A Sample JFE Paper

Richard Stanton

U.C. Berkeley

Abstract

There's nothing very interesting here, but the format (achieved using the file `jfe.sty`) makes it suitable for publication in the *Journal of Financial Economics* even if the content doesn't. Here's a nice, informative, single-spaced abstract.

JEL classification: XXX, YYY.

Keywords: LaTeX; papers with no content.

1 Introduction

This paper investigates the interpemporal productivity choice that a firm has to make when deciding whether to make incremental or abrupt product innovations. As firms pick innovation paths, they can either innovate their products incrementally, though there is naturally an exhaustion of incremental possibilities that can be added, or abruptly. In the latter case, innovation brings a radical change to production, which take a productivity toll. However, the invention of a new product (or a new generation of the product) opens a whole horizon for new incremental innovations. Hence, firms have different R&D incentives: right after a radical innovation, firms want to benefit from all the possible new incremental innovation steps that they can make. As they exhaust the realm of incremental ideas, firms turn to abrupt innovation to restart the cycle.

This firm behavior prediction has important aggregate implications. At any given point, some share of firms will be trading current productivity for future one as they choose to engage in abrupt R&D. However, this productivity decline, which should impact the aggregate measure of TFP in the economy, is not related with sluggish efficiency. It is simply the result of an arbitrage decision that would, ideally, pay off in the future.

This paper is linked to the literature on endogenous growth, where creative destruction by firms is the main driver of growth. In this sense, firms engage in R&D both to increase their mark-up and to steal the products of other firms. Although the recent literature on this subject has taken into account the idea of R&D heterogeneity (Akcigit and Kerr, 2016) and diminishing returns of incremental innovation (Acemoglu, Akcigit and Celik, 2014), innovation is always good for productivity, even in the short-term. Moreover, since the focus of Acemoglu et al. (2014) was in the relationship between management and R&D, they do not endogenize the innovation decision. Additionally, to our knowledge, there is no model that accounts for the possibility of firms innovating abruptly internally, which ignores the temporal dimension of the decision to innovate.

On the empirical side, this paper joins the effort of using patent and patent citation data to discipline productivity in an endogenous framework (e.g. Griliches, 1990). The idea behind this identification is that a patent with more citations has a bigger impact on productivity. Using the rich citation distribution available from the empirical work of Hall, Jaffe and Trajtenberg (2001), we discipline our parameters by imposing patent and citation patterns and quantify the impact of each possible innovation path.

The paper is structured as follows: section 2 specifies the databases that were used to estimate the parameters and presents some motivating stylized facts. Section 3 lays down the theoretical framework used to highlight the R&D decisions of firms. Section 4 describes the estimation procedure. Section 5 presents the estimation results and 6 concludes.

2   Databases and Motivating Data

This section provides an overview of the databases that were used in the estimation, as well as some motivating firm-level observations that the theoretical framework proposed in section 3 will highlight.

2.1   Databases

In order to analyze the intrinsic relationship between firm dynamics and R&D behavior through the lens of patent and citation patterns, we will make use of two databases:

- **NBER-USPTO Patent Database:** Data on patents and citations come from

the massive effort undertaken by Hall et al. (2001) in order to count citations of patents granted between 1976-2006. The database has over 3 million patents from the United States Patent and Trademark Office (USPTO), including assignee identification, patent class and citation data. We will limit ourselves to patents granted to US assignees. Moreover, regarding citations, we will use the diffusive correction coefficients from Hall et al. (2001) to adjust for the cutoff effect: as patent applications get closer to the 2006 cutoff, the citation count is, naturally, lower, simply because it did not have enough time to receive citations. The data can also be dynamically matched to the Compustat database;

- **Compustat North American Fundamentals:** Compustat is a database of financial, statistical and market information on active and inactive companies. The database has information over the period 1950-2015 on the US economy. Since privately held firms do not have to disclose financial statements, the database has only publicly-listed firms. However, because those firms tend to be bigger than private ones, they have a significant weight in the whole economy. The main variables are SIC code, employment, net income, net sales, cost of goods sold (COGS), R&D expenditure, plant, property and equipment and total assets. When price deflators are needed, we use the NIPA tables. Moreover, we remove from the data heavily regulated sectors (e.g. utilities), financial firms and observations in which a firm has made acquisitions larger than 5% of the value of their total assets.

2.2 Empirical Observations

The idea of analyzing the intertemporal productivity choice that firms face when deciding whether to continue on the diminishing returns path of incremental innovations or to successfully innovate a product line abruptly, at the expense of a temporary productivity slump, comes from two observations in the data.

First, one should note that, since the 1970s (but more strongly in the 1980s), there has been a surge in the number of what we will denote as *"sprinter firms"*: firms that report negative profits, but grow faster than their entire sector of activity (e.g. Amazon in the early 2000s). Figure 1 highlights the increasing share of sprinter firms in the US economy. This fact is robust if, instead, we weight each company by their employment share. Impressive in itself, this observation is doubly interesting by the fact that those firms invest more in R&D, when comparing median values, than the entire economy. Figure 2 shows the evolution of this ratio for the US economy.

3

This observation is robust if we restrict comparisons by SIC Code.

If we assume that firms, at some point, should start making profits, this means that a significant share of firms are entering in a temporary R&D "sprint race": they increase their innovation effort in order to succeed. The theoretical framework behind this rationale is that, after making an abrupt innovation, a firm is impacted by a productivity decline. However, since a radical innovation opens a new production potential, returns to R&D are bigger, hence companies have an incentive to make R&D efforts.

Secondly, the productivity slump from abrupt innovation is observed at the firm level. Table 3 shows a persistent and significant labor productivity decline 4 years after a firm files an abrupt patent application (taken, here, as the top 10% of patents in number of citations). This is done by regressing firm-level labor productivity growth with lagged dummy variables that capture the (lagged) arrival of an abrupt patent. We also control for firm-level fixed effects and yearly fixed effects that could be related with macro technological shifts. This goes in line with the idea that those type of patents require a major shift in production (e.g. labor, machines), imposing a productivity toll since firms have to rebuild their accumulated knowledge in the new production.

In section 3, we propose a model that will highlight the R&D pattern observed in both of these stylized facts.

3   The Model

The baseline theoretical framework used here is similar to the one employed in Ackigit and Kerr (2016) and Acemoglu et al. (2014). A firm is defined as a portfolio of product lines $\mathscr{F}_j$ as in Klette and Kortum (2004) and can engage in both internal and external R&D. Differently from the previous literature, the goal here is to highlight each firms' choice of investing in incremental R&D, subject to diminishing returns, and internal abrupt innovation, which causes a temporary productivity slump, along the quality ladder of a product. As a simplification, R&D does not scale with firm size.

Because the estimation concerns only the innovation effort of firms, we will limit ourselves to the partial equilibrium.

3.1 Production

We consider a continuous-time setup where preferences are determined over a unique final good $Y(t)$, produced by labor and a continuum of intermediate goods $j \in [0,1]$ with the CES production technology defined by:

$$Y(t) = \frac{L^{\beta}}{1-\beta} \int_0^1 q_j^{\beta}(t) k_j^{1-\beta}(t) dj \tag{1}$$

where $q_j(t)$ is the productivity of product j, k_j its quantity and L is the production labor supply, which is supplied inelastically.

Since each product line owned by a firm is independent of the others, prices are set in order to maximize profits in each line. Each patent allows for monopoly pricing in each product line by the leading-edge firm, which faces a constant marginal cost ζ. As a simplification, when there is an abrupt innovation, we will assume that the production of the old generation variety is shut-down, being replaced by the new version, even in the presence of a productivity decline.

As such, according to (1), firms face, for each j, an iso-elastic inverse demand defined as (time arguments were suppressed for clarity):

$$p_j = L^{\beta} q_j^{\beta} k_j^{-\beta} \tag{2}$$

The profit-maximization problem, then, can be written as:

$$\Pi(q_j) = \max_{k_j \geq 0} \left\{ L^{\beta} q_j^{\beta} k_j^{1-\beta} - \zeta k_j \right\} \tag{3}$$

The first-order condition for 3 yields, for each intermediate good j:

$$p_j = \frac{\zeta}{1-\beta} \text{ and } k_j = \left[\frac{1-\beta}{\zeta}\right]^{\frac{1}{\beta}} L q_j \tag{4}$$

Hence, equilibrium profits for each product line are linear in quality q_j:

$$\Pi(q_j) = \beta \left[\frac{1-\beta}{\zeta}\right]^{\frac{1-\beta}{\beta}} L q_j \equiv \pi q_j \tag{5}$$

3.2 Research and Development Dynamics

Firms have profit-incentives to innovate in their own product lines through *internal innovation*, thus increasing productivity. Moreover, firms have incentives to do abrupt innovation on other the products of other firms, "stealing" the product line of another firm by developing a new product generation through *external innovation*.

For internal R&D, firms can invest in *incremental innovation*, thus increasing, with diminishing returns, the productivity of the current product generation, and in *internal abrupt innovation*, "killing" the current generation and resetting the cycle for returns of incremental innovation (meaning a new generation can achieve higher productivity levels) at the expense of reducing its current productivity. Here, differently from other models, the process of creative destruction is limited to the development of a new product generation over the current version belonging to another firm. The entire process is determined endogenously.

This R&D dynamic is summarized in figure 3 and will now be described in detail.

Incremental R&D: Firms can improve their current product $j \in \mathscr{F}_j$, within their current technological generation (e.g. from an iPhone 3G to an iPhone 3GS), by investing $C_{inc,k}$ units of the final good at each innovative step $k \geq 0$. This cost is defined as:

$$C_{inc,k}(\lambda_{inc,k}, q_j) = \xi_{inc}\lambda_{inc,k}^{\psi_{inc}}q_j \tag{6}$$

where $\lambda_{inc,k} \geq 0$ is the desired incremental innovation instantaneous Poisson flow rate for each level k, ξ_{inc} is the incremental R&D scale and ψ_{inc} is the elasticity of innovation effort to R&D expenditures. When successful, the productivity improves such that $q(t + \Delta t) = q(t)(1 + \sigma\alpha^k)$, where $\sigma > 0$ is a constant multiplicative step-size, $\alpha \in [0, 1]$ is the decay rate of doing sequential incremental steps.

Internal Abrupt R&D: Firms can also improve a product by making a generational step (e.g. from an iPhone 3GS to the iPhone 4). In that sense, they spend $C_{int,abr}$ units of the final good, defined as:

$$C_{int,abr}(\lambda_{int,abr}, q_j) = \xi_{abr}\lambda_{int,abr}^{\psi_{abr}}q_j \tag{7}$$

where $\lambda_{int,abr} \geq 0$ is the internal abrupt innovation instantaneous Poisson flow rate, ξ_{abr} is the abrupt R&D scale and ψ_{abr} is the elasticity of innovation effort to R&D expenditures. When successful, the productivity improves such that $q(t + \Delta t) = q(t)\gamma_{int,abr}$, where $\gamma_{int,abr} \in [0, 1]$ is the productivity slump required for a generational evolution.

External R&D: Finally, firms can engage in creative destruction by innovating (abruptly) on the product line of other firms. This is done by investing C_{ext}, defined

6

as:

$$C_{ext}(\lambda_{ext}, \bar{q}_j) = \xi_{abr}\lambda_{ext}^{\psi_{abr}}\bar{q}_j \tag{8}$$

where $\lambda_{ext} \geq 0$ is the instantaneous Poisson flow rate for external R&D and $\bar{q}_j$ is the average productivity in the whole economy. Similar to the internal innovation effort, the subsequent productivity is given by $q(t + \Delta t) = \bar{q}(t)\gamma_{ext}$, $\gamma_{ext} \in [0, 1]$. In this case, the product line joins the portfolio of the innovating firm and the firm losing its product may exit the market if its portfolio is empty.

Patent and Citation Behavior: The proposed innovation dynamic assumes a structure of patent and citation arrival. Firms can move vertically and horizontally in the quality ladder of each product by creating patents. Each generation of a product is owned by a single firm, who owns the original abrupt patent and can innovate over it incrementally. Latter incremental patents are assumed to cite all previous patents within a generation. Finally, the cycle is rewound when that same firm makes an internal abrupt innovation or when other firms or entrants innovate abruptly on that product line. This pattern implies that abrupt patents receive more citations.

3.3 Entrants

Similar to Klette and Kortum (2004), a mass of entrants engage in R&D in order to enter the economy by successfully innovating abruptly on a product line. Entrant R&D cost C_e is defined as:

$$C_e(\lambda_e, \bar{q}_j) = \nu\lambda_e\bar{q}_j \tag{9}$$

where $\lambda_e \geq 0$ is the innovation flow rate and $\nu \geq 0$ is a scale parameter. As such, the value V_0 of being an entrepreneur can be written as the expectation of successfully innovating:

$$rV_0 - \dot{V} = \max_{\lambda_e} \left\{ \lambda_e\left[E_j V(\{\bar{q}_j\gamma_{ext}, 0\}) - V_0\right] - C_e \right\} \tag{10}$$

where r is the interest rate, $V(q_j)$ is the value of owning one product with productivity q_j and $\dot{V}$ is the partial derivative with respect to time.

Free-entry condition requires that:

$$E_j V(\{\bar{q}_j\gamma_{ext}, 0\}) = \nu\bar{q}_j \tag{11}$$

Hence, all product lines are subject to an aggregate rate of creative destruction τ from all n_f firms in the economy and the outside entrants, where:

$$\tau = n_f \lambda_{ext} + \lambda_e \tag{12}$$

3.4 Incumbents

The incumbents value function determines the amount of R&D effort. Given a product portfolio with n_p products, a vector of productivities $\overrightarrow{q}_f \equiv \{q_{f,j_1}, q_{f,j_2}, ..., q_{f,j_{n_p}}\}$ and a vector of the number of incremental steps $\overrightarrow{k}_f \equiv \{k_{f,j_1}, k_{f,j_2}, ..., k_{f,j_{n_p}}\}$, a firm chooses each innovation flow rate so as to maximize:

$rV(\{q_t, k\}) - \dot{V}(\{q_t, k\}) =$

$$\max_{\substack{\lambda_{inc,k} \\ \lambda_{int,abr} \\ \lambda_{ext}}} \left[\begin{array}{c} \sum_{p}^{n_p} \left[\begin{array}{c} \pi n_p - \xi_{inc} \lambda_{inc,k}^{\psi_{inc}} q_t - \xi_{abr} \bar{q}_t \lambda_{int,abr}^{\psi_{abr}} \\ + \lambda_{inc,k} \left[V(\{q_t^{p-}, k^{p-}\} \cup \{q_{t+\Delta t, j_p}, k_{j_p} + 1\}) - V(q_t, k) \right] \\ + \lambda_{int,abr} \left[E_j \left[V(\{q_t^{p-}, k^{p-}\} \cup \{q_{t+\Delta t, j_p}, 0\}) - V(q_t, k) \right] \right] \\ - \tau \left[V(\{q_t, k\} \setminus \{q_t^p, k^p\}) - V(q_t, k) \right] \end{array} \right] \\ + \lambda_{ext,abr} \left[E_j \left[V(\{q_t, k\} \cup \{q_{t+\Delta t}^{p'}, 0\}) - V(\{q_t, k\}) \right] - \xi_{abr} \bar{q}_t \lambda_{ext}^{\psi_{abr}} \end{array} \right] \tag{13}$$

The right-hand side of the value function is quite intuitive: the first four lines are common to all product lines owned by a firm. The first line represents the instant returns minus the R&D costs of doing incremental and abrupt internal innovation. The second line has the expected return flows from doing incremental innovation on product p, whose productivity changes with the single incremental jump. Similarly, the third line describes the return flows of doing internal abrupt R&D. In this case, the productivity change comes with a reset of the number of incremental steps. Lastly for each product line, the forth line represents the effect of creative destruction by another firm, when the incumbent loses one of its products. Equivalently, the last line is due to the fact that firms can also engage in external abrupt innovation themselves, in which case it adds one product to its portfolio by investing in external R&D.

Since the focus of this paper is on the estimation of the citation distribution and the innovation-related parameters, we will refrain from proposing a solution to the value function here.

3.5 Stationary Distributions of Products and Citations

At the steady state, the economy has a stationary distribution of products at each incremental step. Since abrupt patents are cited by all the subsequent incremental ones within each generation of a product, the citation distribution of abrupt patents and the stationary distribution of products follow the same pattern. We will describe the former now.

Denote η_k the share of abrupt patents that have received k citations (which is equivalent of having k incremental arrivals). For each k in all product lines (which we normalize to measure 1), at the steady state, we must have outflows balancing inflows, i.e.:

$$
\begin{array}{llcl}
Level & Inflow & & Outflow \\
k = 0 & \tau + \lambda_{int,abr} & = & \eta_0(\lambda_{inc,0} + \lambda_{int,abr} + \tau) \\
k \geq 1 & \eta_{k-1}\lambda_{inc,k-1} & = & \eta_k(\lambda_{inc,k} + \lambda_{int,abr} + \tau)
\end{array}
$$

For the first citation level, inflows come from the aggregate arrival of internal abrupt patents from all steps and from external creative destruction, while outflows originate from products at level 0 that get an incremental innovation (i.e. a new citation), from internal abrupt innovation of product lines with no citation and from creative destruction. For all the other citation levels k, inflows come from successful incremental R&D from the previous level $k - 1$, while outflows are either a result of patents with k citations receiving a citation from a subsequent incremental development, from abrupt internal innovation at that level or from outside creative destruction.

Solving for all levels gives:

$$
\eta_0 = \left[\frac{\lambda_{int,abr} + \tau}{\lambda_{inc,0} + \lambda_{int,abr} + \tau} \right] \tag{14}
$$

$$
\eta_k = \eta_{k-1} \left[\frac{\lambda_{inc,k-1}}{\lambda_{inc,k} + \lambda_{int,abr} + \tau} \right] \tag{15}
$$

4 Estimation

4.1 Link Between Patent Citations and Theory

Multiple assumptions are needed to relate the various features of the patent citation dataset to our theoretical model:

- Every patent represents a single innovation.

- Abrupt patents will receive more total citations than incremental patents.

- An abrupt patent receives a citation from a patent within its same classification if and only if the citing patent represents an incremental innovation within the abrupt patents technology cluster.

- Internal innovations will receive more self-citations (citations from future patents by that same company) than external innovations.

Assumptions one, two, and four are made in AK2016 based on empirical evidence (albeit of varying degrees). Assumption three is very important to our estimation scheme. It states that every incremental innovation will cite the abrupt patent that started its technology cluster, and that they will always be within the same patent classification. The provides a clear link between the data and the theoretical distribution in equations 14-15. If an abrupt innovations technology cluster had k incremental innovations, that abrupt innovations patent will receive k same-class citations.

Of course, to estimate the model we also need to identify the abrupt patents. Assumption two tells us that the abrupt patents will have more total citations (same-class plus external-class citations), but it does not give us an exact cutoff. In AK2016 they estimate that the upper ten percent of the total citation distribution contains the abrupt patents, so we use this cutoff in our baseline estimation. Later we test the robustness of our estimates to changes in this value.

Using the ten-percent cutoff, we can generate the distribution of same-class citations to abrupt patents (see figure 4). This empirical distribution will be fit to our theoretical distribution from equations 14-15 to estimate the model. Two important features of the empirical distribution is the large mass of patents at zero citations and the large upper tail.

It should be noted that, even if the four assumptions listed above are completely accurate, we would expect some abrupt patents to be wrongly classified as incremental due to our ten percent cutoff. According to our theoretical model, there is a chance that an abrupt innovations technology cluster will be replace very early on by another abrupt innovation, meaning that it will receive very few same-class citations. While the abrupt patent could still receive out of class citations after being replaced (and in fact, we do see many patents with zero same-class citations but many external citations), it could not receive enough to be classified as abrupt. Thus, we expect the empirical distribution in figure 4 to underestimate the amount of abrupt patents

with low numbers of citations. The impact of this potential bias on our estimation is discussed below.

4.2 MLE Estimation

According to equations 14-15, the theoretical citation distribution of abrupt patents only depends on three parameters: α, $\lambda_{inc,0}$, and $\lambda_{abr,int} + \tau$. In addition, the distribution is identical at equal ratios of the arrival rate parameters, so it actually only depends on two independent parameters: α and $\frac{\lambda_{abr,int}+\tau}{\lambda_{inc,0}}$. Intuitively, α governs rate at which the incremental innovation step size decays after multiple incremental innovations (α equal to one implies a constant step size). The value of $\frac{\lambda_{abr,int}+\tau}{\lambda_{inc,0}}$ is determined by the relative amount of incremental compared with abrupt innovations in the entire economy.

Given a set of parameter values, the resulting citation distribution looks like a gamma distribution, with a large upper tail and the possibility of a mode greater than zero. A larger value of $\frac{\lambda_{abr,int}+\tau}{\lambda_{inc,0}}$ means that there are relatively more abrupt innovations, which means that more product lines will be replaced before their originating patents can accrue a large number of citations. Thus, increasing $\frac{\lambda_{abr,int}+\tau}{\lambda_{inc,0}}$ shifts the distribution to the left. Reducing the value of α means that the firms will invest relatively more in incremental innovation early in the product lifecycle (due to the diminishing returns to incremental innovation). For low enough values of α, this can lead to a peak in the distribution at a citation value greater than zero, as firms almost always invest in the first incremental innovation but invest increasingly less in the late ones. Theoretical graphs for various parameter values are shown in figure 5.

We estimate the two parameters using maximum likelihood estimation. Our theoretical model gives a probability mass distribution for every whole number of citations, but the actual data shows fractional values (this is because, to address the fact that newer patents did not have as long of a time to gain citations, we used a measure of expected citations that allowed for non-whole numbers). To address this, we rounded the expected citation variable to the nearest whole number, and then took the likelihood as the theoretical pmf value at that point.

The resulting parameter estimates are shown in figure 6, and the fit is plotted in figure 7. The estimated ratio of incremental to abrupt patents is about 10:1, which is consistent with the fact that we defined the upper ten percent of patents in the citation distribution as being abrupt. The estimate of α is equal to one. This indicates that firms do not face diminishing returns to initial investment. This has

11

very important theoretical implications, which will be discussed later on.

Looking at figure 7, we see that the data fits most of the distribution closely (including the upper tail, which is not pictured), but significantly underestimates the amount of zero-citation patents. We find that, when observations in the upper tail are eliminated, the estimation better fits the zero-citation data (see figure 8), indicating that the model is unable to completely capture both of those features. Now, it may seem fairly ad-hoc to eliminate observations in the upper tail, but there may be a theoretical justification for doing so. Our model, in keeping with AK2016, only allows for two scales of innovation: abrupt and incremental. This assumption seems to be made only for mathematical simplicity; while there is plenty of evidence that some patents are more valuable than others, there seems to be no evidence of a binary distinction. One might expect that patents in the upper tail of our empirical distribution (it had a max of 741 citations) are significantly more productive than the others. However, under our simple binary model, those patents are identical to the others from a productivity standpoint, and the only reason they gained more citations was through chance. Thus, because a more complete model may place the upper-tail observations in a separate classification, it is reasonable to be suspicious of estimation results that are pinned down by the upper tail.

Multiple methods were used to make sure that the parameter values given in figure 6 were in fact those that maximize the log-likelihood function. First, many initial values were used in the optimization algorithm, and it always gave the same result. We also tried fixing the value of α and running an unconstrained optimization to determine $\frac{\lambda_{abr,int}+\tau}{\lambda_{inc,0}}$, and this also gave the same result (the motivation being that the unconstrained algorithm is sometimes less sensitive to initial conditions). Finally, the 3d plot of the negative log-likelihood clearly shows a minimum at our parameter values (see figure 8), and a basin hopping algorithm found no other minimum.

Recall that our empirical distribution likely underestimates the number of abrupt patents with a low number of citations. This implies that our theoretical model may actually be even worse at fitting the zero-citation density than figure 7 would suggest. It also implies that our estimate of $\frac{\lambda_{abr,int}+\tau}{\lambda_{inc,0}}$ may be biased downward. Importantly, however, because lowering α leads to a decrease in the theoretical density at zero, this downward bias is probably not impacting our estimate of α; if the bias was corrected, we would still expect α to be equal to one.

4.3 First Robustness Check: GMM Estimation

Before drawing strong conclusions from our estimation, we ran several robustness checks. These were meant to test the sensitivity of our estimates both to changes in the estimation procedure and to changes in theoretical assumptions. We are mainly interested in seeing if α remains equal to one (no diminishing returns to incremental innovations) and if the model's fit to the data is affected.

First, we re-estimated the parameters using an alternative estimation scheme: Generalized Method of Moments. The adopted strategy was to estimate $\hat{\theta}_{GMM} = (\tau + \lambda_{int,abr}, \lambda_{inc,0}, \alpha)$ by solving equation (16) with the actual and the estimated moments ($m^d(x)$ and $m^e(x|\theta_{GMM})$, respectively, where x is the real data).

$$\mathcal{L}(\theta_{GMM}) = \min_{\theta_{GMM}} e(x|\theta_{GMM})^T W e(x|\theta_{GMM}) \tag{16}$$

where W is the weighting-matrix ($W = I$ here) and $e(x|\theta_{GMM})$ is the moment error function, defined as

$$e(x|\theta_{GMM}) \equiv \frac{m^e(x|\theta_{GMM}) - m^d(x)}{m^d(x)} \tag{17}$$

In order to estimate those 3 parameters over the distribution of citations, a set of 4 moments was used: the average number of citations, the average of the first 5 sequential share ratios (e.g. # of patents with 1 citation/# of patents with 0 citations), the sum of the first two shares and the incremental R&D intensity, defined in equation (18) similarly to Akcigit and Kerr (2016).

$$\texttt{R\&D intensity} = \frac{\beta \xi_{inc} \sum_{k=0}^{\infty} \eta_k \lambda_{inc,k}^2}{share_{inc}\pi} \tag{18}$$

where $\beta = 0.106$ and $\pi = 0.0755$ are calibrated parameters from the general equilibrium and $\xi_{inc} = 0.346$ is the incremental R&D scale (also calibrated).

Although no single parameter can be singled out with one particular moment, each of the proposed moments has an underlying intuition behind it:

- The average number of citations and the R&D intensity: two global measures of mass that, together, allow for the identification of $(\tau + \lambda_{int,abr})$ and $\lambda_{inc,0}$;

- Mean of the initial ratios: given that the initial incremental R&D effort is a few times bigger than the creative destruction $((\tau + \lambda_{int,abr}))$, one would expect

the initial reduction of incremental innovation to be related with the decay parameter α;

- Sum of the first 2 bins: due to the presence of a significant discontinuity at zero citations (after all, "having ideas" is a highly skewed business), this moment adds an extra weight so as to pin down that effect.

The results of the estimation can be seen in Table 1. The overall estimation comes fairly close to matching all the data moments, while the estimated parameters are similar to the ones obtained with the MLE. As before, those results are robust to different initial conditions and to a "basin hopping" search. The estimation plot over the data histogram is available in the Appendix.

Table 1: *Estimated and data moments using the Generalized Method of Moments*

Moment	Data	Model
Mean of citations	10.06	7.39
Mean of first 5 seq. ratios	0.57	0.88
Sum of first 2 bins	0.26	0.22
R&D Intensity	0.068	0.068
Criterion		0.386
$(\hat{\tau} + \hat{\lambda}_{int,abr}, \hat{\lambda}_{inc,0}, \hat{\alpha}) = (0.047, 0.351, 1)$		

To test for robustness on both GMM and MLE, we have also checked the predictions of both models on an untargeted moment: the share of incremental R&D effort over total effort. Results can be seen in table 2. Both models presented a close match to the value observed in the data.

Table 2: *Estimated and actual results for the untargeted share of incremental R&D effort*

Moment	Data	Model
MLE		
Incremental/total	0.90	0.91
GMM		
Incremental/total	0.90	0.88

4.4 Other Robustness Checks

The GMM results suggest that the main features of our MLE estimation are not simply artifacts of that specific estimation strategy. Given our theoretical model, the best fit of the data truly seems to be one where the density at zero is underestimated and where alpha is equal to one.

As far as theoretical assumptions, one of our most arbitrary is the assumption that abrupt patents are those in the upper ten percent of the total citation distribution. While the ten percent value is similar to the one estimated by AK2016, and is used in other papers, those papers use a different theoretical model. Furthermore, there is no defining feature of the citation distribution at the ten percent cutoff (no discontinuity or anything). Considering this, it is necessary to check whether our results are robust to changes in this cutoff.

The MLE estimates for cutoffs of 20% and 5%, alongside the baseline results, are shown in figure 9. We see that the arrival rates move in the direction that you would expect; the relative amount of abrupt patents drops when the cutoff is 5% and increases when the cutoff is 20%. More importantly, we find that α is still equal to one.

Next we focused on the assumption that an incremental patent will always cite the abrupt patent that initiated its technology cluster. This assumption, which we initially made for mathematical simplicity, is inconsistent with the conclusions of AK2016. In their model, they included a certain probability p that an abrupt patent will be cited by subsequent incremental patents, which was estimated to be 75%. Adding this parameter to our model is simple. For a technology cluster of size k, the probability of the abrupt patent having i citations is given by the binomial distribution:

$$Prob_{i,k} = Binom{i, k, p} \tag{19}$$

Thus, given the size distribution η_k from equations 14-15, the citation distribution f_i is given by:

$$f_i = \sum_{k=0}^{\infty} \eta_k * Prob_{i,k} \tag{20}$$

The addition of the probability parameter p gives the model more flexibility. A lower value of p shifts the distribution left.

The MLE estimates for this updated model are presented in table 10 and figure 11. We see that p is equal to one, with the other parameters identical to our baseline model. This seems to validate our original assumption of setting p equal to one.

There are a few caveats, however, with this robustness check. First, as was mentioned above, even in our baseline model there is a problem with the classification of abrupt patents. It is likely that abrupt patents which are displaced early on by another abrupt patent will never gain enough citations to reach our cutoff value. This leads our empirical distribution (shown in figure 2) to underestimate the number of abrupt patents with a low number of citations. When the probability of citing is less than one, we would expect this underestimation to be even greater. This bias could cause us to overestimate the value of p.

While we were able to test the robustness of our estimation to multiple key theoretical assumptions, there are still others that have not been checked. Of these, we think the most in need of support are assumptions 1-2 from the citation section: that abrupt patents will receive a large number of total citations and that same-class citations are those that directly fall within the abrupt patents technology cluster. Intuitively, this seems like it may be a good approximation, and it is similar to other assumptions made in previous Akcigit papers, but we do not have any actual empirical evidence to support it. Considering that our whole estimation strategy rests on those assumptions, further empirical research into their validity is needed.

5   Results

6   Conclusion

References

Acemoglu, Daron, Ufuk Akcigit, and Murat Alp Celik, "Young, Restless and Creative: Openness to Disruption and Creative Innovations," Working Paper 19894, National Bureau of Economic Research 2014.

Aghion, Christopher Harris Peter Howitt Philippe and John Vickers, "Competition, Imitation and Growth with Step-by-Step Innovation," *Review of Economic Studies*, 2001, *68* (3), 467–492.

Akcigit, Ufuk and William R. Kerr, "Growth Through Heterogeneous Innovations," forthcoming 2016.

Griliches, Zvi, "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 1990, *28* (4), 1661–1707.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," NBER Working Papers 8498, National Bureau of Economic Research, Inc 2001.

Klette, Tor Jacob and Samuel Kortum, "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 2004, *112* (5), 986–1018.

A An Appendix

Here's an appendix with an equation. Note that equation numbering continues where it left off in the main body and that the JFE wants the word "Appendix" to appear before the letter in the appendix title. This is all handled in `jfe.sty`.
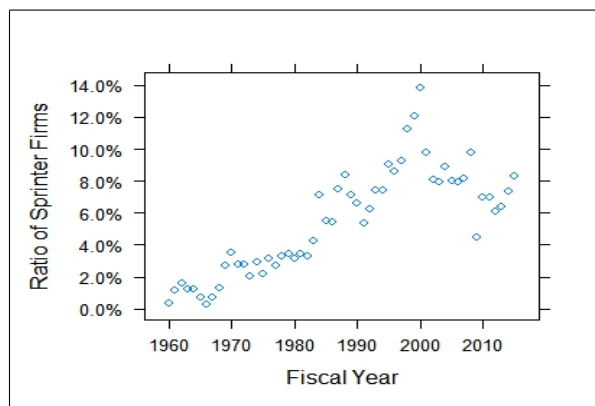

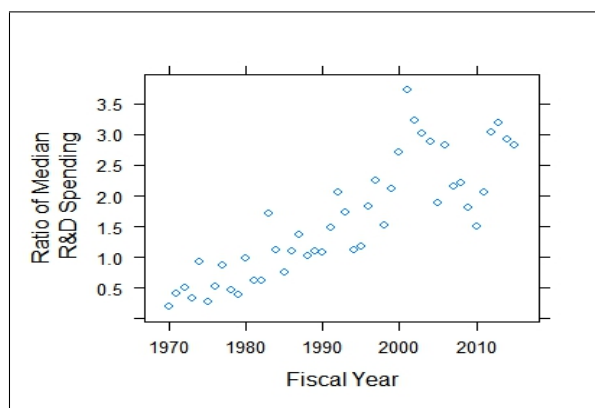*Figure 1: Evolution of the ratio of sprinter firms in the economy.*

Figure 2: Evolution of the ratio of median R&D spending between sprinter firms and the whole economy.

Table 3: Regression Results

	Dependent variable:		
	Labor Productivity		
	(1)	(2)	(3)
Dummy.t	0.005	0.005	0.005
	(0.011)	(0.012)	(0.012)
Dummy.t-1	−0.012	−0.013	−0.013
	(0.012)	(0.012)	(0.012)
Dummy.t-2	0.025*	0.022	0.022
	(0.012)	(0.012)	(0.012)
Dummy.t-3	0.024*	0.021	0.020
	(0.012)	(0.012)	(0.012)
Dummy.t-4	−0.025*	−0.030*	−0.031*
	(0.012)	(0.012)	(0.012)
Dummy.t-5		0.028*	0.026*
		(0.012)	(0.012)
Dummy.t-6			0.013
			(0.013)
Firm fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Observations	38,035	37,549	37,017
R^2	0.0004	0.001	0.001

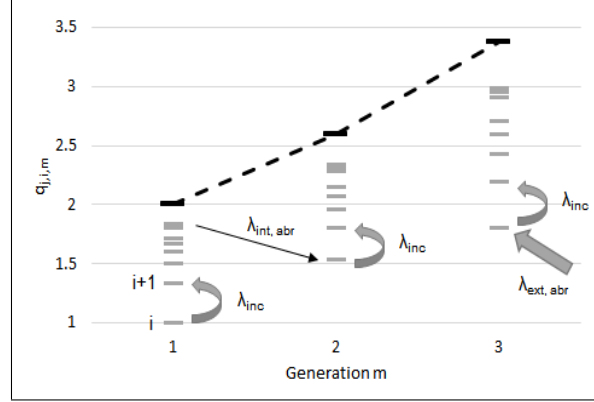*Note:* *p<0.05; **p<0.01; ***p<0.001

Figure 3: Example of sequential innovations in a given product line, for different generations.
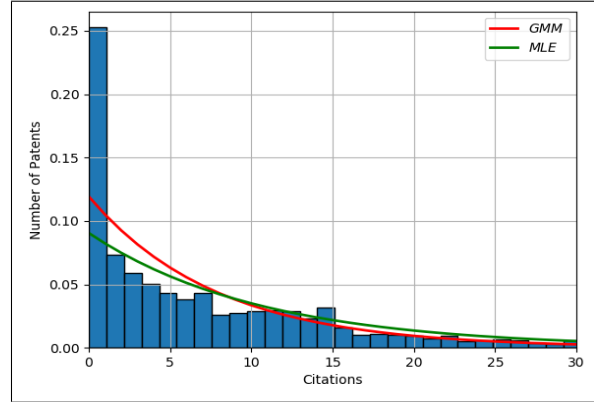

Figure 4: Citation distribution for same-class abrupt patents and estimation results for MLE and GMM.