j)}\right)dj\right\} \\
&= \exp\left\{\int_0^1 \ln(z_{ij,t})-\ln(\omega_t)+\frac{1}{\eta}\ln(\mathcal{S}(j)dj\right\} \\
&= \underbrace{\exp\left\{\int_0^1 \ln(z_{ij,t})dj\right\}}_{\equiv Z_t}\exp\left\{-\ln(\omega_t)\right\}\exp\left\{\frac{1}{\eta}\int_0^1\ln(\mathcal{S}(j))dj\right\} \\
&= \frac{Z_t}{\omega_t}\exp\left\{\frac{1}{\eta}\int_0^1\ln(\mathcal{S}(j)dj)\right\}
\end{aligned}
$$

To derive the expression for the production labor share $\omega_t \equiv \frac{\bar{w}_t}{Y_t}$, we make use of the production labor market clearing condition.

$$
\begin{aligned}
\bar{L} \equiv 1 &= \int_0^1 l_{ijt}^* + l_{-ijt}^* dj \\
&= \int_0^1 \frac{\eta s_{i,jt}(1-s_{i,jt})}{1-\eta s_{i,jt}}\left[\frac{1}{\omega_t}\right]+\frac{\eta s_{i,jt}(1-s_{i,jt})}{1-\eta(1-s_{i,jt})}\left[\frac{1}{\omega_t}\right]dj \\
\omega_t &= \int_0^1 \frac{\eta s_{i,jt}(1-s_{i,jt})}{1-\eta s_{i,jt}}+\frac{\eta s_{i,jt}(1-s_{i,jt})}{1-\eta(1-s_{i,jt})}dj
\end{aligned}
$$

To derive the expression for the steady state growth rate, we first take logs of the expression for Y_t, recovering

$$
\ln(Y_t) = \ln(Z_t) - \ln(\omega_t) + \frac{1}{\eta}\int_0^1 \ln(S(j))dj \tag{34}
$$

Taking time derivatives and observing that along the balanced growth path, the steady state distribution of productivity gaps is fixed and the production labor share is constant, one immediately sees that the growth rate of output is entirely characterized by the growth rate of the productivity frontier, so that

$$\ln(\dot{Y}_t) = \ln(\dot{Z}_t)$$

Now observe that

$$\begin{aligned}
\ln(Z_t) &= \ln\left(\exp\left\{\int_0^1 \ln(z_{ij,t})dj\right\}\right) \\
&= \int_0^1 \ln(z_{ij,t})dj \\
\ln(\dot{Z}_t) &= \ln(\lambda)\left[\int_{\Omega_{x_l}}\int_{\Omega_{x_f}} \left(\epsilon x_l^{\gamma} + \epsilon x_f^{\gamma}\right)\ \mu(0, x_l, x_f)dx_l dx_f + \sum_{m=1}^{\bar{m}}\int_{\Omega_{x_l}}\int_{\Omega_{x_f}} \epsilon x_l^{\gamma}\ \mu(m, x_l, x_f)dx_l dx_f\right]
\end{aligned}$$

The growth rate of the productivity frontier is the average rate at which innovations arrive for the leading firm. When the industry is leveled, both firms are "leading" firms, and hence either generates an improvement in the frontier (which again is the measure of aggregate TFP in this economy) when they innovate.

To calculate steady state welfare from time 0, one departs from the instantaneous preferences of the household at time 0. One then uses goods market clearing and imposes the steady state conditions, before substituting in for steady state output using Equation (34). The remaining step follows from the fact that one can always express output along the balanced growth path in the form $Y_t = Y_0 e^{\bar{g}t}$.

$$
\begin{aligned}
\mathbf{W} &= \int_0^\infty e^{-\rho t} \left[\ln(C_t) - \kappa X_t\right] dt \\
&= \int_0^\infty e^{-\rho t} \left[\ln(Y_t) - \kappa X_t\right] dt \\
&= \int_0^\infty e^{-\rho t} \left[\ln(\bar{Y}) - \kappa \bar{X}\right] dt \\
&= \int_0^\infty e^{-\rho t} \left[\bar{g}t - \ln(\bar{\omega}) + \frac{1}{\eta}\int_0^1 \ln(\mathcal{S}(j))dj - \kappa\bar{X}\right] dt \\
&= \frac{1}{\rho}\left[\frac{1}{\eta}\int_0^1 \ln(\mathcal{S}(j))dj - \kappa\bar{X} - \ln(\bar{\omega})\right] + \frac{\bar{g}}{\rho^2}
\end{aligned}
$$

To derive compensating variation α, we solve the following problem:

$$
\begin{aligned}
\mathbf{W}_2 =& \int_0^\infty e^{-\rho t} \left[\ln(\alpha\bar{C}_1) - \kappa\bar{X}_1\right] \\
=& \frac{1}{\rho}\ln(\alpha) + \frac{1}{\rho}\left[\frac{1}{\eta}\int_0^1 \ln(\mathcal{S}_1(j))dj - \kappa\bar{X}_1 - \ln(\bar{\omega}_1)\right] + \frac{\bar{g}_1}{\rho^2} \\
=& \frac{1}{\rho}\ln(\alpha) + \mathbf{W_1} \\
\mathbf{W}_2 - \mathbf{W}_1 =& \frac{1}{\rho}\ln(\alpha) \\
\alpha =& \exp\left\{\rho\left[\mathbf{W}_2 - \mathbf{W}_1\right]\right\}
\end{aligned}
$$

Compensating variation, defined here as change in the level of output at time 0 that makes the household indifferent between the two consumption sequences thus takes the following form:

$$
\mathbf{C}(\mathbf{W}_2; \mathbf{W}_1) = \exp\left\{\rho\left[\mathbf{W}_2 - \mathbf{W}_1\right]\right\} \tag{35}
$$

A.7 Computation

Solving the model for a given guess of the parameters proceeds in two (nested) steps. The first step consists of solving, for a given guess of the aggregate quantity of research labor X, a representative

industry equilibrium problem. The second step entails simulating the economy wide equilibrium and then examining the invariant distribution implied by the policy functions for consistency with the guess of the aggregate X.

To solve the inner problem, that of an intermediate goods producer, there are two canonical methods. The first is a simultaneous value function iteration, which updates values and policies at each step. The second is a variant of policy function iteration where, *for a given guess* of the policy, the value functions are iterated to convergence before then updating the policies again. Empirically, I have tried both methods and both methods give the same answer, modulo interpolation error. As an initial guess for the policy, I use a decreasing (in the other player's quantity choice) amount of research labor that is invariant across productivity gaps. The initial guesses for the value at adjustment and value while committed are simply the values that would result in the absence of any movement up or down the quality ladder.

Formally, I solve for the industry equilibrium via a Gauss-Jacobi approach where the interpolated policy functions (which are used in the solution of each player's problem) are updated only after one full set of discrete policy updates is completed. I use a tolerance of $1e-5$ for both the policy and the value and then check that $V_i(m, x_{-i}, g_i(m, x_{-i})) \approx W_i(m, x_{-i})$.

1. Guess $g^{*,0}$
2. Generate V^0 and W^0 by iterating to convergence on some V and W using $g^{*,0}$ via mechanical implementation of equations (21) to (24) holding fixed $g^{*,0}$.
3. Given V^0, W^0, and $g^{*,0}$, solve the right hand side of equations (22) and (24). Call the resulting policy $\tilde{g}^{*,1}$.
4. Set $g^{*,1} = \xi\tilde{g}^{*,1} + (1-\xi)g^{*,0}$
5. Update interpolations for g.
6. Update V,W according to equations (21) to (24) once.
7. Update interpolations for V and W.

8. Repeat steps 3 to 7 for the new guess V^1,W^1,$g^{*,1}$ and so on until $|\tilde{g}^{*,N+1} - \tilde{g}^{*,N}| <$ tolerance and $|V^{N+1} - V^N| <$ tolerance.

I simulate the model via an exact method based on inverse transform sampling approaches to sampling time-inhomogeneous Poisson processes. Standard approaches to exact poisson sampling (i.e. using either direct inverse transform sampling or thinning) do not work in this setting. The reason is that the ex-post arrival rates vary at random times to *arrival rates that depend on what the other player's arrival rate is at that point in time*. The algorithm works by updating the processes jointly, one at a time after recursively determining what is the first shock to occur.

To understand the process, it is useful to recall some basic facts about Poisson processes. The first fact to recall is that if a Poisson process is distributed with rate parameter ν, the distribution of expected arrival times has CDF given by:

$$F(t) = 1 - e^{-\nu t} \tag{36}$$

Hence, the natural way to sample arrival times is via inverse transform. This entails sampling from a uniform $(0, 1)$, and then inverting Equation (36) for t, so that:

$$t = -\frac{\ln(1 - U)}{\nu}$$

The independent increments property of Poisson processes then implies that one can continue this process at each event time to generate a sequence of arrival shocks. It is helpful to keep the simple time-homogeneous Poisson in mind, as the core idea when sampling more complicated processes is essentially the same. The natural next case to consider is that of a single time-inhomogeneous Poisson process. In that case, the Poisson process is characterized by its intensity measure so that the distribution of arrival times of that Poisson process at time t are given by:

$$F(t'|t) = 1 - e^{-\int_t^{t'} \lambda(s) ds}$$

Where $\lambda(s)$ is the instantaneous arrival rate of the Poisson process at time s and the object

$\int_t^{t'} \lambda(s)ds$ is referred to as the intensity measure of the process and denoted by $\Lambda(t, t')$. As with the time homogeneous case, inverse transform sampling provides a natural approach to generate arrival times, although this process is substantially more complicated because the inversion requires the arrival time t' which appears in the integral. Fortunately, in my setting, the instantaneous rate function $\lambda(s)$ has the feature that it is piece-wise constant. This means that $\int_t^{t'} \lambda(s)ds$ can be easily computed in a recursive manner. To see how to do this, it is instructive to consider an example. Consider a rate function given by:

$$\lambda(s) = \begin{cases} \lambda_1 & s \in [0, 1) \\ \lambda_2 & s \in [1, \infty) \end{cases}$$

Consider the following algorithm:

1. Draw $U\,\mathcal{U}(0, 1)$
2. Compute
$$t_1 = -\frac{\ln(1 - U)}{\lambda_1}$$
3. If $t_1 > 1$, update
$$t_1 = -\frac{\ln(1 - U) + (\lambda_1 - \lambda_2)}{\lambda_2}$$

Although the entirety of the intensity measure $\Lambda(s)$ can be computed in closed form here, in my setting, the full structure of the rate function $\lambda(s)$ is a function of the sequence of draws, and thus cannot be. To handle this setting, one must observe that the intensity measure keeps track of the average arrival rate over time. This average arrival rate can be computed as a weighted mean of the average arrival rate prior to the current adjustment, and the current arrival rate, with weights given by the relative time in each of the intervals.

There are five sets of shocks to keep track of here.[22] The first one is the quick catchup shock,

[22] I treat the δ shocks as purely behavioral, effectively assuming that whenever the follower is replaced, a deep pocketed investor enters the market and buys all the assets at the market price. This allows me to avoid tracking them, but the algorithm can be easily modified to account for them.

conditional on getting an innovation. This can directly be precomputed by drawing $\mathcal{U}(0,1)$ shocks of sufficient length N_T (which should be chosen so that the collection of shocks will not be exhausted prior to T). The remaining four shocks are outcomes of the coupled Poisson processes. Two are exogeneous and with constant rate ν - these are the Poisson arrival rates of adjustment shocks. The arrival times of these shocks can be computed directly using the fact that the arrival time of a Poisson process with parameter ν follows exponential distribution with (rate) parameter ν. This, in addition to the independent increments property of the Poisson process means we can simply generate a sequence of exponential shocks and cumulatively sum them in order to recover the adjustment times of each firm up to T. The other two Poisson processes have time varying rates where the rates are not known prior to running the simulation. As with the time invariant Poisson processes, these will also have a waiting time distributed exponential, but where the prior distributions had an exponential distribution with (rate) parameter ν, these distributions will be characterized by an intensity measure $\Lambda(t)$.[23] To handle these, I precompute a set of $\mathcal{U}(0,1)$ shocks, which will then be inverted in an appropriate manner to recover the arrival times.[24]

Denote the quantity of research input held by firm i as x_i, the average arrival rate of firm i over a certain time period as $\bar{a}_i$. Denote the time of the last innovation of firm i as $\underline{t}_{I,i}$ and denote the time of last adjustment for firm i as $\underline{t}_{A,i}$.

Collect these times as a vector

$$\underline{\mathbf{t}} \equiv [\underline{t}_I^1,\ \underline{t}_I^2,\ \underline{t}_A^1,\ \underline{t}_A^2]$$

Next, fix a vector of approximate event times

$$\mathbf{t} \equiv [t_I^1,\ t_I^2,\ t_A^1,\ t_A^2]$$

We also need to track

[23] A time varying Poisson process is characterized by its intensity measure $\Lambda(t) \equiv \int_0^t \lambda(s)ds$. The distribution of arrival times, conditional on starting at 0, with then by given by $F(t) = 1 - \exp\{-\int_0^t \lambda(s)ds\}$.

[24] Note that ones does not need to make assumptions on the correlation between the draws ex-ante. The correlation between the arrival times will be fed into the algorithm directly through the inversion.

$$\bar{\mathbf{a}} \equiv [\bar{a}_1,\ \bar{a}_2]$$

$$\mathbf{x} \equiv [x_1,\ x_2]$$

The index of the leader and the productivity difference between the two firms

$$l_i \in \{1, 2\}$$

$$m \in \{0, 1, \ldots, \bar{m}\}$$

And finally a set of indices corresponding to where in the collection of pre-generated shocks we are:

$$\mathbf{i} \equiv [i_{I,1}\ i_{I,2}\ i_{A,1}\ i_{A,2}\ i_{QC}]$$

With the description of the objects in hand, we now turn to the description of the algorithm itself.

In the first step, we initialize $\underline{\mathbf{t}} = \mathbf{0}$, $\bar{\mathbf{a}} = \mathbf{0}$, and $\mathbf{x} = 0.1\iota$. We then set

$$\mathbf{t} = \left[-\frac{ln(U^1_{i_{I,1}}) + (\underline{t}^1_A - \underline{t}^1_I)(\bar{a}_1 - \epsilon x_1^\gamma)}{\epsilon x_1^\gamma}, -\frac{ln(U^2_{i_{I,2}}) + (\underline{t}^2_A - \underline{t}^2_I)(\bar{a}_2 - \epsilon x_2^\gamma)}{\epsilon x_2^\gamma}, \nu^1_{i_{A,1}}, \nu^2_{i_{A,2}} \right]$$

The next step is to find the minimum of these four entries and the associated index. These correspond to the next time an event occurs, as well as the event that occurs.

If the minimum event is the first one, then we first check if the leader index $l_i = 1$. If it is, then we set $m = \min\{m + 1, \bar{m}\}$. If the leader index $l_i = 2$, we then check if $m = 0$. If it is, we then set the leader index $l_i = 1$ and set $m = 1$. If it is not, then we check if $U^{QC}_{i_{QC}} > \phi$. If it is, we then

set $m = 0$ and update $i_{QC} = i_{QC} + 1$, otherwise we set $m = m - 1$ and update $i_{QC} = i_{QC} + 1$. In all cases, we then update $\bar{t}_I^1$ and $\bar{t}_A^1$ to be t_I^1. We then set $\bar{a} = 0$, and update $i_{I,1} = i_{I,1} + 1$.

If the minimum event is the second one, we repeat the same steps above, but with the roles reversed, so that for instance we check if the leader index is 2 and so on.

If the minimum event is the third one, we first update

$$\bar{a}_1 = \frac{\underline{a}_1(\underline{t}_A^1 - \underline{t}_I^1) + (\nu_{i_{A,1}}^1 - \underline{t}_A^1)\epsilon x_1^\gamma}{\nu_{i_{A,1}^1} - \underline{t}_I^1}$$

We then check if the leader index $l_i = 1$. If it is, we then set $x_1 = g(m, x_2)$, else we set $x_1 = g(-m, x_2)$. We then update $\bar{t}_A^1 = \nu_{i_{A,1}}^1$, and update $i_{A,1} = i_{A,1} + 1$. The last step is update

$$t_{I,1} = \underline{t}_I^1 - \frac{ln(U_{i_{I,1}}^1) + (\underline{t}_A^1 - \underline{t}_I^1)(\bar{a}_1 - \epsilon x_1^\gamma)}{\epsilon x_1^\gamma}$$

This final step corrects the arrival time of the innovation shock for the fact that prior to the existing shock, the arrival time itself was changed by the adjustment shock. Note that this can be done because changes to the arrival rate after the adjustment shock can only change the arrival time of the innovation shock, conditional on that waiting time being of sufficient length.[25]

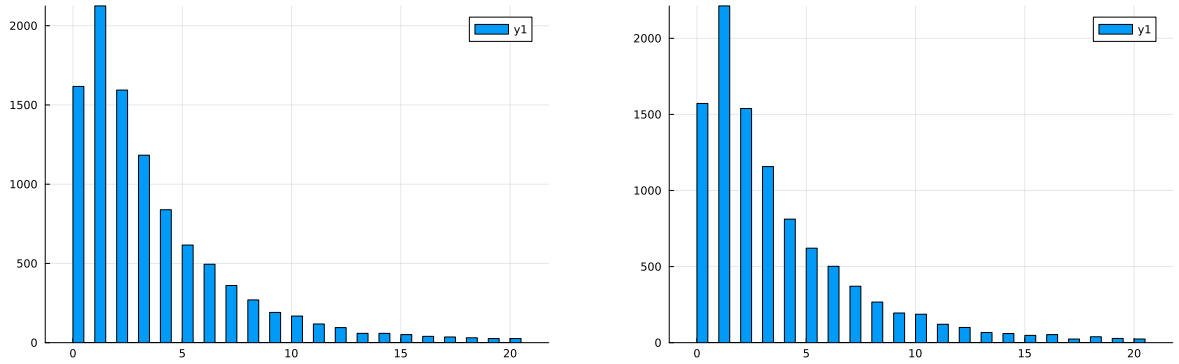
We then repeat until $\min\{\mathbf{t}\} > T$. Because this approach is exact, I reduce error associated with the approximation of a continuous time game with a discrete time one, so that the policy functions in use are exactly the correct ones for the problem.

Notice that at no point have I used information about the death shocks. These death shocks would be important for understanding the age distribution of firms, but I do not use that information anywhere. In essence death shocks in this model function as a kind of behavioral object, whereby the "management" of firms is fired and replaced.

[25]This is a consequence of the intensity measure $\Lambda(t)$ being monotonic in t. Suppose that the adjustment time adjusted the arrival of an innovation prior to the time of the adjustment itself. Then our guess of the time would have been ex-ante correct. The consequence of that is that such a time would necessarily have been chosen when we selected the minimum time, contradicting the claim that the adjustment time occurred prior to the innovation ex-ante.

Table 6: Sensitivity of Moments to 1% Deviations

Parameter	Growth Rate	Average Markup	R&D Share of GDP	Leader Displacement Rate	Variance of Log Wages	ATT on Patent Filing	ATT on Log Wages
λ	30.4%	7.6%	18.4%	3.2%	17.3%	43.1%	63.0%
η	-0.9%	-7.3%	2.7%	-1.6%	14.2%	48.7%	50.1%
ϕ	-0.6%	1.4%	0.3%	-0.4%	11.9%	19.6%	22.7%
ψ	-0.6%	0.3%	0.0%	-0.3%	0.6%	1.2%	1.1%
θ	-0.6%	0.1%	0.1%	0.4%	0.4%	-1.8%	-0.3%
ϵ	1.1%	-0.3%	1.2%	-0.2%	-2.6%	-4.1%	-3.5%
γ	-0.4%	-0.1%	0.7%	-0.9%	1.3%	-1.5%	0.4%



(a) $T = 5000$ (b) $T = 10000$

Figure 13: Productivity gap distribution at two different endpoints.

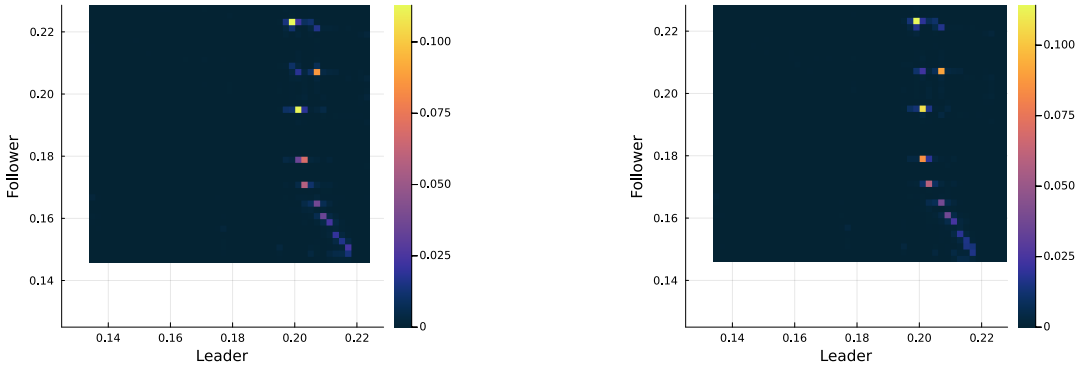
A.8 Sensitivity to Changes in the Estimated Parameters

Table 6 depicts the results of an exercise where I increase each of the estimated parameters by 1%.

A.9 Diagnostic Tests of Convergence

Below I plot unconditional distributions over the distribution of research labor and industry states for two different cutoff times T. I make use of $T = 5000$ in the main text but consider $T = 10000$ for diagnostic purposes here.

One might also worry that these are simply unconditional distributions, but the joint distribution itself might vary in such a way as to escape notice. I evaluate this concern in two ways. The first of these is by directly comparing the calibration moments between the two distributions. The results of that exercise can be seen below in Table 7. The second of these is by computing the average share of the leading firm at each step of the productivity gap distribution, and then plotting the



(a) $T = 5000$ (b) $T = 10000$

Figure 14: Researcher distribution at two different endpoints.

Moment	$T = 5000$	$T = 10000$
Growth Rate	0.8158%	0.8122%
Average Markup	0.2395	0.2403
R&D to GDP	0.02703	0.02696
Sales Leader Change	0.07352	0.07232
Variance of Log Researcher Wages	0.04285	0.04311
ATT on Patent Filing	-0.04016	-0.04071
ATT on Log Wages	0.1067	0.1085

Table 7: Moments At Different Stopping Time Threshold for Ergodic Distribution.

percentage difference between the two times. As can be seen in Figure 15, the deviations in most of the distribution lie substantially below 1%. Even at the edges of the distribution where there is relatively little mass and that are thus more exposed to sampling errors, the deviations lie below 5%. These facts give me confidence that the ergodic distribution is stable.

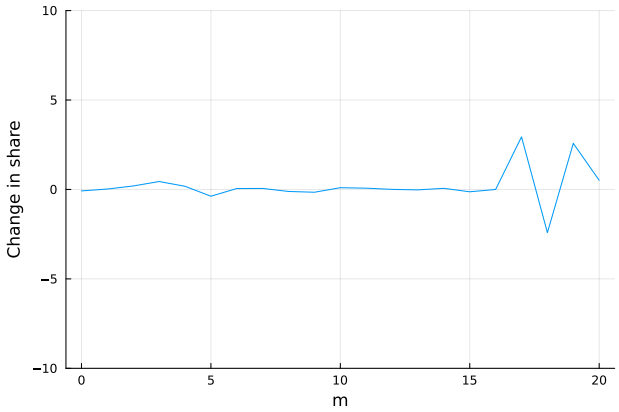


Figure 15: Percentage deviations in leader share of payroll when changing T. This is calculated as a percentage deviation from $T = 5000$ shares.

References

Aghion, Philippe, Christopher Harris, Peter Howitt, and John Vickers (2001) "Competition, imitation and growth with step-by-step innovation," *The Review of Economic Studies*, 68 (3), 467–492.

Aghion, Philippe, Christopher Harris, and John Vickers (1997) "Competition and growth with step-by-step innovation: An example," *European Economic Review*, 41 (3-5), 771–782.

Akcigit, Ufuk and Sina T Ates (2021) "Ten facts on declining business dynamism and lessons from endogenous growth theory," *American Economic Journal: Macroeconomics*, 13 (1), 257–298.

——— (2023) "What happened to US business dynamism?" *Journal of Political Economy*, 131 (8), 2059–2124.

Akcigit, Ufuk, Salomé Baslandze, and Francesca Lotti (2023) "Connecting to power: political connections, innovation, and firm dynamics," *Econometrica*, 91 (2), 529–564.

Akcigit, Ufuk and Nathan Goldschlag (2023) "Where have all the" creative talents" gone? Employment dynamics of us inventors,"Technical report, National Bureau of Economic Research.

Akcigit, Ufuk, Jeremy Pearce, and Marta Prato (2025) "Tapping into talent: Coupling education and innovation policies for economic growth," *Review of Economic Studies*, 92 (2), 696–736.

Altonji, Joseph G and Lewis M Segal (1996) "Small-sample bias in GMM estimation of covariance structures," *Journal of Business & Economic Statistics*, 14 (3), 353–366.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager (2021) "Synthetic Difference-in-Differences," *The American Economic Review*, 111 (12), 4088–4118.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen (2020) "The fall of the labor share and the rise of superstar firms," *The Quarterly Journal of Economics*, 135 (2), 645–709.

Azar, José, Ioana Marinescu, and Marshall Steinbaum (2022) "Labor market concentration," *Journal of Human Resources*, 57 (S), S167–S199.

Azar, José, Ioana Marinescu, Marshall Steinbaum, and Bledi Taska (2020) "Concentration in US labor markets: Evidence from online vacancy data," *Labour Economics*, 66, 101886.

Babalievsky, Fil (2022) "Misallocation in the Market for Inventors."

Berger, David, Kyle Herkenhoff, and Simon Mongey (2022) "Labor Market Power," *American Economic Review*, 112 (4), 1147–1193.

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb (2020) "Are ideas getting harder to find?" *American Economic Review*, 110 (4), 1104–1144.

Crawford, Gregory S., Robin S. Lee, Michael D. Whinston, and Ali Yurukoglu (2018) "THE WELFARE EFFECTS OF VERTICAL INTEGRATION IN MULTICHANNEL TELEVISION MARKETS," *Econometrica*, 86 (3), 891–954, http://www.jstor.org/stable/44955226.

Crouzet, Nicolas, Janice Eberly, Andrea Eisfeldt, and Dimitris Papanikolaou (2022) "Intangible capital, non-rivalry, and growth," *Available at SSRN*.

De Ridder, Maarten (2024) "Market power and innovation in the intangible economy," *American Economic Review*, 114 (1), 199–251.

Ericson, Richard and Ariel Pakes (1995) "Markov-perfect industry dynamics: A framework for empirical work," *The Review of economic studies*, 62 (1), 53–82.

Fernández-Villaverde, Jesus, Yang Yu, and Francesco Zanetti (2024) "Inelastic Supply of Innovators, Monopsony Power, and Creative Destruction."

Garcia-Macia, Daniel, Chang-Tai Hsieh, and Peter J. Klenow (2019) "HOW DESTRUCTIVE IS INNOVATION?" *Econometrica*, 87 (5), 1507–1541.

Hall, Bronwyn H, Adam Jaffe, and Manuel Trajtenberg (2005) "Market value and patent citations," *RAND Journal of Economics*, 36 (1), 16–39.

Hart, Oliver, Jean Tirole, Dennis W Carlton, and Oliver E Williamson (1990) "Vertical integration and market foreclosure," *Brookings papers on economic activity. Microeconomics*, 1990, 205–286.

Jun, Byoung and Xavier Vives (2004) "Strategic incentives in dynamic duopoly," *Journal of Economic Theory*, 116 (2), 249–281.

Klette, Tor Jakob and Samuel Kortum (2004) "Innovating firms and aggregate innovation," *Journal of political economy*, 112 (5), 986–1018.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman (2017) "Technological innovation, resource allocation, and growth," *The Quarterly Journal of Economics*, 132 (2), 665–712.

Maskin, Eric and Jean Tirole (1988) "A theory of dynamic oligopoly, I: Overview and quantity competition with large fixed costs," *Econometrica: Journal of the Econometric Society*, 549–569.

Prato, Marta (2025) "The global race for talent: Brain drain, knowledge transfer, and growth," *The Quarterly Journal of Economics*, 140 (1), 165–238.

Salop, Steven C and David T Scheffman (1983) "Raising rivals' costs," *The American economic review*, 73 (2), 267–271.

——— (1987) "Cost-raising strategies," *The Journal of Industrial Economics*, 19–34.

Von Stackelberg, Heinrich (1934) *Market structure and equilibrium*: Springer Science & Business Media.

Yeh, Chen, Claudia Macaluso, and Brad Hershbein (2022) "Monopsony in the US labor market," *American Economic Review*, 112 (7), 2099–2138.