Firm Size, Innovation Dynamics and Growth*

Ufuk Akcigit†
Massachusetts Institute of Technology

November 26, 2008

JOB MARKET PAPER

Abstract

This paper investigates the relationship between the size of the firm and the quality of innovations of the firm. Much of the previous literature on innovation focuses on innovation frequency with an economy-wide uniform innovation quality. In contrast to the previous literature, this paper allows firms to choose not only the stochastic innovation frequency but also the innovation quality and focuses on how this heterogeneity in innovation quality is affected by the size of the firm. This paper has three distinct contributions:

First, using Compustat firms and their patent applications, I document the following three reduced form facts: *i*) Firm sales growth (both short-run and long-run) is negatively related to the firm size. *ii*) Firm R&D intensity, defined as R&D expenses over sales is negatively related to the firm size. *iii*) The quality of innovation, proxied by the number of citations that a patent receives, is negatively related to the firm size.

Second, I build a tractable general equilibrium growth model that is rich enough to investigate these empirical results. I prove the existence of the equilibrium, characterize its properties and show that the predictions of the theoretical model are consistent with the reduced form evidences mentioned above.

Third, I structurally estimate the theoretical model parameters using Simulated Method of Moments on Compustat firms. Finally, I use these estimated parameters to conduct a macro policy experiment to evaluate the effects of a size-dependent R&D subsidy on different sized firms. In conclusion of this analysis, the optimal size-dependent R&D subsidy policy does considerably better than optimal uniform (size-independent) policy. More interestingly, the optimal (welfare-maximizing) policy provides higher subsidies to smaller firms.

---

*I am particularly indebted to Daron Acemoglu for his insightful guidance. I am also grateful to Philippe Aghion and Mikhail Golosov for their valuable advice. For useful suggestions and comments, I thank George-Marios Angeletos, Suman Basu, Nick Bloom, Ulrich Doraszelski, William Kerr, Josh Lerner, Robert Shimer, MIT Macro Seminar, MIT Macro Lunch and NBER Productivity Group Participants.

†E-mail: uakcigit@mit.edu / Mail Address: Department of Economics, E52-391; Massachusetts Institute of Technology; 50 Memorial Drive; Cambridge, MA 02142.

1 Introduction

In his earlier work, Joseph Schumpeter (1934) claimed that the major source of innovation were small firms operated by wild-spirited entrepreneurs. However, in his later work, Schumpeter (1942) argued that the main innovators in an economy were actually the big firms that posess the required resources for important R&D laboratories and projects. His intriguing theories, also known as the Schumpeter Mark I and Mark II,[1] have raised many policy questions that are still open to debate. In particular, does firm size matter for innovation? Are big firms more R&D intensive as Schumpeter argued in his second theory? What are the implications of these facts for firm growth? Most importantly, how should the optimal R&D policy incorporate these facts? These questions are at the heart of any policy debate which targets technological development through innovation.

As the point of departure, first this paper uncovers the important reduced form facts using Compustat and USPTO patent data. Then, it outlines a theoretical model to understand the microfoundations of the observed reduced form relations. Next, it estimates the structural parameters of the theoretical model using Simulated Methods of Moments (SMM). Finally, this paper conducts a policy experiment of varying R&D subsidy to different sized firms.

The key finding of the reduced form analysis is that smaller firms undertake more radical and original innovations as measured by subsequent patent citations. This fact is also a key property of the theoretical model which furthermore explains other stylized empirical patterns related to the relationship between firm size and firm growth and firm size and R&D intensity. Besides explaining the reduced form facts, this tractable model provides a comprehensive framework to understand the implications of macropolicies on innovative behavior. Hence, this paper concludes its analysis by studying the implications of R&D subsidy policies under various regimes: *i*) uniform (size-independent) R&D subsidy, *ii*) size-dependent two-level R&D subsidy, *iii*) size-dependent linear R&D subsidy. Overall, the paper elaborates on the heterogeneity of innovation quality of different sized firms, documents the related empirical patterns, develops intiution for these results and outlines how to incorporate these important facts into the macro models and policy analysis.

The first reduced form relation considered here is the relationship between firm size and firm growth. The previous literature on firm growth is mainly centered around Gibrat's (1931) claim that the growth rate of a firm is independent of its size. I repeat here a similar exercise with the Compustat dataset in order to generate an independent, but comparable measure of

[1] Nelson and Winter (1982a,b) Kamien and Schwartz (1975, 1982)

the relationship between firm size and growth. Doing so, I also address certain caveats that the previous empirical literature has overlooked. My results tend to corroborate the findings of previous studies which suggested that smaller firms grow faster on average. This effect becomes even sharper once unobserved firm heterogeneity is controlled for. These results are robust to controlling for a survivor bias and for the selection of small firms in the Compustat sample.

The second reduced form relation investigated is that between firm size and R&D intensity, defined as R&D spending over sales. In his seminal book, Joseph Schumpeter (1942) claimed that firm size was crucial for both R&D intensity and innovation suggesting that large firms have a size advantage. The early subsequent empirical studies had mostly supportive conclusions to Schumpeter's theory. However, more recent studies from late 80s and early 90s have utilized larger datasets and addressed several econometric concerns. In the end, these later studies have generally found no systematic relation between R&D intensity and firm size. My results on this question differ significantly from the previous literature. I show that the R&D intensity decreases significantly with firm size, which implies that R&D spending increases less than proportional with firm size. This result is robust, among numerous other specifications, to the sample selection problem of small firms.

The last reduced form analysis addresses the relation between the firm size and innovation quality. Empirical studies that have focused only on innovation frequency, to the detriment of innovation quality, have been inadequate to capture the true effect of innovations on technological progress. One single, major innovation could be much more important for technological progress than many incremental innovations. The main empirical challenge lies in measuring innovation quality. Additional information from patent data can be used to overcome this hurdle. Industrial economists have documented a strong positive correlations between the value of a patent and elements such as the number of citations of that patent by subsequent research, claims and scope.[2] Hence, the number of citations that a patent receives is a good indicator of the quality of an innovation and is therefore used as a proxy for the importance or quality of an innovation in this paper. For robustness, I verify my results with these alternative indicators of patent quality. The reduced form results indicate that the number of citations that a patent receives decreases as the firm size increases. This implies that the firm size is negatively related to innovation quality. The use of the alternative proxies for innovation quality do not dramatically affect the results.

[2] See Lerner (1994) for details.

These three empirical facts constitute the starting point of this study. Next, to understand the microfoundations underlying these results, I build a general equilibrium model with entry and exit. In this model, technologically heterogeneous firms compete for innovations against each other in order to increase operating profits. Each firm is identified by its firm-specific labor productivity which is improved stochastically through R&D and innovation. Every single innovation raises the profits of the firms, by reducing their cost relative to their competitors. A novelty of this model is that firms can endogenously choose not only the probability of innovation, but also the innovation quality which is exogenous in previous papers. I prove the existence and characterize the properties of the equilibrium and show that the model's predictions are consistent with the aforementioned reduced form evidences. My theoretical results rely on two features: 1) The concavity of the profit function and as a result of this, the value function is concave in productivity. 2) There are no strong increasing returns in productivity. Any model with these features would imply that firms with lower productivity will have a greater incentive to increase their productivity. Therefore, they are going to be more innovative, more R&D intensive, choose higher quality innovations and grow faster.

To complete the analysis, I estimate the main structural parameters of the theoretical model using the Simulated Method of Moments (SMM). These parameter estimates allow me to discuss the impact of various macro policies on innovation. In particular, I focus on the effects of a public R&D subsidy for different sized firms by comparing three different regimes; *i*) uniform (size-independent) R&D subsidy, *ii*) size-dependent two-level R&D subsidy (different rates below/above a certain threshold of firm size), *iii*) size-dependent linear R&D subsidy. The results of this analysis document the significant contributions of size-dependent subsidies on welfare. The optimal size-dependent policy provides higher subsidies to smaller firms due to their higher and more influential innovative behavior.

This paper contributes to the empirical literature on firm innovation, R&D and growth. A detailed review of the related empirical literature will be provided in the next section. On the theoretical side, this paper is a contribution to the endogenous growth literature with step-by-step innovation. Previous studies with endogenous R&D decisions have mainly focused on the arrival rate of innovations as a choice of the firm, treating the size of each individual innovation as an exogenous parameter of the model (Aghion, Harris and Vickers, 1997; Aghion Harris, Howitt and Vickers, 2001; Acemoglu and Akcigit, 2008). This theoretical specification leaves out a very important empirical observation; that is, different innovations have different quality and technological impacts. The ultimate goal of the current work is to shed light on this unexplored subject both in terms of theory and empirics.

The results of this paper is also related to the R&D subsidy literature. It is widely argued that the social return to R&D is greater than private return, which suggests that the equilibrium amount of private R&D investment is suboptimal and R&D subsidies could be used to make up for the private underinvestment in R&D (Spence, 1984; Griliches, 1992, 1995; Jones and Williams, 1998). This paper contributes to this literature by introducing a size-dependent R&D subsidy policy and documenting the substantial welfare gain from this policy.

Finally this paper is also a contribution to the recently growing literature on size-dependent policies. This literature typically argues that the size-dependent policies create detrimental effects on the economy by misallocating the resources (Guner, Ventura and Xu, 2008; Klenow and Rodriguez-Clare 1997; Hall and Jones 1999; Caselli 2004, Restuccia and Rogerson, forthcoming). On the other hand, in the context of Intellectual Property Rights Protection, Acemoglu and Akcigit (2008) argues that the gain from size-dependent policies could be substantial. This paper adds to this literature by introducing size-dependent R&D subsidy policy.

The rest of the paper is organized as follows: Section two reviews the related empirical literature and describes the reduced form analysis; section three builds the theoretical model and presents its results; section four does the structural estimation of the model with SMM; section five investigates the size-dependent R&D subsidy policy and section six concludes.

2 Empirical Evidence

The goal of this section is to document some empirical facts about growth, R&D and innovation activity of different sized firms. In particular, the three relations estimated are

- Firm size vs firm growth
- Firm size vs R&D intensity
- Firm size vs innovation size.

2.1 Related Literature

The literature on firm growth has been deeply influenced by Gibrat's law. This theory has however been mostly refuted by empirical evidence, in particular for the US economy. Most studies showed that firm size in the U.S. is negatively correlated with firm growth (Mansfield (1962), Hall (1987), Amirkhalkhali and Mukhopadhyay (1993), Botazzi and Secchi (2003) for quoted US manufacturing firms and Evans (1987a), Gabe and Kraybill (2002) for non-quoted firms). Nevertheless, several studies have still argued that Gibrat's law holds for firms above a certain size threshold (Mowery (1983), Caves (1998)).

Similarly, the early literature on firm level R&D has been inspired by Schumpeter's claim and showed that the R&D spending increases more than proportionally in relation to firm size (Horowitz (1962), Hamberg (1964), Scherer (1965 a,b, 1980), Comanor (1967), Meisel and Lin (1983), Kamien and Schwartz (1982)). However, these studies focused mainly on the 500-1000 largest firms, possibly due to limitations in data availability. In addition, the unobserved sector and firm heterogeneities, which were potential sources of bias, were not taken into account in these studies. Hence, it is not surprising that more recent studies from the late 80s and early 90s, which have utilized relatively larger datasets and controlled for sector level heterogeneity, have contradicted earlier findings. Indeed, they showed that the increase in firm size in relation to R&D is either proportional (Cohen et al (1987)) or slightly less than proportional below a certain firm size threshold (Bound et al 1984).

The innovation patterns of different sized firms have been analysed by a number of studies. Kamien and Schwartz (1975) argue that small firms are more innovative in highly concentrated industries. Mansfield (1968) and Mansfield et al (1971) document that the major innovation per R&D expense is higher for smaller firms. Cooper (1964) claims that small firms have an organizational advantage in innovative activities. Other studies argue that major innovations come from small firms because they make use of the innovation opportunities whereas large firms might supress such opportunities for various reasons (Blair, 1972; Pavitt and Wald, 1971; Kamien and Schwartz, 1975). These empirical studies are based on case studies which prevent us from drawing broader conclusions.

Finally the literature on patents have generated fruitful work for detecting firm level innovations and identifying their heterogenous values. Trajtenberg (1990) shows the positive relation between the citations a patent receives and its social value. Harhoff et al (1996) shows the same relation but for the private value of a patent. Hall et al. (2001) argue that a firm's market value is not positively correlated with patent counts but rather with the "citation weighted patent portfolio", a measure which considers both patent counts and the number of citations per patent. Similarly, Shane and Klock (1997) document a positive relation between patent citations and Tobin's q. In addition, Shane (1999a, 1999b) suggests that the more citations a patent receives, the more likely it is to get licensed. Sampat (1998) and Sampat and Ziedonis (2004) present the positive correlation between number of citations and the licensing revenues. Hence, it seems that the number of citations that a patent receives is a good indicator of the quality of an innovation and is therefore used as a proxy for the importance or quality of an innovation in this paper. In addition, other studies have introduced alternative indicators for innovation quality. For instance, Lanjouw and Schankerman (1997, 1999, 2001) point out that the number of patent claims is a significant predictor of whether the patent will be renewed or litigated. Lerner (1994) shows that an index, proxying for the scope of a patent, is positively correlated with firm value. For robustness, I verify my results with these alternative indicators of patent quality as well.

2.2 Data

The data used is the Standard and Poor's Compustat database which includes information on all publicly trade companies in the US since 1950. Besides the tremendous increase in the sample size, the average firm size in Compustat data has decreased drastically over time (see figure 1) which allows the sample data to capture a wider range of firm sizes. Compared to earlier studies using Compustat, this could potentially make the current sample more representative of the whole population of US manufacturing firms. Following a common practice in the literature, I exclude from the sample non-manufacturing firms or non-domestic firms. [3]

The data on innovation is taken from the NBER/USPTO patent dataset, described in detail in Hall et al. (2001) .It includes all utility patents granted in the USA between 1963 and 2002. By definition, patents grant their holder a monopoly for the use of the innovation. Patents identify "prior art" through *citations* so as to clearly mark the boundaries of that monopoly power. This variable, namely the number of citations will be used to proxy for the innovation size. It is available in the NBER Patent Data set for all patents granted after 1975.

Another important component of the patent data set is the match between patent and Compustat firms (Hall et al, 2001). Even though this match is a great source for linking the patents to firms, it is necessary to be cautious because it includes the Compustat firms that existed in 1989. Firms that were established after 1989 are not included in the patent-match data.

The last two useful variables that I am going to use in this study are the *number of claims* and the *originality index*. The former is the number of "components" or the "main pieces" of the patent, therefore, this number could be an indication of the "scope" or "width" of the innovation (Jaffe and Trajtenberg, 2002, p. 432). The latter is an index calculated for each patent i with the following formula: $Originality_i \equiv 1 - \Sigma_j^{n_i} s_{ij}^2$. In this formula, s_{ij} is the percentage of citations made by patent i to a patent that is in patent class j and n_j is the

[3] that do not have their headquarters in the US.

number of patent classes. This implies that if a patent cites a wider set of patent classes, its originality index is higher.

Table 1 contains the descriptive statistics for the key variables.

2.3 Firm Size vs Firm Growth

2.3.1 Benchmark Specification

To investigate the link between firm growth and firm size, the following regression is estimated using OLS

$$g_{it} = \beta_0 + \beta_1 \ln Sales_{it} + \delta_j + \delta_t + \varepsilon_{ijt} \tag{1}$$

where $g_{ijt} \equiv Sales_{ij(t+1)}/Sales_{ijt} - 1$ is the growth rate of firm i in industry j at time t, $Sales_{ijt}$ is net sales and δ_i and δ_t are respectively the industry and sector fixed effects.[4] This regression does not include the final period's growth rate of the firms that exit the sample. If Gibrat's law holds, β_1 should be zero.

The results are reported in the first column of Table 2. The coefficient β_1 is estimated as -0.037 with a standard deviation of 0.001 which is a considerable departure from Gibrat's law. This is both statistically significant and economically large. An increase of one standard deviation in $\ln Sales$ $(= 0.412)$ decreases the growth rate by 1.5 percentage points. Since the average growth rate is 12 percent, this amounts to a sizeable 13 percent decrease in growth rate relative to the mean. However, simple OLS results may not be reliable enough and further refinements to the estimation are presented next.

Before the discussion of the results continues, it is important to note that the reduced form analysis on R&D intensity and innovation quality will use similar specifications to (1). To avoid repetition, the detailed discussion on the common caveats related to all three specifications will be provided in the first section and only be mentioned briefly in later sections.

2.3.2 Survival Bias

Firms that exit the market do not report any data in Compustat which disables the observation of the decline in $Sales_t$ in period t when the exit happens. In general, the exit rate is higher among smaller firms, so that excluding exiters from the sample biases the results in favor of small firms. To address this problem, there are two different methods.

The first method consists in merely assuming that the output of a firm drops to zero right before the exit, so that the growth rate in their last period is -100 percent. The results of this

[4] All nominal variables are deflated by the GDP deflator in the corresponding year

method are presented in column 2 of Table 2. The parameter estimate has now increased from -0.037 to -0.025 confirming the suspicion that the benchmark OLS is biased downward due to survival bias. A one standard deviation increase in $\ln Sales$ leads to a 20 percent decrease relative to the average growth rate of this sample (0.057). Even though this method has generated both statistically and economically significant estimates, it is likely to overcorrect for the bias, since -100 percent is a lower bound for the actual growth rate in the last period before exit.

The second possible solution is to apply a Heckman two-step selection estimation (Wooldridge, 2000). The selection equation can be written as,

$$z_{it}^{*} = \gamma_0 + \gamma_1 \ln Sales_{it} + \boldsymbol{\gamma}\mathbf{X} + \epsilon_{it} \qquad \text{(Selection Equation)}$$

where z^{*} is a latent variable interpreted as "the propensity to exit", which depends on firm size and some other explanatory variables $\mathbf{X}$. Even though the latent variable is unobserved, a binary variable z_{it} is observable which is defined as:

$$z_{it} = \begin{array}{l} 1 \text{ if } z_{it}^{*} > 0, \text{ (firm remains in the sample)} \\ 0 \text{ if } z_{it}^{*} < 0 \text{ (the firms exits)} \end{array}$$

The outcome equation is then

$$g_{it} = \beta_0 + \beta_1 \ln Sales_{it} + \delta_j + \delta_t + \varepsilon_{ijt} \qquad \text{(Outcome Equation)}$$

The growth rate g_{it} is observed only when $z_{it} = 1$. A consistent estimator of β_1 (1) can then be obtained from the conditional regression

$$\mathbb{E}\left(g_{it} \mid z_{it}^{*} > 0\right) = \beta_0 + \beta_1 \ln Sales_{it} + \beta_2 \lambda_{it} + \delta_j + \delta_t + \varepsilon_{ijt} \tag{2}$$

where

$$\lambda_{it} = \frac{\phi\left(\gamma_0 + \gamma_1 \ln Sales_{it} + \boldsymbol{\gamma}\mathbf{X}\right)}{\boldsymbol{\Phi}\left(\gamma_0 + \gamma_1 \ln Sales_{it} + \boldsymbol{\gamma}\mathbf{X}\right)}$$

is the "Inverse Mills Ratio". The parameters of $\gamma_0, \gamma_1, \boldsymbol{\gamma}$ are obtained with a probit regression. The Heckman two-step model requires an exclusion restriction (instrument) otherwise the model is identified solely on distributional assumptions. The novel instrument used here is the ownership status of the firm. Indeed, firm exits are significantly affected by the ownership status of the firms. Besides other factors, once the parent company is financially constrained, it might decide to sell its subsidiary to a third party through carve-out, spin-off or sell-off (Draho, 2004, p.156). Such transactions would increase the likelihood of getting dropped from the Compustat sample for subsidiaries. This fact makes the ownership status of the firm a novel instrument in the selection equation.

Columns 3a,b and 4 of Table 2 report the estimation results of this two-step selection model. The significance of the inverse Mills ratio confirms the existence of a selection bias. Column 3a shows that the probability of surviving in the sample increases significantly in firm size. In addition, the sign of the coefficient on the ownership status dummy[5] implies that subsidiaries have a lower probability of survival. Column 3b reports the same results in Column 3a as the marginal effects at the sample means. These transformed results suggest that the marginal effect at the sample mean of $\ln Sales$ is 0.008 (standard error=0.000) which implies that the marginal change in $\ln Sales$ from the average increases the probability of surviving in the sample by 0.8 percent. The new, corrected coefficient on firm size in Column 4 is now -0.028 with a standard error of 0.002. As expected, it lies between the simple OLS coefficient and the coefficient obtained when assuming zero output in the last period. This is because, as explained, the OLS estimation favored small firms whereas the latter method favored the large firms. With this new corrected estimate, one standard deviation increase in $\ln Sales$ is associated with a 10 percent decrease in the growth rate, which is economically sizeable. Since the coefficient on firm size remains significantly negative, these results show that the survival bias alone cannot account for the negative relationship between firm size and growth.

2.3.3 Measurement Error

Another common concern for estimating (1) is the errors-in-variables problem, as firm sales are a noisy proxy for firm size. Consider a measurement error in $Sales_{it}$ in period t. This will create a spurious negative relation between firm growth and firm size because it affects the denominator of the dependent variable as well. Therefore the existence of a measurement error would bias the estimate of β_1 downwards. The problem can be solved through an instrumental variables approach, which uses the one year-lagged values of sales, $\ln Sales_{t-1}$ as an instrument for $\ln Sales_t$.

The results reported in column 5 of Table 2 show a less negative coefficient of -0.028 (standard deviation=0.001) on firm size, confirming the presence of some error-in-variables bias. Again, this suggests that one standard deviation increase in $\ln Sales_t$ is associated with 1.2 percentage points decrease in growth rate which corresponds to a 10 percent decrease of the growth rate relative to the sample average. Therefore this implies that errors-in-variables was not the sole driver of this relation.

[5] This dummy becomes 1 when the observation is a subsidiary.

2.3.4 Sample Selection

A final problem is the sample selection bias inherent in the Compustat sample. Suppose for example that most big firms are present in the sample, even though they might have low or medium productivities, but that only the small firms with the highest productivities are admitted in the sample. There are two potential reasons for this. First, small firms might be more productive due to some unique unobserved individual characteristics. Secondly, as part of their life-cycle, at the time that they are selected into the sample, they might be on the increasing side of an inverse U-shaped growth/age trajectory which might lead to high initial growth rate and lower growth rates over time.

These two issues require different approaches. The former concern is related to unobserved firm heterogeneities which can be captured by the fixed-effects regression. Column 6 of Table 2 reports the fixed effects results including time effects. Contrary to the expectations, adding firm fixed effects sharpens the contrast between the growth rates of small and large firms by reducing the coefficient down to -0.175 (standard error=0.003) which is more than 5-fold in absolute terms of the benchmark OLS estimate. This suggests that a one standard deviation increase in firm size is associated with an approximately 61 percent decrease in growth rate relative to the sample average. Even though this result seems to be surprising initially, it shows that avoiding the firm fixed effects generates an omitted variable bias which favors the large firms. These omitted variables are likely to bias the results in favor of large firms, since these are the factors which have made these firms "large firms" at the first place. The very same factors would make the firm grow faster than others in general and not controlling for them would make the firm size take the credit for the additional growth of the large firms. Unfortunately, previous literature has generally overlooked this aspect, possibly due to data limitations.

The second issue regarding the position of the firm on its growth trajectory is more related to the age or the time elapsed since when the firm first shows up in the sample. A way to address this issue is to control for the number of years that the firm has spent in Compustat sample. Column 7 controls for this variable which is denoted by "Age". The point estimate of the coeffcient on Age is -0.003 (standard error=0.000). This significant negative coefficient on Age illustrates that on average, firms are growing faster earlier in their life-cycle. This is also economically significant since this implies that in 10 years (≅one standard deviation of Age, 11.3), the growth rate of a firm with the same beginning and end-of-period sizes will decrease by 25 percent relative to the average growth rate of the sample. On the other hand, the coefficient of the $\ln Sales$ is close to the OLS estimate with a small increase. Controlling for age, the benchmark OLS estimate increases by only 0.007 percentage point to -0.03 (standard error=0.001), which is still highly significant and economically large. The tiny increase suggests that only a small portion of the OLS estimate could be attributed to the sample selection problem of the second kind.

Another strategy to address the second caveat is to focus only on the firms that have been long enough in the sample. For that purpose Column 8 consider the firms that have been in the Compustat sample for more than 10 years. This approach is expected to eliminate the initial heterogeneity of the firms that might have qualified them to go public. The drawback of this approach is the significant decrease in the sample size (around 50%) and reducing the attention only on the healthy firms which manage to stay for more than 10 years in the sample. The average firm size increases by more than 15 folds (from 0.07 to 1.06) eliminating the young and typically small firms. Nevertheless, this method would be suggestive about an upper bound of the actual coefficient estimate on $\ln Sales$. This specification produces an estimate of -0.011 (standard error=0.001) which is about a third of the OLS estimate. This decline was expected as the nature of the sample has changed, yet in spite of this, the result from this restricted sample still indicates a smaller but very significant negative relationship between firm size and growth. Moreover, this negative relation is also economically significant. A one standard deviation increase in $\ln Sales$ ($\cong$2.154) is associated with 2.37 percentage points decrease in the growth rate, which amounts to a 42 percent decrease in growth rate relative to the average growth rate of 5.6 in the sample.

Therefore the conclusion of all these different specifications is that the statistically and economically significant negative relation between firm growth and firm size is robust to sample selection problem.

2.3.5 Additional Caveats

Time-varying Sector Heterogeneities could be a potential issue. The take-off of the computer industry in the mid-80s for example could be attributed to such an unobserved sector level change. In order to prevent the results from being affected by these changes, I introduce an interaction term between the sector and time dummies. Column 9 shows that the impact of firm size on growth is almost identical as in the benchmark OLS estimation, showing that time-varying sector effects do not significantly affect the benchmark results.

Short-Run vs Long-Run. The short-run relationship between firm size and growth might differ from the true long-run relationship , because of transitory and cyclical movements in Sales. Therefore, the 10-year average growth rate of firms between 1995-2005 is regressed on the initial firm size of 1995. This is done first by excluding firms which have exited over the period (Column 10) and then by applying Heckman's 2-step selection method (Columns 11a,b and 12). Column 10 indicates that the long-run results, which includes only the firms that survived for 10 years, are 0.009 percentage point higher than the short-run benchmark OLS. This should be expected since this sample excludes the exiters which are typically smaller firms. The average firm size of the survivors is 0.171 whereas the average firm size of the exiters during this period is -0.599. Since this might cause a survivor bias on the estimate, columns 11-12 employ the Heckman selection method. Column 11a shows that the probability of surviving is increasing in firm size whereas subsidiaries have a lower probability of surviving. Column 11b reports the marginal effects at the sample means. These estimates imply that the marginal deviation from the average firm size increases the probability of surviving by 5 percent. Similarly, being a subsidiary company decreases the probability of surviving by 41 percent. However, the coefficient β_2 of the inverse Mills ratio in (2) is not significant indicating that the survival bias does not affect the OLS results systematically. Column 12 reports the results of the second step. The corrected estimate on $\ln Sales$ is now -0.026 (standard deviation=0.006). This is not very different from the OLS estimate in Column 10 since the survivor bias is not in effect. These alternative specifications indicate that even the long-run relationship between firm size and growth is not consistent with Gibrat's law.

In conclusion, the benchmark regression, together with the robustness checks[6] show the following result.

Fact 1 (Firm Growth) *Firm size is negatively related to firm growth.*

2.4 Firm Size vs R&D Intensity

2.4.1 Specification.

The next empirical question is the relation between firm size and R&D intensity. Bound et al. (1984) had used a much older version of the Compustat data including only the firms that existed in 1976. With this cross-section, they found the size elasticity of R&D to be 0.97 suggesting that the R&D intensity is independent of the firm size.

To analyse the same relation, I will consider the following benchmark specification,

$$\ln \left(R\&D_{it}/Sales_{it}\right) = \beta_0 + \beta_1 \ln \left(Sales_{it}\right) + \delta_j + \delta_t + \varepsilon_{ijt} \tag{3}$$

[6] I also checked the robustness of the results with Compustat segment level data and I did not find any significant difference between the results of the two datasets.

In this reduced form equation $R\&D_{it}$ stands for the total R&D spending of a company in a given year.

The results of this benchmark regression are in column 1 of Table 3 . In contrast to most previous studies, a 10 percent increase in firm size is associated with 2.65 percent decrease in R&D intensity. In other words, the elasticity of R&D spending with respect to the firm size is 0.735 which is significantly less than unity. To see the practical importance of this effect, consider an average sized firm ($\ln Sale = 0.073$) with the average R&D intensity ($\ln(R\&D/Sale)$ =-3.205). Consider also a 10 percent annual growth in real sales for 10 years. If the R&D intensity was independent of the firm size, the new real R&D spending would have increased from 4.4 to 11.4 percent at the end of the period. However, with the current estimate, the real R&D spending goes up to 8.9 percent which is 22 percent lower than the case where R&D intensity is independent of the firm size.

Unfortunately, about 30% of the observations are excluded due to missing entries, which is a problem that is addressed next.

2.4.2 Missing Observations

Due to the SEC reporting requirements[7] and the definitions in Compustat data collection rules, R&D entries of some firms are recorded as either 0 or as blank even though firms have conducted positive amount R&D. Following Bound et al (1984), I will treat both missing and null observations as signifying *"not reporting positive R&D"* (see Bound et al (1984) for a detailed discussion of this assumption). Consequently, a two-step Heckman selection procedure is applied.

Table 3, columns 2a,b and 3 report the results of the selection model. The significance of the inverse Mills ratio indicates as before the existence of a selection bias. Column 2a shows that the probability of reporting positive R&D is increasing in firm size. Column 2b reports the marginal effects at the sample means. It shows that a marginal deviation from the sample average of firm size increases the probability of reporting positive R&D by 0.04 (standard error=0.001). The corrected estimates in Column 3 indicate that the R&D intensity decreases by 2.3 percent if the firm size increases by 10 percent, which corresponds to an elasticity of 77%, again significantly less than unity. To see its economic significance, the same exercise as above gives 19 percent lower real R&D spending in the final period relative to the case where the R&D intensity is independent of the firm size. Therefore the coefficient is still economically large.

[7] Securities and Exchange Commision

2.4.3 Measurement Error

Similar to section 2.3.3, measurement errors could be a concern. I follow the same steps as before and instrument the sales at time t by its lagged value. The estimates, presented in Column 4 of Table 3, show the coefficient on $\ln Sales$ is precisely estimated as -0.225 (standard error=0.003). In other words, the results were not strongly biased by the measurement error, since the R&D elasticity increases from 74 percent to only 78 percent.

2.4.4 Sample Selection

This problem has been discussed in the previous section in detail. Briefly, if the Compustat sample selects only the highest R&D intensive firms among the small firms, and if this bias is not so severe for large firms, this factor might derive the observed results in the benchmark OLS. Therefore I will address this issue again in two ways.

First, the permanent unobserved heterogeneity is captured through fixed effects. The result of this method is presented in column 5. Interestingly, the coefficient estimate of this specification is lower than the benchmark OLS estimate, namely -0.384 with a standard error of 0.007, and the size elasticity of R&D is now only 62% , marking even more the contrast to the previous literature. The possible explanation for this result is that excluding firm characteristics from these regressions is likely to give all the credits for higher R&D intensity to firm size. For example firm specific know-how or products could affect the size of the firm through making it capture the market and grow faster. The very same factors could make the firm invest in R&D to develop them further.

Secondly, the temporary unobserved heterogeneity caused by the firm's life-cycle can be captured by the number of years spent in the Compustat sample. The results from this estimation, in Column 6 confirms that firms are more R&D intensive when recently added to the sample, but does not have a dramatic effect on the benchmark estimates. To address the same problem in a different way, one can restrict the sample to only those firms which have been surveyed by Compustat for more than 10 years. The resulting estimate in Column 7 is now -0.092, which is higher than the benchmark. As it has been mentioned in the previous section, the possible explanation for this result is the exclusion of small firms from the sample. However, even among mature firms, the negative relation between R&D intensity and firm size remains, though less stark. Consider a firm that has the average size and R&D intensity in this new sample ($\ln Sale$ =1.027, $\ln(R\&D/Sale$ =-3.705) and grows with 10 percent for 10 years. The end-of-period real R&D spending was going to be 0.028 if R&D intensity was independent of the firm size, but is now only 0.021, that is, 8% lower.

2.4.5 Additional Caveats

The robustness of the results to alternative specifications can also be checked. The methods are the same as the ones employed in section 2.3.5.

Changing Sector Heterogeneities. The results using a year×sector interaction effect are presented in column 8 and are very close to the benchmark OLS results.

Short-Run vs Long-Run. The long-run estimate of -0.298 (standard error=0.014) in column 9 is 10% lower than the short-run estimate. This coefficient denotes that the elasticity of R&D with respect to the firm sale is 0.70. As a result, the negative relation between R&D intensity and firm size is sharper in the long-run.

Overall, the following fact can be summarized from the data.[8]

Fact 2 (R&D Intensity) *Firm size is negatively related to R&D intensity.*

2.5 Firm Size vs Innovation Size

2.5.1 Specification.

The final and most innovative reduced form estimated is the one between firm size and innovation quality, as proxied by the number of forward citations. The benchmark reduced form regression considered is:

$$Citations_{jit} = \beta_0 + \beta_1 \ln (Sales_{jt}) + \delta_k + \delta_t + \varepsilon_{ji} \tag{4}$$

where $Citations_{jit}$ denotes the number of citations that patent i of firm j receives within 7-years after its grant date t, $Sales_{jt}$ is the sales of firm j in year t, and δ_k is a fixed effect for 4-digit International Patent Classification (IPC) k. In chosing the window size, one faces a trade-off because a larger window captures more citations per patent, but also reduces the number of observations available. Since the number of cumulative citations increases mostly until 5 to 6 years before evening out, after the grant date, using a 7-year period seems optimal.[9]

The first column in Table 4A reports the benchmark result, using the full sample of patents and firms. The coefficient estimate is precisely estimated as 0.154 (standard error = 0.008). This estimate is economically sizable too. A one-standard deviation increase in the log of

[8] I also checked the robustness of the results with Compustat segment level data and I did not find any significant difference between the results of the two datasets.

[9] I checked for the robustness of the results with 8 and 10-year windows and verified that the results are not significantly sensitive to those variations.

firm sale (=1.844) is associated with a decrease of 0.284 citations, which is 6 percent of the sample mean. However, there might be a biased sample problem. The patent data sample was matched only to firms which existed in 1989. Hence, the firms that entered after 1989 are not represented. This causes the sample to get smaller, "older" and the average firm size to get "larger" over time; hence, the sample could get biased as time elapses.

To account for this, the second column focuses only on the patents obtained between 1988-1992. In this case, the coefficient doubles to -0.324. This result should not be surprising because the previous sample included presumably large and successful firms from later years in addition to the current sample, which would bias the result in favor of large firms

A possible flaw with the OLS regression is that it does not take into account the discrete nature of the data, as well as the overdispersion in citations (that is, the standard deviation of citations is greater than the mean). To correct for this, a negative binomial regression is performed, the results of which are reported in Column 3. In column 4 the analysis is extended to a zero-inflated negative binomial to correct for the fact that a lot of observations cluster around zero, that is many patents receive zero citations. The conclusion of all these alternative regressions is that the patents of small firms receive more citations and hence that smaller firms generate better quality patents.

However, there remain some important issues to be addressed.

2.5.2 Patenting Decisions

An important concern is the differential patenting behavior of small and large firms. One could argue that due to strategic reasons large firms might patent more heavily. Indeed, one can consider that there is a "crown" patent, which is the main patent for an important innovation, but that large firms also patent other peripheral innovations around it, so that those can be used against opponents in cases of litigation, to make a strong case against a competitor coming even close to the market niche of the firm. On the other hand, if the costs of patenting are important to small firms, they might register their innovations only if the innovation is significantly high quality and prefer to save on the expense of patenting minor quality patents. This would cause the quality of patents of small firms to be higher. This concern can be addressed by performing the regression not on the whole universe of patents, but rather on only the best patent (the one with the highest number of citations) of each company. However, this strategy tends to overestimate the quality of patents of larger firms that apply for a big number of patents.[10] For that purpose, I will use the total number of patents of a firm as a control variable.

Column 5 shows the results of this approach. The results confirm the existence of a possible bias since the coefficient of the negative binomial regression has increased from -0.045 to -0.029. However, it is reassuring to see that the bias due to patenting behavior of firms is not the sole driver of the negative relation between firm size and innovation quality.

2.5.3 Alternative Quality Measures

Next, I consider alternative indicators of patent quality. Column 6 considers an originality index, the construction of which was explained in a previous section ranging from 0 to 1, and increasing in the width of technologies cited by the patent. For any given number of citations, the wider is the range of cited patent classes, the higher the value of this index will be.[11] The results on column 6 indicates that as the firm size doubles, its innovations combine a smaller number of technologies and the originality index decreases by 0.001.

In Column 7 the results from regressing the number of claims of a patent on the firm size are presented. The number of claims could be seen as an alternative indicator of the value of an innovation since each claim has an additional cost and the patent holder would demand so many claims that the marginal cost is equated to the marginal expected return to that innovation. This result of this regression shows that the number of claims goes down by 0.6 as the firm size doubles, which implies again that the value of an innovation decreases in firm size.

In conclusion, I proxied the innovation quality by patent citations, by patent originality and by the number of patent claims and all these cases generated the following result.

Fact 3 (Innovation Size) *Firm size is negatively related to innovation size.*

3 Model

3.1 Demographics, Preferences and Technology

Consider the following discrete time economy. The representative household maximizes its sum of expected discounted utility over an infinite horizon, with intertemporal preferences of the

[10] Assume firm i receives only 1 patent in a given patent class in year t and firm j obtains 10 patents in the same patent class and year. If we focus on the best patents, firm j will be more advantegous since its best patent will be the best of 10 patents whereas firm i has only one patent which will count as the best patent.

[11] The calculation of this index was described in the data description section.

following form,

$$U_t = \mathbb{E}_t \sum_{\tau=t}^{\infty} \beta^{\tau} \log C_{\tau}. \tag{5}$$

where C_t denotes consumption at time t. In this specification, $\beta \in (0,1)$ is the discount factor and $\mathbb{E}_t$ the expectation operator conditional on the information at time t. Logarithmic form measures the per-period utility derived from the consumption of the final good Y_t which is produced using a continuum of differentiated goods indexed by $i \in I$ and a numeraire good. Specifically,

$$Y_t = \left[\alpha y_{0,t} \int_{i\in I} y_{i,t} di - \frac{\gamma}{2} \int_{i\in I} y_{i,t}^2 di - \frac{\delta}{2} \left(\int_{i\in I} y_{i,t} di \right)^2 \right]^{1/2} \tag{6}$$

where $y_{0,t}$ and $y_{i,t}$ denote the quantities of the numeraire good and variety i at time t, respectively and $\alpha, \gamma, \delta > 0$ capture the substitution patterns among the varieties and the numeraire good. Variations of the specification in (6) is commonly used in the trade literature (Ottaviano et al (2002), Melitz and Ottaviano (2008), Corcos et al (2007)). The specification in (6) features *love for variety* and this is measured by γ. Having $\gamma = 0$ implies that the varieties are perfect substitutes. Without loss of generality, the price of the final good is set equal to 1. The household is allocated with a continuum of 1 unit of labor which will be supplied inelastically for production and R&D. The representative household also owns a balanced portfolio of all the firms in the economy which implies the following budget constraint

$$C_t \leq \int_{i\in I_A} \Pi_{i,t} di + w_t$$

where I_A is the set of active firms in the economy.

3.1.1 Final Good Sector

The final good producer who is a profit maximizer uses a set of differentiated goods $[y_i]_{i\in I}$ and a numeraire $y_{0,t}$ as inputs. Each of the differentiated input good is being produced by a monopolist $i \in I_A$ who charges $p_{i,t}$ for each unit it sells. While mapping the model into real life, one can think of each variety as a different brand of a product, say i and j as the Home Computing goods of Hewlett Packard (HP) and Dell. To capture the reality that HP produces and develops only HP products, in specification (6) firm i will compete with other firms from its own product line. In other words, it will produce, advance and price products that are only in its own product line i. The representative household, which can be thought as the whole economy being a big single household, provides demand for every type of brands.

In equilibrium, the final good producer will demand more of the lower priced products which will push the firms into competition over the cost of production.

The numeraire good, on the other hand, is purchased from an outside market at a price $p_{0,t} > 0$. The final good producer does not have a storage technology, therefore it does a period-by-period optimization. The problem of the final good producer can be stated as

$$\pi_{Y,t} = \max_{y_0,[y_i]_{i \in I}} \left\{ Y_t \left(y_{0,t}, [y_i]_{i \in I} \right) - \int_{i \in I} p_{i,t} y_{i,t} di - p_{0,t} y_{0,t} \right\} \tag{7}$$

This maximization will generate final good producer's demands for the intermediate goods, $y_{i,t}^d \left(p_{i,t} \right)$.

3.1.2 Intermediate Goods Sector

This sector is composed of monopolistically competitive firms which decide for production, R&D, entry and exit. Each of these decisions is described below:

Production. Monopolist i in the intermediate goods sector operates with the following linear technology,

$$y_{i,t} = q_{i,t} l_{it}$$

where $l_{i,t}$ is the labor hired by firm i for production at time t and $q_{i,t} \in Q_t \equiv [q_{\min,t}, q_{\max,t}]$ is a firm specific productivity level with a distribution function $\Psi_t \left(Q_t \right)$. Productivity $q_{i,t}$ can be stochastically improved through R&D. The bounds of Q_t imply that in every period there is a threshold below which the technology is so outdated that it does not have any productive value relative to the current level of technology and there is also an upper bound for the feasible technologies. For mathematical convenience, the lower bound for the technology is assumed to be $q_{\min,t} \geq 6 \alpha w_t p_{0t} / (\alpha^2 - 2 \delta p_{0t}^2)$. This assumption ensures three points: First, the production of the monopolist never assumes negative values. Second, since in steady state, p_{0t} will be constant and w_t will grow at a constant rate, the lower bound for the minimum quality (the threshold for the outdated technologies) in the economy would grow at the same constant rate. Third, the profit function is concave in labor productivity $q_{i,t}$, which is crucial for the results of the model.

On the production side, monopolist i decides how much to produce and what price to charge for its product, given its own technology and the demand from the final goods sector for its good. Its production decision is simply

$$\max_{p_{i,t},y_{i,t}} \{(p_{i,t} - c_{i,t})\, y_{i,t}\} \quad \text{subject to} \quad y_{i,t} = y^d_{i,t}(p_{i,t}) \tag{8}$$

R&D. The marginal cost of the monopolist is equal to the ratio of the current wage rate to the firm-specific productivity, $c_{i,t} = w_t/q_{i.t}$. This has two implications. First, for any given wage rate, higher productivity means lower marginal cost and higher profit, which generates an incentive to innovate. Second, for any given constant productivity $q_{i,t} = q_{i,t+1}$ between periods t and $t+1$, the change in the marginal cost is determined by the growth rate of the wage g_t which the monopolist takes in his problem as given.

Firms operate in a Schumpeterian environment, in which they can improve their state of technology $q_{i,t}$ through *additive* step-by-step innovations. The outcome of R&D is uncertain, so that a firm cannot directly choose to make an innovation, but only a probability of success, that is, achieving an innovation. The novelty of the current model is that firms can endogenously choose not only this probability of success, but also the size of the innovation. Hence, each R&D decision is splitted into two components:

- the size of the innovation that the firm aims for, $\lambda_{i,t} \in \left[0, \bar{\lambda}_t\right]$, and
- the probability of success, $\phi_{i,t} \in [0,1]$.

Let $q_{i,t}$ be the technology of firm i at time t. It is assumed that any successful innovation of size $\lambda_{i,t}$ improves the technology by this amount with probability $\phi_{i,t}$. As the overall technology in the economy improves, the maximum step size $\bar{\lambda}_t$ will grow as well. For simplicity, this upper limit will be set as $\bar{\lambda}_t = w_t$.[12] Let $\left(\lambda_{i,t}, \phi_{i,t}\right)$ be firm i's R&D decision in period t. Consequently, the state of technology for firm i in the subsequent period will be

$$q_{i,t+1} = \begin{cases} q_{i,t} + \lambda_{i,t} \text{ with probability } \phi_{i,t} \\ \\ q_{i,t} \text{ with probability } 1 - \phi_{i,t} \end{cases} \tag{9}$$

The parameter $\lambda_{i,t}$ is the improvement on top of the current technology. This specification assumes that there are constant returns to scale to quality improvements, in other words, the improvements are independent of the current quality of the firm. The alternative specification,

[12] Since the payoff relevant state variable is q/w the payoff relevant improvement will become λ/w. Setting the upper limit as $\bar{\lambda} = w$ bounds the payoff relevant innovation size between 0 and 1, i.e., $\lambda/w \in [0,1]$.

which has been used by several studies in the literature (Aghion, Harris and Vickers; Aghion, Howitt, Harris and Vickers, 2000; Acemoglu and Akcigit, 2008) is that the improvements are proportional to the current quality, yet this would introduce increasing returns to scale to quality improvements.[13]

The theoretical results of the current model will rely on two features: First, the concavity of the value function in quality and second, the absence of any strong increasing returns in quality. These two features would imply that firms with lower quality, will have a greater incentive to increase their productivity. Therefore, they are going to choose higher quality innovations and grow faster. The assumptions on the aggregator in (6) and the structure of the quality improvement in (9) guarantee those two aforementioned features.

The specification for the R&D cost function is as follows: Let

$$h_{i,t} = h\left(\frac{\lambda_{i,t}}{w_t}, \phi_{i,t}\right) \tag{10}$$

denote the amount of labor required to undertake an R&D project of size λ_i with a success probability of ϕ_i at time t. In a growing economy, firms benefit from the improvement of the overall technology through (i) spillovers ((Jaffe, 1989; Jaffe, Trajtenberg, and Henderson, 1993; Anselin, Varga, and Acs, 1997; Anselin, Varga, and Acs, 2000)) (ii) improvement in labor substituting capital in R&D (computers, for instance) (iii) increase in the human capital. The reduced form (10) captures such positive externalities. In steady-state, w_t is a fraction of Y_t, therefore, (10) implies that a given size of innovation requires less labor as the overall technology in the economy advances. The function h is assumed to be strictly convex, with $h_x(\cdot,\cdot), h_{xx}(\cdot,\cdot), \ h_{xz}(\cdot,\cdot) \geq 0$; $\frac{h_x(\cdot,\cdot)}{z}$ decreases in z, for $x, z \in \{\lambda, \phi\}$ and $x \neq z$, and the Inada conditions are assumed to hold $h(0,\cdot), h(\cdot,0) = 0$, $h(1,\cdot), h(\cdot,1) = \infty$. As a result, the cost of an R&D project $(\lambda_{i,t}, \phi_{i,t})$ is simply the R&D labor expense, $w_t h_{i,t}$. Having labor as the only input for both R&D and production implies the following resource constraint for the economy,

$$C_t \leq Y_t - p_{0,t} y_{0,t}.$$

Free Exit. In every period, firms are assumed to have an outside option of $\Lambda_t = \nu_\Lambda Y_t$ where $\nu_\Lambda \in (0,1)$. This means when the market value of a firm falls below Λ_t, it would be optimal to exit and utilize the outside option. The decision to exit will be denoted by $\kappa_{i,t} = 1$ if the firm decides to exit and $\kappa_{i,t} = 0$, otherwise. In terms of timing of the model, firms decide whether

[13] One way to kill this increasing returns to scale would be to make the R&D cost function proportional to the current quality.

to stay for the next period or exit the market, after the stochastic R&D outcome is realized and before period t ends. In summary, the decision variables of a monopolist firm are i) output $y_{i,t}$, ii) price $p_{i,t}$, iii) innovation size $\lambda_{i,t}$, iv) innovation intensity $\phi_{i,t}$, and v) exit $\kappa_{i,t}$.

Free Entry. The market has an outside pool of potential entrants, which is large enough to have entry as long as it is profitable. These outside firms are ready to pay an entrance fee $\chi_t = \nu_\chi Y_t$ where $\nu_\chi \in (0,1)$ and determine their entry level productivity $q_{i,t}$ with a draw from the last period's productivity distribution $\Psi_{t-1}(q)$. In the beginning of period t, these potential entrants pay the fee and draw a productivity. As a result of this draw $q_{i,t} \sim \Psi_{t-1}(q)$, firms with a high productivity draw will enter the market and the firms with a low productivity draw will find it more profitable to stay out of the market and utilize the outside option. The outside firms will enter the market as long as the expected value of entry is greater than the entry fee χ_t. This process will determine the equilibrium measure of the firms in the market, N_t.

Labor Market. Labor is being employed by monopolist i both in production $(l_{i,t})$ and R&D $(h_{i,t})$. Given the measure of the labor supply is 1, the wage rate has to adjust to satisfy the following labor market clearing condition,

$$1 \geq \int_{i\in I} \left[l_{i,t}(w_t) + h_{i,t}(w_t)\right] di$$

If the demand for labor is lower than the supply, the wage rate w_t would go down until the inequality holds with equality.

Finally, to review the model, the timeline of the model in period t can be summarized as follows:

beginning of period t	$\rightarrow$	draw of a productivity, $q_{i,t} \sim \Psi_{t-1}(q)$ and decision to enter	$\rightarrow$	i) production $y_{i,t}$ ii) R&D, $(\lambda_{i,t}, \phi_{i,t})$ iii) labor market clearance	$\rightarrow$	R&D outcome realized, $q_{i,t+1}$	$\rightarrow$	decision to stay or exit, κ_t

3.2 Steady-State Equilibrium

Before starting to characterize the equilibrium, I provide the defition of an allocation in this economy.

Definition 1 (Allocation) *An allocation in this economy consists of i) consumption levels $[C_t]_{t=0}^{\infty}$, ii) prices, quantities, innovation sizes, innovation intensities and exit decisions of mo-*

nopolists $\left\{ \left[p_{i,t}, y_{i,t}, \lambda_{i,t}, \phi_{i,t}, \kappa_{i,t} \right]_{t=0}^{\infty} \right\}_{i \in I}$, *iii) quantities of the numeraire good* $[y_{0,t}]_{t=0}^{\infty}$, *iv) the measure of available product types,* $[N_t]_{t=0}^{\infty}$, *v) the distribution of productivities,* $\left\{ [\Psi_{q,t}]_{q \in Q_t} \right\}_{t=0}^{\infty}$ *and vi) wage rates,* $[w_t]_{t=0}^{\infty}$.

Throughout the model, I will focus on the Markov Perfect Equilibrium which makes the payoffs a function of the payoff relevant state variable. For each individual firm, the state variables are its own productivity and the productivity distribution in the economy. However, in the firm's optimization problem, the productivity distribution reflects itself in the wage rate, therefore the only pay-off relevant state variable can be summarized as the ratio of the firm's productivity to the wage rate, $\hat{q}_i \equiv q_i/w_t$. The formal definition of the Markov Perfect Equilibrium goes as follows.

Definition 2 (Equilibrium) *A Markov Perfect Equilibrium is an allocation* $\langle [C_t^*]_{t=0}^{\infty}, \left\{ \left[p_{i,t}^*, y_{i,t}^*, \lambda_{i,t}^*, \phi_{i,t}^*, \kappa_{i,t}^* \right]_{t=0}^{\infty} \right\}_{i \in I}, [y_{0,t}^*]_{t=0}^{\infty}, [N_t^*]_{t=0}^{\infty}, [w_t^*]_{t=0}^{\infty} \rangle$ *such that i)* $\left[p_{i,t}^*, y_{i,t}^* \right]_{t=0}^{\infty}$ *solves monopolist's profit maximization, ii)* $\left[\lambda_{i,t}^*, \phi_{i,t}^* \right]_{t=0}^{\infty}$ *solves optimal R&D investment problem, iii)* $\left[\kappa_{i,t}^* \right]_{t=0}^{\infty}$ *solves monopolist's exit problem, iv)* $\left[y_{0,t}^* \right]_{t=0}^{\infty}$ *is consistent with the final good producers maximization problem, v)* $[C_t^*]_{t=0}^{\infty}$ *is consistent with the household's optimization, vi)* $[N_t^*]_{t=0}^{\infty}$ *makes the free-entry condition hold as an equality, vii) the evolution of the productivity distribution* $\left\{ [\Psi_{q,t}]_{q \in Q_t} \right\}_{t=0}^{\infty}$ *is consistent with the R&D, entry and exit decisions of the firms and viii)* $[w_t^*]_{t=0}^{\infty}$ *is consistent with the labor market clearing condition.*

Having provided the definition of the equilibrium for this economy, I start to solve for it with the production decision of the monopolist.

Production The final good producer's demand for the numeraire good in equation (7) is

$$y_{0,t} = \left(\frac{\alpha}{4p_{0,t}^2} + \frac{\delta}{2\alpha} \right) \bar{y}_t N_t + \frac{\gamma \sigma_t}{2\alpha} \tag{11}$$

where $\bar{y}_t \equiv (1/N_t) \int_{i \in I^*} y_{i,t} di$, $\sigma_t \equiv \int_{i \in I^*} y_{i,t}^2 di / \int_{i \in I^*} y_{i,t} di$ and N_t is the total number of firms that are actively producing in the market. The demand for the numeraire good is increasing in both the average consumption of the differentiated goods and the dispersion of differentiated goods σ. Similarly, the inverse demand for each variety $i \in I$ is

$$p_{i,t} = \frac{\alpha y_{0,t} - \gamma y_{i,t} - \delta \bar{y}_t N_t}{2Y_t}. \tag{12}$$

The monopolist i with a marginal cost $c_{i,t} = w_t/q_{i,t}$ takes this inverse demand for its variety as given and maximizes its profit as in (8). Since the pay-off relevant state variable is the ratio of the technology to the wage rate, I will adopt the following notation $\hat{q}_{i,t} \equiv q_{i,t}/w_t = 1/c_{i,t}$. The maximization of the monopolist delivers the following output and price decisions

$$y_i = \frac{Y}{\gamma}\left(\frac{2\gamma p_{0,t}}{\alpha N} + \overline{\left(\frac{1}{\hat{q}}\right)_t} - \frac{1}{\hat{q}_{i,t}}\right) \tag{13}$$

$$p_i = \frac{1}{2}\left(\frac{2\gamma p_{0,t}}{\alpha N} + \overline{\left(\frac{1}{\hat{q}}\right)_t} + \frac{1}{\hat{q}_{i,t}}\right) \tag{14}$$

The optimal price set by the monopolist increases in its marginal cost $c_{i,t} = 1/\hat{q}_{i,t}$. Note that with the aggregator in (6) the mark-ups are decreasing in the number of competitors and in marginal cost. This implies that an increase in marginal cost is always shared by both the monopolist and consumer. The empirical findings of Hopenhayn and Campbell (2002) and Syverson (2004, 2007) empirically support these features.

Next, using the optimal pricing and output decisions of the monopolists, its profit becomes the following expression

$$\pi_t\left(\hat{q}_{i,t}\right) = \frac{Y_t}{2\gamma}\left(\frac{2\gamma p_{0,t}}{\alpha N} + \overline{\left(\frac{1}{\hat{q}}\right)_t} - \frac{1}{\hat{q}_{i,t}}\right)^2 \tag{15}$$

where $\overline{(1/\hat{q})}$ is the mean of the average marginal cost in the economy. Profits decrease in marginal costs, providing firms with an incentive to reduce their costs through innovations. Another key point regarding (15) is that it is concave in $\hat{q}_{i,t}$ on the set $\hat{Q}_t \equiv Q_t/w_t$.

Next, using the zero-profit condition of the final good producer and equation (14), the price of the numeraire good is pinned down as

$$p_{0,t} = \frac{\frac{\alpha^2}{2}\left[1 + \frac{\gamma}{2Y^2}\int_{i\in I} y_{i,t}^2 di\right] - 2\alpha\overline{(1/\hat{q})}_t}{\left(\frac{4\gamma}{N} + \delta\right)}.$$

Value Functions, R&D and Exit Starting from the end of period t, we can formulate the maximization of firm i backwards. Let $q_{i,t}$ be the technology of firm i during the production in period t and let $q_{i,t+1}$ be the productivity after the R&D outcome is realized at the end of period t. Denoting the beginning-of-period and end-of-period values of firm i by $V_t(\cdot)$ and $W_t(\cdot)$, respectively, and recalling that the pay-off relevant state variable is $\hat{q}_t \equiv q_t/w_t$, the exit problem can be stated as

$$W_t\left(\hat{q}_{i,t+1}\right) = \max_{\kappa_i\in[0,1]}\left\{\kappa_{i,t}\Lambda_t + (1-\kappa_{i,t})\, V_{t+1}\left(\frac{\hat{q}_{i,t+1}}{1+g_t}\right)\right\}. \tag{16}$$

where Λ_t is the outside option and g_t is the growth rate of the wage between time t and $t+1$ which firm i takes it as given. Now going backwards, conditional on the fact that the firm starts the period with a productivity level of $q_{i,t}$, the program for the R&D decision is summarized by the following Bellman equation,

$$V_t\left(\hat{q}_{i,t}\right) = \max_{\substack{\lambda_i \in \left[0,\bar{\lambda}_t\right] \\ \phi_i \in [0,1]}} \left\{ \begin{array}{c} \pi_t\left(\hat{q}_{i,t}\right) - w_t h\left(\hat{\lambda}_{i,t}, \phi_{i,t}\right) + \\ \frac{1}{1+r}\left[\phi_{i,t} W_t\left(\hat{q}_{i,t} + \hat{\lambda}_{i,t}\right) + \left(1-\phi_{i,t}\right) W_t\left(\hat{q}_{i,t}\right)\right] \end{array} \right\} \tag{17}$$

where $\hat{q} \equiv q/w$ and $\hat{\lambda} \equiv \lambda_t/w_t$. The first term in this equation is the gross profit, the second expression is the R&D expenditure due to the hired R&D workers for an innovation project with a size $\lambda_{i,t}$ and a success probability $\phi_{i,t}$. The expression in the bracket is the expected end-of-period value of firm i taking into account that the R&D investment will be successful with the endogenous probability $\phi_{i,t}$. r is the interest rate.

Entry In the beginning of period t, potential entrants pay an entrance fee $\chi_t = \nu_\chi Y_t$ and determine their starting technology with a draw from the previous period's distribution, $q \sim \Psi_{t-1}\left(q\right) = w_{t-1}\Psi_{t-1}\left(\hat{q}\right)$. This means, since the fee is sunk at the moment of the draw, firms with bad draws $q_{i,t} < q_{exit,t-1}$ will not enter the market even though they would already pay the fee. Firms will attempt to enter, as long as the expected value of entry is higher than the entry fee. This *free-entry condition* reads as,

$$\int V_t\left(q/w_t\right) d\Psi_{t-1}\left(q\right) \geq \chi_t. \tag{18}$$

When the expected value of entry is greater than the entry fee, the measure of firms in the market, N will implicitly increase leading to a uniform decrease in the value function. As a result, the expectation on the left-hand side of (18) will decrease until the inequality holds with equality.

Labor Market Labor is being employed both in the production and in the R&D sector. The demand for labor in the former; derived from equations (13) is:

$$l_t\left(\hat{q}\right) = \frac{Y_t}{\gamma q_t}\left(\frac{\alpha y_{0,t} - \delta N_t \bar{y}_t}{2Y} - \frac{w_t}{q_t}\right) \text{ for } \hat{q}_t \in \hat{Q}_t \tag{19}$$

Previous section showed that firm i's labor demand for R&D activity is $h\left(\hat{\lambda}_{i,t}\left(\tilde{w}_t\right), \phi_{i,t}\left(\tilde{w}_t\right)\right)$. Setting the supply of labor in the economy to 1, the labor market clearing condition is

$$1 = \frac{N_t}{\tilde{w}_t \gamma}\left[\left(\frac{\alpha y_{0,t}}{2Y_t} - \frac{\delta}{\alpha}\right)\overline{\left(\frac{1}{\hat{q}_t}\right)} - \overline{\left(\frac{1}{\hat{q}_t^2}\right)}\right] + N_t \bar{h}_t$$

where $\bar{y}$, $\overline{(1/\tilde{q})}$ and $\bar{h}$ stand for the average values of differentiated goods, normalized marginal costs and the workers employed by firms for R&D, respectively. From this equation, it is clear that the relevant variable for the labor market clearing is the labor share, $\tilde{w}_t = w_t/Y_t$. If the demand for labor is lower than the supply, the normalized wage rate $\tilde{w}$ would adjust to bring the market into equilibrium.

The focus, henceforth, will be on the steady-state, in which all aggregate variables grow at the same rate $g_t > 0$.

3.3 Steady-State and Theoretical Results

I will start this section normalizing the over-time-growing variables by Y_t in order to transform the problem into a stationary problem. Throughout this section, the normalized value of some generic variable x will be denoted by $\tilde{x} \equiv x/Y$.

Definition 3 (Steady-State Equilibrium) *A Steady-State Equilibrium is a tuple* $\langle \left[\lambda^*_{\hat{q}}\right]_{\hat{q}\in\hat{Q}^*}, \left[\phi^*_{\hat{q}}\right]_{\hat{q}\in\hat{Q}^*}, \left[\kappa^*_{\hat{q}}\right]_{\hat{q}\in\hat{Q}^*}, \left[\Psi^*_{\hat{q}}\right]_{\hat{q}\in\hat{Q}^*}, N^*, \tilde{w}^*, g^*\rangle$ *such that i)* $\left[\lambda^*_{\hat{q}}\right]_{\hat{q}\in\hat{Q}^*}, \left[\phi^*_{\hat{q}}\right]_{\hat{q}\in\hat{Q}^*}, \left[\kappa^*_{\hat{q}}\right]_{\hat{q}\in\hat{Q}^*}$ *solve the normalized versions of the value functions in* (16), *ii)* $\left[\Psi^*_{\hat{q}}\right]_{\hat{q}\in\hat{Q}^*}$ *forms an invariant distribution over the state space* $\hat{Q}$, *iii)* $\tilde{w}^*$ *clears the labor market, iv)* N^* *is constant and consistent with steady-state free-entry condition, v) the aggregate variables* Y_t, w_t, *and the aggregate technology index* $\Theta_t \equiv \int_{i\in I^*} \frac{1}{q_{i,t}} di$ *grow at the steady-state rate* g^* *which is consistent with the steady-state R&D choices* $\left[\lambda^*_{\hat{q}}\right]_{\hat{q}\in\hat{Q}^*}, \left[\phi^*_{\hat{q}}\right]_{\hat{q}\in\hat{Q}^*}$.

Next I characterize the steady-state equilibrium. The normalized steady-state versions of the value functions in (17) take the following form

$$\tilde{V}(\hat{q}_i) = \max_{\substack{\lambda_i\in[0,1]\\ \phi_i\in[0,1]}} \left\{ \begin{array}{c} \tilde{\pi}(\hat{q}_i) - \tilde{w}h\left(\hat{\lambda}_i, \phi_i\right) + \\ \\ \frac{1}{1+r}\left[\phi_i \tilde{W}\left(\hat{q}_i + \hat{\lambda}_i\right) + (1-\phi_i)\tilde{W}(\hat{q}_i)\right] \end{array} \right\} \tag{20}$$

where $\tilde{V} \equiv V/Y$, $\tilde{W} \equiv W/Y$, $\tilde{\pi} \equiv \pi/Y$, $\tilde{w} \equiv w/Y$ are the normalized aggregate variables and $\hat{q} \equiv q/w$ and $\hat{\lambda} \equiv \lambda/w$ are the pay-off relevant state and R&D variables. Similary the end-of period value function becomes

$$\tilde{W}(\hat{q}_i) = \max_{\kappa_i\in[0,1]} \left\{ \kappa_i \tilde{\Lambda} + (1-\kappa_i)\tilde{V}\left(\frac{\hat{q}_i}{1+g^*}\right)\right\} \tag{21}$$

where $\tilde{\Lambda} \equiv \Lambda/Y$. Since the transformed problem is autonomous, the time subscripts are dropped. A point worth noting is that the growth of the wage rate g reflects the pressure of the rest of the firms on firm i as follows. The relative technology (inverse of the marginal cost) of firm i deteriorates as the overall economy innovates because the increase in aggregate technology reflects itself through an increase in wages. This implies that for any given $g > 0$, if firm i never innovates, it will have to exit the market eventually. Therefore, in addition to the usual cost reduction incentive, there is an additional incentive to innovate, namely to survive in the market.

The optimal steady-state exit decision in (21) is

$$\begin{aligned} \kappa_i^* &= \begin{cases} 1 \text{ if } \tilde{\Lambda} > \tilde{V}\left(\frac{\hat{q}_{i,t+1}}{1+g^*}\right) \\ 0 \text{ if } \tilde{\Lambda} < \tilde{V}\left(\frac{\hat{q}_{i,t+1}}{1+g^*}\right) \end{cases} \\ \kappa_i^* &\in [0,1] \text{ otherwise} \end{aligned}$$

This simple condition implies that if the firm value is below the outside option $\tilde{\Lambda}$, the firm will find it optimal to exit.

The next proposition characterizes the value function in (20). Its proof will be presented in the Appendix of the paper, together with all other proofs.

Proposition 1 *Consider the dynamic optimization of an individual firm. For any given steady-state values* $\left(\left[\Psi_{\hat{q}}^*\right]_{\hat{q}\in\hat{Q}^*}, N^*, \tilde{w}_t, g^*\right)$

i) the value function in (20) *exists, is unique, continuous, strictly increasing, differentiable and strictly concave.*

ii) Optimal policy functions $\lambda^*\left(\hat{Q}\right), \phi^*\left(\hat{Q}\right)$ *exist and they are continuous functions.*

This proposition documents that the steady-state value function is increasing in the technology of the firm. The following lemma is immediate from propostion (1).

Lemma 1 *Let* $\hat{q}_{exit}$ *be implicitly defined as* $\tilde{\Lambda} = \tilde{V}\left(\hat{q}_{exit}/\left(1+g^*\right)\right)$. *Then*

i) firm i's exit decision is a cut-off rule such that

$$\begin{aligned} \kappa_i^* &= 1 \textit{ if } \hat{q}_i < \hat{q}_{exit} \\ \kappa_i^* &= 0 \textit{ if } \hat{q}_i > \hat{q}_{exit} \\ \kappa_i^* &\in [0,1] \textit{ otherwise} \end{aligned}$$

ii) potential entrants will follow the same cut-off rule in their entry decision.

Recall that the relative technology $\hat{q}_i = q_i/w$ decreases if the firm fails to innovate since $w_t = w_0 (1+g^*)^t$ grows at the rate g^*. Together with the cut-off rule, this implies that the firms below a certain threshold $\hat{q}_i \in [\hat{q}_{\min}, \hat{q}_{dist}]$ where $\hat{q}_{dist} \equiv (1+g^*)\,\hat{q}_{exit}$ will exit the market with certainty if they fail to innovate in the current period. I will call the active firms below this threshold, $\hat{q}_i \in [\hat{q}_{\min}, \hat{q}_{dist}]$ as *distressed* firms.

Next, having generated the differentiability and concavity of the value function in proposition 1, we can use the first order conditions to pin down the optimal steady-state R&D decisions:

$$\hat{\lambda}_i \quad : \quad \frac{(1+g^*)\,\tilde{w}^* h_{\hat{\lambda}}\left(\hat{\lambda}_i^*, \phi_i^*\right)(1+r)}{\phi_i^*} = \tilde{V}'\left(\frac{\hat{q}_i + \hat{\lambda}_i^*}{1+g^*}\right) \tag{22}$$

$$\phi_i \quad : \quad \tilde{w}^* h_{\phi_i}\left(\hat{\lambda}_i^*, \phi_i^*\right)(1+r) = \begin{cases} \tilde{V}\left(\frac{\hat{q}_i+\hat{\lambda}_i^*}{1+g^*}\right) - \tilde{V}\left(\frac{\hat{q}_i^*}{1+g^*}\right) & \text{if } \hat{q}_i > \hat{q}_{dist} \\ \tilde{V}\left(\frac{\hat{q}_i+\hat{\lambda}_i^*}{1+g^*}\right) - \tilde{\Lambda} \text{ otherwise} \end{cases} \tag{23}$$

The key economic force for the innovation size is the marginal value of the new state that firm i is going to end up when it is successful. On the other hand, the incentive for the innovation intensity is the difference between the values of the successful state and the failure state, which is simply the private value of innovation. The larger is this private value, the more intensively firm i is going to try to innovate.

Let $\left[\lambda_{\hat{q}}^*\right]_{\hat{q}\in\hat{Q}^*}, \left[\phi_{\hat{q}}^*\right]_{\hat{q}\in\hat{Q}^*}$ denote the steady-state R&D decisions of all the firms in (17). Then the aggregate steady-state growth rate is characterized as follows.

$$g^* = \frac{\int_{i\in I^*} \frac{1}{\hat{q}_i} di}{\int_{i\in I^*} \frac{\hat{q}_i + \hat{\lambda}_i^*(1-\phi_i^*)}{\hat{q}_i^2 + \hat{\lambda}_i^* \hat{q}_i} di} - 1 \tag{24}$$

where $\hat{q}_i = q_i/w^*$ and $\hat{\lambda}^* = \lambda_i^*/w^*$.

The growth rate is determined not only by the heterogenous innovation intensities but also by the heterogenous innovation sizes, λ_i^*. Therefore this proposition has an important implication: A mechanism, such as R&D tax credit, which targets to improve the aggregate growth rate or boost the aggregate technological innovation should also take into account the heterogeneity of the innovation sizes of different firms.

Next, I return to the discussion on the firm entry. After the firms with $\hat{q}_i \geq \hat{q}_{exit}$ enter the market, firms invest in R&D, choose (λ_i, ϕ_i) and hire the R&D workers accordingly. At any point in time, firms are assumed to have an outside value of $\tilde{\Lambda} > 0$. $\tilde{\Lambda}$ is assumed to be binding such that firms find exiting more profitable rather than staying in the market with a technology which is below the minimum acceptable level, $\hat{q}_{\min}$.

Let us denote the measure of the firms with $\hat{q}$ by $\Psi(\hat{q})$. When the economy has a strictly positive growth rate $g > 0$, the relative productivity of firm i will depreciate in every period. If the firm's step size is bigger than the impact of the depreciation, i.e., $\hat{\lambda}_i > g^* \hat{q}_i$, then the firm's relative technology will improve. From Proposition (1) we know that the step size of innovation, $\hat{\lambda}(\hat{q})$ is a strictly decreasing function converging to 0. Therefore, we have the following lemma.

Lemma 2 *Let $\hat{\lambda}(\hat{q})$ be the optimal choice of the step size. Then*

i) there exists $\check{q} \in \hat{Q}$ such that $\hat{\lambda}(\check{q}) = g^* \check{q}$, and

ii) for $\hat{q}_i > \min\{\check{q}, \hat{q}_{\max}\}$, $\hat{q}_{i,t+1} \leq \hat{q}_{i,t}$.

This lemma implies that there exists a threshold level of technology above which all the states are transient. This will be crucial for the existence of the invariant distribution. The formal statement for the existence of the invariant distribution is provided with the proposition below.

Proposition 2 *Consider the above model and let the state of outside firms be denoted by $\hat{q}_{out}$. Assume also that the steady-state growth rate of the economy is strictly positive, $g^* > 0$. For the given R&D decisions in (22), (23) there exists a unique steady-state distribution of industries, $\Psi^*(\hat{q})$ such that $\Psi^*\left(\hat{Q}\right) = 1$ with an atom at $\hat{q}_{out} \in \hat{Q}$. Moreover, the invariant distribution is continuous in its transition probabilities.*

Now we have all the necessary tools for the existence of the equilibrium. I establish the existence of the equilibrium by using Brouwer-Schauder-Tychonoff Fixed Point Theorem. The road map for the proof is that the equilibrium will be shown to be the fixed point of a continuous operator that takes elements from an infinite dimensional compact convex set and maps it *into* the same set. Let $m = \left[\hat{q}_{exit,} N, \tilde{w}, p_0, g, \lambda\left(\hat{Q}\right), \phi\left(\hat{Q}\right), \kappa\left(\hat{Q}\right), \Psi, \tilde{V}\left(\hat{Q}\right), \tilde{W}\left(\hat{Q}\right)\right] \in \mathbf{M}$ be a generic argument of the following operator

$$\boldsymbol{\Phi}(m) : \mathbf{M} \to \mathbf{M} \tag{25}$$

such that

$$\boldsymbol{\Phi}(m) \equiv \left[\Phi_{\hat{q}_{exit}}(m), \Phi_N(m), \Phi_{\tilde{w}}(m), \Phi_{p_0}(m), \Phi_g(m), \Phi_{\tilde{V}}(m), \Phi_{\tilde{W}}(m), \Phi_{\boldsymbol{\Psi}}(m), \Phi_\lambda(m), \Phi_\phi(m), \Phi_\kappa(m)\right]^T$$

The details of the set $\mathbf{M}$ and the individual entries of the mapping $\mathbf{\Phi}(m)$ is provided in the Appendix. Then I show that all of the individual operators in $\mathbf{\Phi}(m)$ are continuous and $\mathbf{M}$ is a compact convex set.

These steps lead to the following proposition.

Proposition 3 *Consider the economy described above. A steady-state equilibrium* $m^* \equiv \left[\hat{q}^*_{exit,} N^*, \tilde{w}^*, p^*_0, g^*, \lambda^*(.), \phi^*(.), \kappa^*(.), \Psi^*, \tilde{V}^*(.), \tilde{W}^*(.)\right]$ *of this economy exists. Moreover, the steady-state growth rate is strictly positive* $g^* > 0$.

The next theorem delivers the model's prediction on the relation between innovation size and firm size. It is consistent with the third reduced form evidence from section 2.

Theorem 1 (Innovation Size) *Let* $\hat{\lambda}(\hat{q}) : \hat{Q} \to [0,1]$ *be the policy function as described in Proposition* $1 - ii$. *For* $\hat{q} > \hat{q}_{dist} \in \hat{Q}$, $\hat{\lambda}(\hat{q})$ *is a monotonically decreasing function such that* $\hat{\lambda}(\hat{q}_{\max}) = 0$.

This theorem states that among the healthy (non-distressed) firms, innovations become more incremental as firm size increases. The intution for this result comes from the shape of the value function. The shape of the value function is driven both by the shape of the marginal cost, $c_i = 1/\hat{q}_i$ and the concavity of the profit function in the relative technology. Since the marginal value of innovation is diminishing, the incentives for drastic innovation decreases as firm size increases.

Next we turn to the relation between the R&D intensity and the firm size.

Theorem 2 (R&D Intensity) *Let* $\mathfrak{R}(\hat{q}) : \hat{Q} \to \mathbb{R}_+$ *be the R&D intensity function defined as*

$$\mathfrak{R}(\hat{q}) = \frac{w^* h(\hat{q})}{y(\hat{q})}.$$

For $\hat{q} > \hat{q}_{\min} \in \hat{Q}$, $\mathfrak{R}(\hat{q})$ *is a monotonically decreasing function.*

This result is in line with the second reduced form evidence from section 2. The intuition for this result comes from the fact that the firm incentives for radical innovations diminishes as the firm size increases. Firms turn to more incremental innovations as their size increases and this in turn decreases the R&D intensity for larger firms.

Firm Growth The engine of both firm level and aggregate growth in this economy is the productivity enhancing innovation. The model delivered two main reasons for R&D; to increase profits and to survive in the market. As long as the improvement in the productivity dominates the increase in the wage rate, the marginal cost of production goes down and firms start to obtain a higher share in the market and grow. The following proposition chracterizes the expected growth rate of firm i.

Proposition 4 *Consider the model above and let $g^* > 0$ be the aggregate steady-state growth rate. Then, firm expected growth rate can be expressed as*

$$\mathfrak{G}(\hat{q}) = g^* + \frac{\frac{[1+g^*]\hat{\lambda}^*(\hat{q})\phi^*(\hat{q})}{\hat{q}+\hat{\lambda}^*(\hat{q})} - g^*}{\left(\alpha\tilde{y}_0 - \frac{2\delta}{\alpha}\right)\hat{q} - 1}.$$

This theorem documents the role that the innovation size plays in the expected firm growth. Intiutively, larger innovation size will bring higher growth, and since the innovation size decreases in firm size, the following theorem documents that the model generates the same result that we obtained in the reduced facts, namely the expected growth rate is smaller for larger firms.

Theorem 3 (Firm Growth) *Let $\mathfrak{G}(\hat{q}) : \hat{Q} \to \mathbb{R}$ be the function of expected growth rate as defined in Proposition 4. For $\hat{q} > \hat{q}^* \in \hat{Q}$, $\mathfrak{G}(\hat{q})$ is a monotonically decreasing function.*

This final theorem concludes the theoretical section. This model achieved to explain the real facts that we observe from the data. It predicted that above a certain size threshold the firms in the model behaved consistently with the data. Since the Compustat firms are relatively larger compared to the whole population in the US manufacturing sector, this threshold firm size also fitted well into the main picture. The next section estimates the structural parameters of the model.

4 Structural Estimation

The next goal of this paper is to use the theoretical model in order to analyze the effects of various R&D subsidy policies. The model generated important and empirically relevant predictions about the firms and their R&D investments, growth dynamics, the distribution of their productivities and their transitions within this distribution. The model can be simulated with different parameter values, taken to the real data and experimented with different macroeconomic policies. In this respect, a simulation-based estimation technique, like Simulated Method of Moments (SMM) is particularly useful because it enables us to focus on the relevant moments of the model and the data. In this section, first, I will provide the background information on SMM. Then I will describe the chosen moments and the computational strategy. Finally, I will conclude the section by presenting the estimates. The next section will conduct the policy experiment using these estimated parameters.

For this estimation, a parametric form of the R&D cost function needs to be specified and is chosen such as to satisfy all the assumptions previously imposed, namely:

$$h\left(\hat{\lambda}, \phi\right) = B_{\lambda}\frac{\hat{\lambda}^{\eta_{\lambda}}}{1-\hat{\lambda}} + B_{\phi}\frac{\phi^{\eta_{\phi}}}{1-\phi}$$

where $\eta_{\lambda}, \eta_{\phi} > 1$ and $B_{\lambda}, B_{\phi} > 0$. Seperability is a natural bechmark assumption in this context since we do not have strong priors about whether these two dimensions of innovation are complements or substitutes.

To specify the parameters to be estimated, the following simplifications are made: The interest rate is set at 5% per annum. The entry fee $\tilde{\chi}$ is also set a priori, such that the total measure of firms in the economy is equal to 1 because it is hard to find an informative moment condition for this parameter. The outside option $\tilde{\Lambda}$ determines the cut-off level for the exiters $\hat{q}_{\min}$. Alternatively, one can determine the cut-off level and compute the corresponding outside option. Since the mapping between the two is monotonic and continuous, this does not affect the estimates. For computational simplicity, I will follow the latter option. This choice does not matter for the estimates because any change in the estimates since the choice As a result, the vector of parameters of length $k = 8$ to be estimated, within the set $\mathbf{\Theta}$ of feasible values is:

$$\boldsymbol{\theta} = \begin{bmatrix}\alpha & \gamma & \delta & B_{\lambda} & B_{\phi} & \eta_{\lambda} & \eta_{\phi} & \hat{q}_{\min}\end{bmatrix}^{T} \in \mathbf{\Theta}$$

4.1 Simulated Method of Moments (SMM)

The rationale for using the SMM method lies in the lack of a closed form expression for the parameters in terms of the data moments. The idea of SMM is as follows. Let $\left(o_i^A\right), i = 1, ..., n$ be i.i.d real observations and $\mathbb{M}^A$ denote the vector of the $l \geq k$ selected moments from the actual data. Let $\mathbb{M}^S(\boldsymbol{\theta})$ denote the vector of the corresponding moments that are generated from the simulation of the model for a given set of parameters $\boldsymbol{\theta}$. In addition, let $\mathbf{R}^2(\boldsymbol{\theta})$ denote the weighted sum of squred deviation (WSSD) between the data and their simulated counterparts,

$$\mathbf{R}^2(\boldsymbol{\theta}) = \left[\mathbb{M}^A - \mathbb{M}^S(\boldsymbol{\theta})\right]^T \mathbb{W}\left[\mathbb{M}^A - \mathbb{M}^S(\boldsymbol{\theta})\right] \tag{26}$$

where $\mathbb{W}$ is the optimal weighting matrix. SMM calculates the estimate $\hat{\boldsymbol{\theta}}$ by minimizing the distance between the data and the model moments,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathbf{R}^2 (\boldsymbol{\theta})$$

Due to possible discontinuities in the objective function, *Simulated Annealing Algorithm* is used for this minimization (See Goffe, 1993 or Wu and Wang, 1998 for details of this algorithm).[14]

Let Ω denote the variance-covariance matrix of $(1/\sqrt{n})\left[\mathbb{M}^A - \mathbb{M}^S(\boldsymbol{\theta})\right]$. The optimal weighting matrix $\mathbb{W}$ is the inverse of the variance-covariance matrix $\mathbb{W} = \Omega^{-1}$ (Adda and Cooper, 2003, p. 88). Hence, observations with higher variance are given less weight. To calculate Ω, I follow Bloom (2008) and use a block bootstrap with replacement on the data.

Gouriéroux and Monfort (1996, p. 29) show that $\hat{\boldsymbol{\theta}}_n(\mathbb{W})$ is consistent when n tends to infinity and that

$$\sqrt{n}\left[\hat{\boldsymbol{\theta}}_n(\mathbb{W}) \rightarrow \boldsymbol{\theta}_0\right] \underset{n\rightarrow\infty}{\xrightarrow{d}} N(0, Q(\mathbb{W}))$$

where

$$Q(\mathbb{W}) = \left[\frac{\partial \mathbb{M}^T}{\partial \theta}\mathbb{W}\frac{\partial \mathbb{M}}{\partial \theta^T}\right]^{-1}. \tag{27}$$

Therefore the magnitudes of the standard errors of estimates are determined both by the variance-covariance of the moments and the sensitivity of the moments to the parameters, with a greater sensitivity leading to smaller standard errors.

4.2 Data and Moments

In this estimation, I use the full Compustat manufacturig sample between 1980-2005. The identifying moment conditions are generated as follows: First, the medians of the variables of interest are calculated. For each variable, the deviations from a linear time trend and from sector averages are taken. These deviations are added back to the corresponding variable medians to obtain variables centered around their medians, but purged from time and sector effects. After dropping outliers (the top and bottom 2.5 percentiles for each of the three variables firm growth, R&D intensity, firm sales), firms are ranked in terms of their size and divided into two bins of "small firms" and "large firms". The first six moment conditions match the median firm growth rates, R&D intensities and labor productivities for each bin to their simulated counterparts. The next two moment conditions match the yearly transition

[14] It samples a $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ during each iteration and accepts the current $\boldsymbol{\theta}$ as the new candidate for the global minimizer with certainty if it reduces the WSSD and with some probability if the WSSD is increased with the current $\boldsymbol{\theta}$.

rates between the two bins and the last one matches the median firm value (normalized by assets). This way, 9 moment conditions are obtained for the 8 parameters.

4.3 Identification

The necessary condition for identification is that $\frac{\partial \mathbb{M}}{\partial \theta^T}$ should have full column rank which means that the objective function $\mathbf{R}^2(\boldsymbol{\theta})$ has a unique local minimum attained for the true parameter value. For the efficiency of the estimator, it is essential to use informative moment conditions, that is the moments must be sensitive to changes in parameter values as shown in (27).

Hence, for identifying the cost parameters $(B_\lambda, B_\phi, \eta_\lambda, \eta_\phi)$, the moments used are the growth rate, R&D intensity and transition rates which are directly determined by the R&D choices. To identify the parameters of the aggregator in (5) and the cut-off productivity $(\alpha, \gamma, \delta, \hat{q}_{\min})$, firm value/asset and labor productivities are most useful.

4.4 Computational Strategy

The computational solution of the model consists of a nested fixed point problem. The outer layer of the nest consists of four variables, namely the aggregate growth rate g, the price and quantity of the numeraire good (p_0, y_0) and the labor share $\tilde{w}$. To solve the model, I start with some initial guess for these parameters. Next taken these values as given, inside the nest, firms' value functions are solved. Since the problem features continuous state space, the numerical methodology relies on *cubic splines collocation method* to approximate the value function exactly at $n = 150$ collocation nodes (Judd (1998, p. 225), Miranda and Fackler (2002, p.230)). This method uses a series of cubic polynomial segments spliced together to form a twice continuously differentiable function that aproximates the value function at n nodes. The routine for solving the model is as follows.

1. Start with an initial argmax $\hat{\boldsymbol{\theta}}$ and set $\hat{\mathbf{R}}^2 = \infty$.

2. Sample a set of parameters $\boldsymbol{\theta}_{guess} \in \Theta$.

  3. Start with a guess for $[g, p_0, y_0, \tilde{w}]_{guess}$

    4. Solve the value functions

    5. Generate the R&D efforts $[\lambda(\hat{q}), \phi(\hat{q})]_{\hat{q} \in \hat{Q}}$

    6. Using $[\lambda(\hat{q}), \phi(\hat{q})]_{\hat{q} \in \hat{Q}}$ generate the invariant distribution of firms $\boldsymbol{\Psi}(\hat{q})$.

    7. Solve for the new values of $[g, p_0, y_0, \tilde{w}]_{new}$

8. If $\left\|[g, p_0, y_0, \tilde{w}]^T_{guess} - [g, p_0, y_0, \tilde{w}]^T_{new}\right\| < criteria$, stop. Else, update $[g, p_0, y_0, \tilde{w}]_{guess}$ and go back to line 4.

9. Calculate the moments of the model $\mathbb{M}^S(\boldsymbol{\theta}_{guess})$. If $\mathbf{R}^2(\boldsymbol{\theta}_{guess}) < \hat{\mathbf{R}}^2$, then $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_{guess}$ and $\hat{\mathbf{R}}^2 = \mathbf{R}^2(\boldsymbol{\theta}_{guess})$. Else set $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_{guess}$ and $\hat{\mathbf{R}}^2 = \mathbf{R}^2(\boldsymbol{\theta}_{guess})$ with some probability which decreases as the number of iteration increases.

10. Repeat this loop as many times as possible. The more it is repeated, the more likely it will reach the global minimum.

4.5 Results

The following table reports the moments of the actual data and the model regarding the median growth rate g_i, R&D intensity $\mathfrak{R}_i$ and the labor productivity q_i in bin i; the transition rates from bin i to j, denoted exit_i, and the median firm value in the sample V.

Moment Conditions

	g_1	g_2	$\mathfrak{R}_1$	$\mathfrak{R}_2$	q_1	q_2	exit_1	exit_2	V
Data	0.025	0.009	0.068	0.031	1.202	1.462	0.103	0.043	1.053
Model	0.015	0.014	0.035	0.027	0.851	1.194	0.129	0.073	1.063

Table 5

This table shows that the median growth rate and R&D intensity is higher among smaller firms. Also, the transition rate from small firms to large firms is higher than the transition rate of large to small firms . On the other hand, the labor productivity is higher among large firms. All these qualitative facts have been matched by the simulated model. The following table reports the parameter estimates and their standard errors.

Parameter Estimates

	α	γ	δ	B_λ	B_ϕ	η_λ	η_ϕ	$\hat{q}_{\min}$
Estimate	79.389	0.118	1.569	2.984	0.016	5.958	1.948	17.725
St Dev	(0.425)	(0.011)	(0.119)	(0.212)	(0.007)	(0.107)	(0.114)	(0.125)

Table 6

Figure 3 shows the plots of the value function, R&D efforts, expected growth rate and the firm size distribution at the estimated parameter values as a function of the firm size. The value function is consistent with (1) and concave. Again consistent with their theorems, innovation quality, R&D intensity and expected firm growth rate are decreasing in firm size, for firms above a certain firm size threshold ($\cong$0.1). Next section utilizes these parameters in a policy experiment.

5 Policy Analysis: R&D Subsidy

The suboptimality of private R&D investment has spawned a heated debate both among the academicians and policymakers. It is widely argued that the social return to R&D is greater than private return, which suggests that the equilibrium amount of private R&D investment is suboptimal (Griliches, 1992, 1995; Jones and Williams, 1998). To align the social and private returns of R&D, policymakers in the U.S. and in the E.U. have used R&D subsidy programs. The UK, for instance, is one of the countries which employs size-dependent R&D policies, reserving higher subsidies to SMEs,[15] (Bloom et al 2001).

The implications of such R&D policies on the economy are substantial. Therefore it is

[15] SME: Small and Medium Sized Enterprises are defined as firms with less than 250 employees.

essential to have a good framework to investigate their effects. This section will use the estimated theoretical model to analyse size-dependent R&D policies.

5.1 Theory

In a decentralized equilibrium, firms do not take into account the externalities and the monopoly distortions that they generate. As a result, the market equilibrium does not match the first-best level. The R&D subsidy gives the policymaker a partial ability to align the private and the social incentives. In the following setup, the government provides a volume-based R&D subsidy and finances $\tau_i \in [0,1]$ portion of the whole R&D spending of the firm.[16] The subsidy rate τ_i is allowed to be size dependent and since the firm size is a monotonic transformation of the state-variable $\hat{q}$, the tax scheme that I will consider is

$$\tau_i = \tau\left(\hat{q}_i\right), \quad \forall \hat{q}_i \in \hat{Q}$$

As a result, firm i will pay only $\left[1-\tau\left(\hat{q}_i\right)\right] wh\left(\hat{q}_i\right)$ instead of the full R&D expenditure.

I will assume that the government finances these subsidies through lump sump corporate taxes, $\mathcal{T} \geq 0$. With this policy, the value function of firm i becomes

$$\tilde{V}\left(\hat{q}_i\right) = \max_{\substack{\lambda_i \in [0,1] \\ \phi_i \in [0,1]}} \left\{ \begin{array}{c} \tilde{\pi}\left(\hat{q}_i\right) - \mathcal{T} - \left[1-\tau\left(\hat{q}_i\right)\right] \tilde{w} h\left(\hat{\lambda}\left(\hat{q}_i\right), \phi\left(\hat{q}_i\right)\right) + \\ \\ \frac{1}{1+r}\left[\phi\left(\hat{q}_i\right) \tilde{W}\left(\hat{q}_i + \hat{\lambda}\left(\hat{q}_i\right)\right) + \left(1-\phi\left(\hat{q}_i\right)\right) \tilde{W}\left(\hat{q}_i\right)\right] \end{array} \right\}.$$

The government follows a balanced budget. Therefore it must be the case that

$$\mathcal{T} = \int \tau\left(\hat{q}\right) wh\left(\hat{q}\right) d\mathbf{\Psi}\left(\hat{q}\right)$$

In what follows, I will focus on optimal (welfare-maximizing) subsidy schedules.

[16] Some policies considers subsidies for non-employee R&D spendings only. I will not focus on that.

Welfare The discounted sum in (5) in the steady state can be expressed as

$$\begin{aligned} \text{Welfare}_0 &= \sum_{t=0}^{\infty} \beta^t \ln \left(C_0 \left(1+g^*\right)^t \right) \\ &\cong \frac{\ln C_0}{1-\beta} + \beta^t t g^* \end{aligned} \tag{28}$$

where C_0 is the consumption at time 0. From the resource constraint of the economy we have

$$C_t = Y_t - y_{0,t} p_{0,t}. \tag{29}$$

Summing over y_is in (13) ,

$$\frac{\alpha y_{0,t} - \delta \int y_{i,t} di}{2Y_t} - \overline{\left(\frac{w}{q_{i,t}}\right)} = \frac{\gamma \int y_{i,t} di}{N_t Y_t} \tag{30}$$

Combining (29) and (30),

$$C_0 = \left[\frac{\frac{\alpha \tilde{y}_0^*}{2} - \frac{p_0^*}{\alpha}\left(\delta + \frac{2\gamma}{N^*}\right)}{\tilde{w}^* \overline{(1/q_{i,0})}} \right] \left[1 - \tilde{y}_0^* p_0^*\right].$$

In steady state, $\tilde{y}_0^*, p_0^*, \tilde{w}^*$ and N^* are constant and obtained from the normalized economy as in section 3.3. Therefore in order to compare the welfare levels of two steady-state economies, the only variable that needs to be specified exogenously is $\overline{(1/q_{i,0})}$ which is the inverse quality index at time 0. In the analysis below, $\overline{(1/q_{i,0})}$ is set to 0.1 and equation (28) is used to calculate welfare.

5.2 Alternative Regimes

Next, I use the estimated structural parameters of the model and introduce different policy regimes. However before describing different policies, I initially report the results with no R&D subsidy. This simply means with the new notation that $\tau(\hat{q}) = 0$ and $\mathcal{T} = 0$. The following table reports the values of the subsidy rate, lump-sum tax $\mathcal{T}$, the average innovation quality $\overline{\left(\hat{\lambda}\right)}$, the average innovation probability $\bar{\phi}$, the initial consumption C_0, the aggregate steady-state growth rate of the economy g^* and the resulting welfare in (28) :

No Subsidy

τ	$\mathcal{T}$	$\overline{\left(\hat{\lambda}\right)}$	$\bar{\phi}$	C_0	g^*	Welfare
0	0	0.421	0.760	0.959	0.015	9.540

Table 7

This table indicates that without any R&D subsidy, the average innovation quality in the economy is equal to 0.421 and the average probability of innovation is 0.76. The average growth rate is 1.5 percent and the welfare is reported as 9.54.

5.2.1 Uniform Subsidy

Next I turn to the uniform R&D subsidy policy, where the subsidy rate does not depend on firm size. This is the most commonly observed subsidy policy around the world. Formally, this corresponds to $\tau\left(\hat{q}\right)=\tau\in[0,1]\,,\ \forall\hat{q}\in\hat{Q}$. Under optimal uniform subsidy policy regimes, the model generates the following results:

Optimal Uniform Subsidy

τ	$\mathcal{T}$	$\overline{\left(\hat{\lambda}\right)}$	$\bar{\phi}$	C_0	g^*	Welfare
0.748	0.046	0.517	0.874	0.814	0.021	15.746

Table 8

Without any subsidy, the growth rate of the economy is 1.5%. The optimal uniform subsidy rate of 75% boosts growth to 2.1%, at the expense of a 16% lower initial consumption. In addition, the government collects a payment of 0.05 units worth of the final good as a lump-sum tax to finance these subsidies. Initial consumption is lower as scarce resources (labor in this model) are being diverted from the production sector into the R&D sector, but the overall effect on welfare is strongly positive, as welfare rises by 65%.[17]

5.2.2 Size-Dependent Two-Level Subsidy

A two-level subsidy is close to the policy used in the UK, where the government provides two different levels of R&D subsidy. The policy sets a certain firm size threshold $y^{\ddagger} = y\left(\hat{q}^{\ddagger}\right)$ and for the firms below and above, two different subsidy rates are applied:

$$\tau\left(\hat{q}\right) = \begin{cases} \tau_s \text{ if } \hat{q} < \hat{q}^{\ddagger} \\ \\ \tau_b \text{ otherwise} \end{cases}, \quad \tau_s, \tau_b \in [0,1], \; \hat{q} \in \hat{Q}.$$

Clearly this policy nests the previous case. I choose the size threshold as the mean of $\hat{q}$ in the no-subsidy case, i.e, $\hat{q}^{\ddagger} \equiv \int \hat{q} d\Psi^{ns}\left(\hat{q}\right)$ where Ψ^{ns} corresponds to the invariant productivity distibution in the no-subsidy case. The results of this policy is reported below.

Optimal Two-level Subsidy

τ_s	τ_b	$\mathcal{T}$	$\overline{\left(\hat{\lambda}\right)}$	$\bar{\phi}$	C_0	g^*	Welfare
0.714	0.020	0.037	0.506	0.869	0.857	0.021	17.150

Table 9

The main result of this table is the heterogenous subsidy levels of small and large firms. The findings indicate that the optimal policy provides 71% of subsidy to small firms whereas this rate is only 2% for large firms. As in the uniform case, the subsidy improves the welfare of the economy through higher growth rate. However, this time the reduction in the inital consumption is mitigated because of the lower subsidy provided to large firms. This is mainly due to the fact that in equilibrum, the marginal innovations of small firms are of higher

[17] It should be noted that the percentage increase in welfare is sensitive to the assumption about the magnitude of the initial inverse quality index (0.1). Different numbers would generate the same qualitative results but quantitative results on welfare would be different.

quality which means that the same amount of subsidy translates into bigger and more frequent innovations. Compared to the uniform subsidy regime, the same growth gain is achieved at lower cost in terms of initial consumption, and leads to a 9% additional welfare gain on top of the uniform subsidy regime.

5.2.3 Size-Dependent Linear Subsidy

The final policy regime that I consider is a linear subsidy policy, providing a differential subsidy rate depending on size (productivity):

$$\tau(\hat{q}) = \tau_1 + \tau_2 \hat{q}, \quad \tau_1, \tau_2 \in \mathbb{R}, \ \hat{q} \in \hat{Q}$$

The results for this policy are reported in the table below

Optimal Linear Subsidy

τ_1	τ_2	$\mathcal{T}$	$\overline{\left(\hat{\lambda}\right)}$	$\bar{\phi}$	C_0	g^*	Welfare
1.323	-0.027	0.046	0.514	0.871	0.833	0.022	18.172

Table 10

The main finding relates to the negative sign of τ_2 which indicates that the optimal size-dependent subsidy rate is decreasing in firm size. The optimal subsidy rate for the smallest firm is 85%, and every unit increase in $\hat{q}$ reduces the subsidy rate by 0.027 so that the optimal rate becomes 0% for the largest firm. This policy generates the same average innovation quality (0.51) and probability (0.87) as in the case of the uniform policy with 2% less reduction in the initial consumption. The intuition for this is that the same average R&D efforts as in the uniform policy regime are maintained, but this effort is now mainly exerted by the small firms which are encouraged to do more R&D. The growth rate increases by 0.1 percentage point and welfare increases to 18.17 which is a 6% increase on top of the two-level subsidy policy.

Overall, the results show that size-dependent R&D subsidy policies have a significantly positive impact on welfare. By providing higher subsidies to smaller firms, the private under-investment in R&D can be mitigated. Even though the cost of such policies tends to reduce consumption initially, the higher steady-state growth compensates for this and yields a higher welfare.

6 Conclusion

This paper investigated the innovation dynamics of different sized firms and studied the relevant policy implications. It carried out its analysis in three distinct steps. First, using Compustat firms and their patent applications, it uncovered three major reduced form facts. The key finding of the reduced form analysis was that smaller firms undertake more radical and original innovations as measured by patent citations. The other two stylized empirical patterns were that firm size is negatively related to R&D intensity and sales growth.

Second, this paper outlined a theoretical general equilibrium model destined to explain the microeconomic causes of the observed reduced form relations. In this model technologically heterogeneous firms compete against each other for innovations in order to increase their operating profits. A novelty of this model is that firms can endogenously choose both the probability of innovation and the innovation quality. Inline with the reduced form results, the key fact of the model is that smaller firms undertake more radical innovations. Furthermore the model also explains other stylized empirical patterns related to the relationship between firm size and firm growth and firm size and R&D intensity. My theoretical results rely on two features: 1) The concavity of the profit and the value functions, and 2) the absence of strongly increasing returns to productivity. These two features provide greater incentives for smaller firms to increase their productivity. Therefore, they are more innovative, more R&D intensive, choose higher quality innovations and grow faster.

Third, the structural parameters of the model were estimated using the Simulated Method of Moments. These estimated parameters were then used in testing the effects of macro policies, in the form of public R&D subsidy for different sized firms on innovation. Three different regimes were compared: *i*) uniform (size-independent) R&D subsidy, *ii*) size-dependent two-level R&D subsidy (different rates below/above a certain threshold of firm size), *iii*) size-dependent linear R&D subsidy. The results of this analysis documented significant gains from moving to size-dependent R&D subsidy policies. Introducing size-dependent two-level R&D subsidy increased the welfare of the economy by 9 percent over the optimal uniform (size-independent) subsidy. More interestingly, this welfare gain is achieved by providing R&D subsidy only to small firms. Next, moving from two-level subsidy policy to a linear subsidy policy increased the welfare by 6 percent on top of the two level policy. Similarly, this policy also provides higher subsidy to smaller firms.

The empirical and theoretical results in this paper documented the additional contributions of the small firms to the aggregate innovation and growth. As a result, size-dependent policies have the potential of taking advantage of this heterogeneity among different sized firms. The focus of the paper has been only on R&D subsidy policy and as a possible extension of the current work,would be a study of the implications of entry and production subsidies on welfare. The challenge is to determine how close these alternative policies could bring the economy to the first best outcome.

7 Appendix A: Tables & Figures

Figure 1

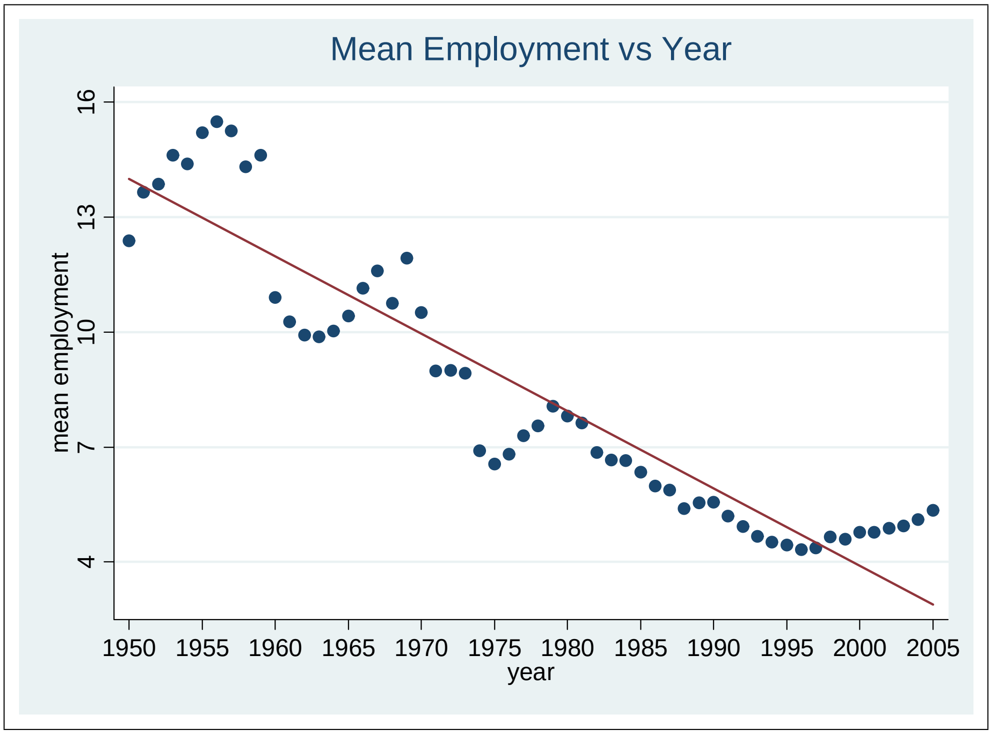

Figure 2

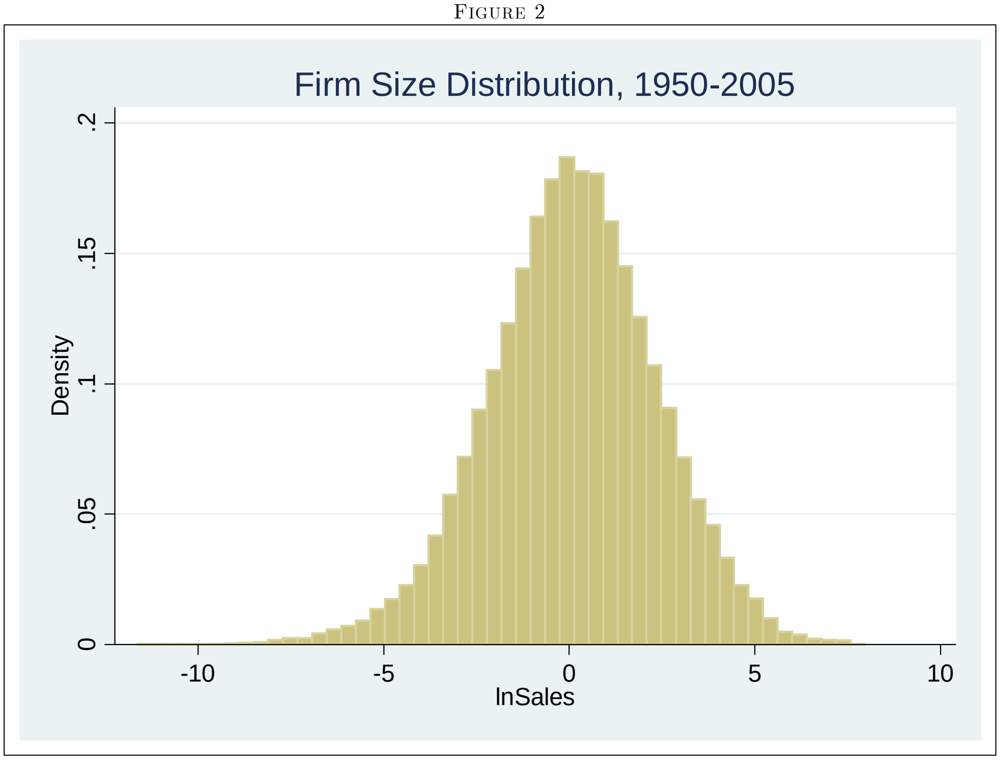

Figure 3

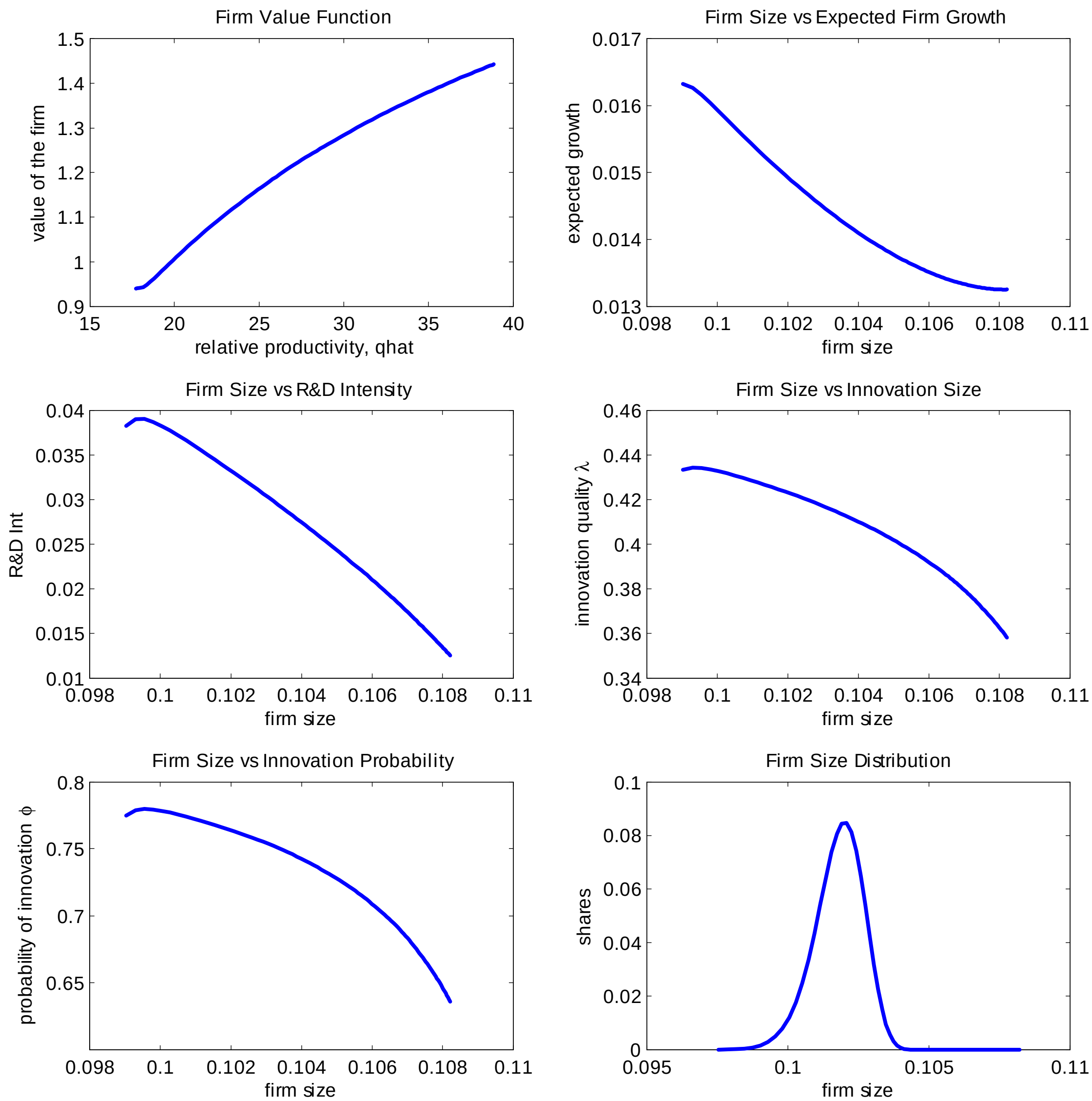
Firm Value Function
value of the firm
relative productivity, qhat
Firm Size vs Expected Firm Growth
expected growth
firm size
Firm Size vs R&D Intensity
R&D Int
firm size
Firm Size vs Innovation Size
innovation quality λ
firm size
Firm Size vs Innovation Probability
probability of innovation ϕ
firm size
Firm Size Distribution
shares
firm size

TABLE 1. SUMMARY STATISTICS

Variable	Obs	Mean	St Dev
Growth_t	111,869	0.118	0.412
$\text{Log}(\text{Sales}_t)^*$	118,343	0.073	2.294
$\text{Log}(\text{R\&D}_t)^*$	64,214	-3.228	2.231
Age	118,343	12.012	11.329
Subsidiary Dummy	118,343	0.019	0.134
Citations - 7 year	517,572	5.077	7.402
Originality	517,572	0.348	0.289
Claims	517,572	12.743	10.914

*Sale and R&D data is normalized by the GDP deflator

Table 2a. Growth Regressions

	OLS	OLS	Heckman 2-Step			IV	OLS
		(-100%)	survive	mfx	growth		
Variable	(1)	(2)	(3a)	(3b)	(4)	(5)	(6)
$\text{Log}(\text{Sales}_t)$	-0.037***	-0.025***	0.076***	0.008***	-0.028***	-0.028***	-0.175***
	(0.001)	(0.001)	(0.003)	(0.000)	(0.002)	(0.001)	(0.003)
Subsidiary Dummy			-0.422***	-0.062***			
			(0.046)	(0.009)			
Constant	0.331***	0.235***	0.6.583***		0.304***	0.241***	0.008***
	(0.026)	(0.030)	(0.106)		(0.052)	(0.022)	(0.014)
$\lambda-$ Inv Mills Ratio					0.666***		
					(0.098)		
Sector Dummy	Yes	Yes	Yes		Yes	Yes	Yes
Year Dummy	Yes	Yes	Yes		Yes	Yes	Yes
Firm Dummy	No	No	No		No	No	Yes
R^2	0.067	0.040				0.051	0.297
Obs	111,869	118,343	118,343 (6588 censored)			102,532	111,755

Note: The technique used for each estimation is reported on top of its column.

Dependent variable is sales growth=$\text{Sale}_{t+1}/\text{Sale}_t-1$.

Heteroskedasticity robust standard errors are reported in paranthesis.

*,**,*** indicate 10%, 5%, 1% significance.

Table 2b. Growth Regressions

	OLS	OLS	OLS	OLS	Heckman 2-Step		
		(Age>10)		LR	LR, survive	LR, mfx	LR, growth
Variable	(7)	(8)	(9)	(10)	(11a)	(11b)	(12)
Log(Sales$_t$)	-0.030***	-0.011***	-0.037***	-0.028***	0.116***	0.046***	-0.026***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.010)	(0.004)	(0.006)
Age	-0.003***						
	(0.000)						
Subsidiary Dummy					-1.605***	-0.412***	
					(0.373)	(0.035)	
Constant	0.222***	0.154***	0.124***	0.217***	0.492**		0.191***
	(0.015)	(0.013)	(0.001)	(0.024)	(0.234)		(0.142)
Sector Dummy	Yes	Yes	No	Yes	Yes		Yes
Year Dummy	Yes	Yes	No	No	No		No
Firm Dummy	No	No	No	No	No		No
Year*Sector D.	No	No	Yes	No	No		No
R^2	0.072	0.044	0.044	0.221			
Obs	111,755	47,942	111,755	1,489	3,289 (1800 censored)		

Note: The technique used for each estimation is reported on top of its column.

Dependent variable is sales growth=Sale$_{t+1}$/Sale$_t$$-1$.

Heteroskedasticity robust standard errors are reported in paranthesis.

*,**,*** indicate 10%, 5%, 1% significance.

Table 3a. R&D Intensity Regressions

	OLS	Heckman 2-Step			IV
		report	mfx	log(R&D/Sales$_t$)	
Variable	(1)	(2a)	(2b)	(3)	(4)
Log(Sales$_t$)	-0.265***	0.101***	0.040***	-0.234***	-0.225***
	(0.003)	(0.002)	(0.001)	(0.005)	(0.003)
Constant	-4.307***	-2.241***		-6.092***	-4.491***
	(0.157)	(0.094)		(0.333)	(0.157)
$\lambda-$ Inv Mills Ratio				0.669***	
				(0.102)	
Sector Dummy	Yes	Yes		Yes	Yes
Year Dummy	Yes	Yes		Yes	Yes
R^2	0.465				0.443
Obs	65,230	120,238 (55,008 cnsrd)			60,710

Note: The technique used for each estimation is reported on top of its column.
Dependent variable is log R&D intensity= $\ln(\text{R\&D}_t/\text{Sale}_t)$.
Heteroskedasticity robust standard errors are reported in paranthesis.
*,**,*** indicate 10%, 5%, 1% significance.

Table 3b. R&D Regressions

Variable	OLS (5)	OLS (6)	OLS (Age>10) (7)	OLS (8)	OLS LR (9)
Log(Sales$_t$)	-0.384***	-0.248***	-0.092***	-0.257***	-0.298***
	(0.007)	(0.003)	(0.004)	(0.003)	(0.014)
Age		-0.006***			
		(0.001)			
Constant	-4.560***	-4.324***	-3.909***	-3.227***	-1.594***
	(0.148)	(0.155)	(0.098)	(0.005)	(0.280)
Sector Dummy	Yes	Yes	Yes	No	Yes
Year Dummy	Yes	Yes	Yes	No	No
Firm Dummy	Yes	No	No	No	No
Year*Sector D.	No	No	No	Yes	No
R^2	0.881	0.466	0.390	0.483	0.456
Obs	65,230	65,230	30,831	65,230	2,418

Note: The technique used for each estimation is reported on top of its column.
Dependent variable is log R&D intensity= ln(R&D$_t$/Sale$_t$).
Heteroskedasticity robust standard errors are reported in paranthesis.
*,**,*** indicate 10%, 5%, 1% significance.

Table 4a. Patent Citation Regressions

	OLS	OLS	Negative Binomial	0-Inf'd Neg Binomial
	(75-95)	(88-92)	(88-92)	(88-92)
Variable	(1)	(2)	(3)	(4)
$Log(Sales_t)$	-0.154***	-0.324***	-0.045***	-0.040***
	(0.008)	(0.018)	(0.002)	(0.002)
Count				
Constant	3.119***	8.146***	2.425***	0.120***
	(0.137)	(0.080)	(0.012)	(0.976)
IPC4 Dummy	Yes	Yes	Yes	Yes
Year Dummy	Yes	Yes	Yes	Yes
R^2	0.116	0.142		
Obs	346,719	123,473	91,055	91,055

Note: The technique used for each estimation is reported on top of its column.
Dependent variable is the number of citation received by each patent
Heteroskedasticity robust standard errors are reported in paranthesis.
*,**,*** indicate 10%, 5%, 1% significance.

Table 4b. Patent Citation Regressions

Variable	Negative Binomial Best, (88-92) (5)	OLS Originality (6)	OLS Claims (7)
Log(Sales$_t$)	-0.029***	-0.001***	-0.601***
	(0.003)	(0.000)	(0.021)
Count	0.068***		
	(0.001)		
Constant	2.789***	-0.299***	16.297***
	(0.903)	(0.002)	(0.120)
IPC4 Dummy	Yes	Yes	Yes
Year Dummy	Yes	Yes	Yes
R^2	0.045	0.111	0.050
Obs	30,865	30,865	92,718

Note: The technique used for each estimation is reported on top of its column.
Dependent variable is the number of citation received by each patent
Heteroskedasticity robust standard errors are reported in paranthesis.
*,**,*** indicate 10%, 5%, 1% significance.

8 Appendix B: Proofs

Proof of Proposition (1)**.**

Fact 4
The relevant state space is $\hat{Q} = [\hat{q}_{\min}, \hat{q}_{\max}]$. *This implies that* $\hat{Q}$ *is a Borel set in* $\mathbb{R}$.

Fact 5 $\phi_i \in Z \equiv [0,1]$. *Z is a compact Borel set in* $\mathbb{R}$. *In addition, the transition function is*

$$\mathbf{P}\left(\hat{q}_i, \hat{q}'\right) = \begin{cases} \phi_i \text{ if } \hat{q}' = \frac{\hat{q}_i + \hat{\lambda}_i}{1+g} \\ 1-\phi_i \text{ if } \hat{q}' = \frac{\hat{q}_i}{1+g} \\ 0 \text{ otherwise} \end{cases}.$$

In order to show that the transition function has the Feller property, for now we will make the following conjecture and later on we will show that the conjecture is, in fact, true.

Conjecture 6 *Firm i's optimal R&D decisions,* ϕ_i *and* $\hat{\lambda}_i$ *are continuous in* $\hat{q}_i$.

Now we can continue with our discussion on $\mathbf{P}$. *Since* ϕ_i *and* $\hat{\lambda}_i$ *are continuous in* $\hat{q}_i$ *and* Ω *is also a continuous function,*

$$T\Omega\left(\hat{q}_i\right) = \phi_i \Omega\left(\frac{\hat{q}_i + \hat{\lambda}_i}{1+g}\right) + \left(1-\phi_i\right)\Omega\left(\frac{\hat{q}_i}{1+g}\right)$$

is also continuous. This shows that $\mathbf{P}$ *has the Feller property.*

Fact 7 $\hat{\lambda}_i, \phi_i \in \Gamma \equiv [0,1]^2$ *and* Γ *is clearly nonempty, compact-valued and continuous.*

Fact 8 $\tilde{\pi}\left(\hat{q}_i\right) = \frac{1}{2\gamma}\left(\frac{\gamma\bar{y}}{Y} + \overline{\left(\frac{1}{\hat{q}_i}\right)} - \frac{1}{\hat{q}_i}\right)^2$, *is bounded on* $\hat{Q}$ *by* $\pi_{\max} = \frac{1}{2\gamma}\left(\frac{\gamma\bar{y}}{Y} + \overline{\left(\frac{1}{\hat{q}_i}\right)}\right)^2$ *and is continuous and* $\beta \in (0,1)$.

Given Facts $(4)-(8)$ the existence and uniqueness of the value function follows from Theorem 9.6 in Stokey-Lucas (1989).

Fact 9 $\pi\left(\hat{q}_i\right) = \frac{1}{2\gamma}\left(\frac{\gamma\bar{y}}{Y} + \overline{\left(\frac{1}{\hat{q}_i}\right)} - \frac{1}{\hat{q}_i}\right)^2$, $\hat{q} \in \hat{Q}$ *is a strictly increasing function.*

Fact 10 Γ *is increasing.*

The fact that the value function is strictly increasing follows from the Facts $(4)-(10)$ and Theorem 9.7 in Stokey-Lucas $(1989\ p.264)$.

Fact 11 *Since* $\pi\left(\hat{q}_i\right)$ *is strictly concave in* $\hat{q}_i$, *we have*

$$\pi\left(\theta\hat{q}_i + (1-\theta)\frac{\hat{q}_i + \hat{\lambda}_i}{1+g}\right) > \theta\pi\left(\hat{q}_i\right) + (1-\theta)\,\pi\left(\frac{\hat{q}_i + \hat{\lambda}_i}{1+g}\right).$$

Fact 12 Γ *is convex.*

Strict concavity of the value function follows from the Facts $(4) - (8)$ and $(11) - (12)$ and Theorem 9.8 in Stokey-Lucas (1989 p.265).

The fact that $\hat{\lambda}(\hat{q}_i)$ and $\phi(\hat{q}_i)$ are continuous and single valued functions also follow from the same proposition. Note that this proves our conjecture that $\hat{\lambda}(\hat{q}_i)$ and $\phi(\hat{q}_i)$ are continuous. ■

Derivation of (24) **.** In steady state $\overline{(1/\hat{q}_i)}$ has to be constant. This implies that

$$\overline{\left(\frac{1}{\hat{q}_i}\right)} = \frac{1}{N}\int_{i\in I^*}\frac{1}{\hat{q}_i}di$$

$\int_{i\in I^*}(1/\hat{q}_i)\,di$ has to be constant. Therefore

$$\int_{i\in I^*}\left[\phi_i\frac{w(1+g)}{q_i+\lambda_i} + (1-\phi_i)\frac{w(1+g)}{q_i}\right]di = \int_{i\in I^*}\frac{w}{q_i}di$$

where the left-hand side is the next period's aggregate cost and the right-hand side is today's. After some simplification

$$(1+g)\int_{i\in I^*}\left[\frac{\phi_i}{q_i+\lambda_i} + \frac{(1-\phi_i)}{q_i}\right]di = \int_{i\in I^*}\frac{1}{q_i}di$$

$$g = \frac{\int_{i\in I^*}\frac{1}{q_i}di}{\int_{i\in I^*}\frac{q_i+\lambda_i(1-\phi_i)}{q_i^2+\lambda_i q_i}di} - 1$$

■

Proof of Proposition (2)**.** Since $\phi(\hat{q})$ is continuous, it is bounded on $\hat{Q}$. Therefore $0 < \phi_{\min} \leq \phi\left(\hat{Q}\right) \leq \phi_{\max} < 1$ thanks to the Inada conditions. Let's consider $\hat{q}_{out}$ as the state where all exiters are collected. Then $\exists n^* \equiv \min\{n \in \mathbb{N}\} \geq 1$ such that

$$\frac{\hat{q}_{\max}}{(1+g)^n} < \hat{q}_{\min}$$

equivalently

$$\frac{\log(\hat{q}_{\max}) - \log(\hat{q}_{\min})}{\log(1+g)} < n.$$

Then for any $\hat{q} \in \hat{Q}$

$$P^{n^*}(\hat{q}, \{\hat{q}_{out}\}) \geq (1-\phi_{\max})^{n^*} > \epsilon > 0$$

Since for all $A \subset 2^{\hat{Q}}$, either $\hat{q}_{out} \in A$ or $\hat{q}_{out} \notin A$, we have that

$$P^{n^*}(\hat{q}, A) \geq P^{n^*}(\hat{q}, \{\hat{q}_{out}\}) > \epsilon$$

or

$$P^{n^*}(\hat{q}, A^c) \geq P^{n^*}(\hat{q}, \{\hat{q}_{out}\}) > \epsilon.$$

This prove the existence of $\epsilon > 0$, an integer $n^* \geq 1$ such that for any $A \subset 2^{\hat{Q}}$, either $P^{n^*}(\hat{q}, A) \geq \epsilon$ or $P^{n^*}(\hat{q}, A^c) \geq \epsilon \ \forall \hat{q} \in \hat{Q}$. This is *Condition M* in Stokey-Lucas (1989) p.348.

Then the existence of the invariant distribution follows from Theorem 11.12 in Stokey-Lucas (1989 *p*.350). ■

Proof of Proposition (3)**.** Let $m = \left[\hat{q}_{exit,} N, \tilde{w}, p_0, g, \lambda(Q), \phi(Q), \kappa(Q), \Psi, \tilde{V}(Q), \tilde{W}(Q)\right] \in \mathbf{M}$ be a generic argument of the following operator

$$\mathbf{\Phi}(m) : \mathbf{M} \to \mathbf{M} \tag{31}$$

such that

$$\mathbf{\Phi}(m) \equiv \left[\Phi_{\hat{q}_{exit}}(m), \Phi_N(m), \Phi_{\tilde{w}}(m), \Phi_{p_0}(m), \Phi_g(m), \Phi_{\tilde{V}}(m), \Phi_{\tilde{W}}(m), \Phi_{\mathbf{\Psi}}(m), \Phi_\lambda(m), \Phi_\phi(m), \Phi_\kappa(m)\right]^T$$

I define the elements of the operator $\mathbf{\Phi}$ as follows. $\Phi_{\tilde{V}}(m)$ is the right-hand side of the Bellman equation in (20), $\Phi_{\tilde{W}}(m)$ is as in (21), The R&D decisions $\Phi_\lambda(m), \Phi_\phi(m), \Phi_\kappa(m)$ are the maximizers of the $\tilde{V}$ and $\tilde{W}$ and $\mathbf{\Psi}$ is the invariant distribution obtained for any given m whose existence is proven in proposition (2). The values for all $\Phi_\lambda(m), \Phi_\phi(m), \Phi_\kappa(m), \Phi_{\mathbf{\Psi}}(m) \in [0,1] \ \forall q \in Q$ belong to the compact convex set. Similarly the values $\Phi_{\tilde{V}}(m), \Phi_{\tilde{W}}(m) \in \left[\Lambda, \frac{\gamma \tilde{q}_{\max}^2}{2(1-\beta)}\right] \ \ \forall i \in I$ are also in a compact and convex set.

The remaining parts of the operator $\Phi_N(m), \Phi_{\hat{q}_{exit}}(m), \Phi_{\tilde{w}}(m), \Phi_{p_0}(m), \Phi_g(m)$ are as follows. The first operator is for the total number of firms, N. Using the the free entry condition, the Bellman equation and the profit function as expressed in (**??**)

$$\begin{aligned} N &= \max\left\{0, \min\left\{N_{\max};\ \frac{N}{\tilde{\chi}} \int \left\{ \begin{array}{c} \frac{Y}{2\gamma}\left(\frac{2\gamma}{\alpha N} + \overline{\left(\frac{1}{\hat{q}}\right)} - \frac{1}{\hat{q}_i}\right)^2 - \tilde{w} h\left(\frac{\lambda(\hat{q})}{Y}, \phi(\hat{q})\right) \\ + \frac{1}{1+r}\left[\phi(\hat{q}) \tilde{W}\left(\hat{q} + \hat{\lambda}(\hat{q})\right) + (1 - \phi(\hat{q})) \tilde{W}(\hat{q})\right] \end{array} \right\} d\Psi(\hat{q}) \right\}\right\} \\ &\equiv \Phi_N(\mathbf{m}) \end{aligned}$$

Clearly this mapping assumes values from the set $[0, N_{\max}]$ which is compact and convex.

Then for the minimum level of productivity $q_{\min}$, we use the free exit condition

$$\Lambda = V(\hat{q})$$

Since $\tilde{V}$ is strictly increasing in $\hat{q}$, we can define the following function

$$\begin{aligned} \hat{q}_{exit} &= \min\left\{q_{\max}, \max\left\{q_{\min}, \tilde{V}^{-1}\left(\tilde{\Lambda}\right)\right\}\right\} \\ &\equiv \Phi_{\hat{q}_{exit}}(m) \end{aligned}$$

Again, since $\tilde{V}$ is continuous in $\hat{q}$ so is its inverse and $\tilde{V}$ being continuous in m implies that $\hat{q}_{exit}$ is continuous in m. In addition, it maps into a compact set[18].

[18]The minimum value $\tilde{q}_0$ can take is when $\frac{\gamma \bar{q}}{Y} - \overline{\left(\frac{1}{\hat{q}}\right)}$ is maximum which is simply $\gamma \tilde{q}_{\max} - \frac{1}{\tilde{q}_{\max}} > 0$ given that assumption that $q_{\max}^2 > 1/\gamma$.

The operator for the growth rate g is defined using its expression in Theorem **??**

$$\begin{aligned} g &= \min\left\{\frac{1}{\hat{q}_{\min}}, \frac{\int_{i\in I^*}\frac{1}{q_i}di}{\int_{i\in I^*}\frac{q_i+\lambda_i(1-\phi_i)}{q_i^2+\lambda_i q_i}di} - 1\right\} \\ &\equiv \Phi_g(m) \end{aligned} \tag{32}$$

This mapping assumes values from $[0, 1/\hat{q}_{\min}]$ which is both compact and convex.

The operator for the price of the numeraire is

$$\begin{aligned} p_0 &= \min\left\{p_{0,\min}, \max\left\{0, \frac{\frac{\alpha^2}{2}\left[1+\frac{\gamma}{2Y^2}\int_{i\in I} y_{i,t}^2 di\right] - 2\alpha\overline{(1/\hat{q})}_t}{\left(\frac{4\gamma}{N}+\delta\right)}.\right\}\right\} \\ &\equiv \Phi_{p_0}(m) \end{aligned}$$

Finally, for the relative wage, I use labor market clearing,

$$\begin{aligned} \tilde{w} &= \max\left\{0, \min\left\{\tilde{w}_{\max}; \frac{\left(\frac{2}{N}\overline{\left(\frac{1}{\tilde{q}}\right)} + \gamma\bar{h} - \frac{\gamma}{N}\right)}{\overline{\left(\frac{1}{\tilde{q}^2}\right)}}\right\}\right\} \\ &\equiv \Phi_{\tilde{w}}(m) \end{aligned}$$

which takes values only from $[0, \tilde{w}_{\max}]$

Define $\mathbf{M} \equiv [\hat{q}_{\min}, \hat{q}_{\max}] \times [0, N_{\max}] \times [\tilde{w}_{\min}, \tilde{w}_{\max}] \times [0, g_{\max}] \times \Pi_{i\in I}[0,1]^4 \times \Pi_{i\in I}\left[0, \tilde{V}_{\max}\right]^2$ and a typical element of the set to be $m = \left\langle \hat{q}_{exit}, N, \tilde{w}, g, \lambda, \phi, \kappa, \Psi, \tilde{V}, \tilde{W}\right\rangle \in \mathbf{M}$. In order to show the existence of the equilibirum, I need to show that

$$\Phi : \mathbf{M} \to \mathbf{M}$$

that was defined in (31) has a fixed point $\Phi(m^*) = m^*$. First, $\mathbf{M}$ is a compact in the product topology since it is the cartesion product of infinitely many compact sets (Aliprantis and Border, 1999, Theorem 2.57, p.52). Clearly $\mathbf{M}$ is non-empty and convex. Moreover $\mathbf{M}$ is a subset of locally Hausdorf space (Aliprantis and Border, 1999, Lemma 5.54, p.192).

Next, I will consider them individually and define their maximum values.

Fact 13 ($\tilde{w}$) *Note that the amount of Y has to be bigger than the expenditure made on all the differentiated goods, $Y \geq \int q_i p_i di$. In addition, the labor expense at the intermediate sector (both production and R&D) must be met by the revenues, $\int q_i p_i di > \int w l_i di = w$ since the total labor supply is 1. These two observations imply*

$$1 \equiv \tilde{w}_{\max} > w/Y \equiv \tilde{w}.$$

On the other hand, $\int q_i di < \int q_{\max} l_i di < w\hat{q}_{\max}$. Dividing everything by Y

$$\tilde{w}_{\min} \equiv \frac{2}{\alpha\hat{q}_{\max}} < \tilde{w}.$$

Fact 14 **(g)** *The operator takes nonnegative values in* (32)*. On the other hand, the highest increases will be less than when all the firms are in the smallest technology and innovate the maximum size with probability 1,*

$$g < \frac{\int_{i \in I^*} \frac{1}{\hat{q}_{\min}} di}{\int_{i \in I^*} \frac{1}{\hat{q}_{\min}+1} di} - 1 \equiv g_{\max} = \frac{1}{\hat{q}_{\min}}$$

Fact 15 **($\tilde{V}$)** *The value of a firm is bounded from below by 0. On the other hand, the upper limit for the profit that can be generated is*

$$\tilde{\pi}_i < \tilde{\pi}_{\max} \equiv \frac{1}{\gamma} \left(\frac{\alpha}{2} - \frac{\delta}{\alpha} - \frac{1}{\hat{q}_{\max}} \right)^2$$

Therefore

$$0 < \tilde{V}_i < \tilde{V}_{\max} \equiv \frac{1+r}{\gamma r} \left(\frac{\alpha}{2} - \frac{\delta}{\alpha} - \frac{1}{\hat{q}_{\max}} \right)^2$$

Since Φ maps compact convex set $\mathbf{M}$ into itself and since it is continuous, by Brouwer-Schauder-Tychonoff Fixed Point Theorem (Aliprantis and Border, 1999, p.550), the fixed point exists. ■

Proof of Theorem (1)**.** Let us recall the first order condition,

$$\hat{\lambda}_i \quad : \quad \frac{\beta \phi_i}{1+g} v' \left(\frac{\hat{q}_i + \hat{\lambda}_i}{1+g} \right) = \tilde{w} h_{\hat{\lambda}} \left(\hat{\lambda}_i, \phi_i \right) \tag{33}$$

$$\phi_i \quad : \quad \beta \left[v \left(\frac{\hat{q}_i + \hat{\lambda}_i}{1+g} \right) - v \left(\frac{\hat{q}_i}{1+g} \right) \right] = \tilde{w} h_{\phi} \left(\hat{\lambda}_i, \phi_i \right) \tag{34}$$

We proceed case by case.

Case 1. Assume an initial equilibrium with $\left(\hat{q}_{old}, \hat{\lambda}_{old}, \phi_{old} \right)$ and by contradiction assume a new equilibrium $\left(\hat{q}_{new}, \hat{\lambda}_{new}, \phi_{new} \right)$ such that $\hat{q}_{new} > \hat{q}_{old}$, $\hat{\lambda}_{new} \geq \hat{\lambda}_{old}$ and $\phi_{new} \leq \phi_{old}$. Then (33) implies

$$\begin{aligned} \frac{\beta}{1+g} v' \left(\frac{\hat{q}_{new} + \hat{\lambda}_{new}}{1+g} \right) &< \frac{\beta}{1+g} v' \left(\frac{\hat{q}_{old} + \hat{\lambda}_{old}}{1+g} \right) \\ &= \frac{\tilde{w} h_{\hat{\lambda}} \left(\hat{\lambda}_{old}, \phi_{old} \right)}{\phi_{old}} \\ &\leq \frac{\tilde{w} h_{\hat{\lambda}} \left(\hat{\lambda}_{new}, \phi_{old} \right)}{\phi_{old}} \\ &\leq \frac{\tilde{w} h_{\hat{\lambda}} \left(\hat{\lambda}_{new}, \phi_{new} \right)}{\phi_{new}} \end{aligned}$$

Here the first line used the fact that v is concave (therefore its slope is a decreasing function). The second line is just the initial equilibrium condition and the last two lines are using the assumptions of the $h(.,.)$ function.

Case 2. Similarly assume an initial and new equilibria with $\hat{q}_{new} > \hat{q}_{old}, \hat{\lambda}_{new} \leq \hat{\lambda}_{old}, \phi_{new} \geq \phi_{old}$. Define a function of

$$F(\hat{q}) \equiv \left[\frac{v\left(\frac{\hat{q}+\hat{\lambda}}{1+g}\right) - v\left(\frac{\hat{q}}{1+g}\right)}{\hat{\lambda}} \right].$$

This implies 2 cases: 1) $F(\hat{q}_{new}) < F(\hat{q}_{old})$ 2) $F(\hat{q}_{new}) > F(\hat{q}_{old})$. Consider case 1. Then

$$\begin{aligned} F(\hat{q}_{new}) &< F(\hat{q}_{old}) \\ &= \frac{\tilde{w}h_{\phi}\left(\hat{\lambda}_{old}, \phi_{old}\right)}{\beta\hat{\lambda}_{old}} \\ &\leq \frac{\tilde{w}h_{\phi}\left(\hat{\lambda}_{old}, \phi_{new}\right)}{\beta\hat{\lambda}_{old}} \\ &\leq \frac{\tilde{w}h_{\phi}\left(\hat{\lambda}_{new}, \phi_{new}\right)}{\beta\hat{\lambda}_{new}} \end{aligned}$$

which is a contradiction.

Now consider case 2. Then since $\hat{\lambda}$ and ϕ are continuous in $\hat{q}$, $\exists \hat{q}^* > \hat{q}$ such that (case 2.1) $F(\hat{q}^*) = F(\hat{q}_{old})$ and $\phi^* > \phi_{old}$ or (case 2.2) $F(\hat{q}^*) \geq F(\hat{q}_{old})$ and $\phi^* = \phi_{old}$. Consider case 2.1 :

$$\begin{aligned} F(\hat{q}^*) &= F(\hat{q}_{old}) \\ &= \frac{\tilde{w}h_{\phi}\left(\hat{\lambda}_{old}, \phi_{old}\right)}{\beta\hat{\lambda}_{old}} \\ &< \frac{\tilde{w}h_{\phi}\left(\hat{\lambda}_{old}, \phi^*\right)}{\beta\hat{\lambda}_{old}} \\ &\leq \frac{\tilde{w}h_{\phi}\left(\hat{\lambda}^*, \phi_{new}\right)}{\beta\hat{\lambda}^*} \end{aligned}$$

which is a contradiction to the fact that $\left(\hat{q}^*, \hat{\lambda}^*, \phi^*\right)$ is a solution.

Next consider case 2.2 : $F(\hat{q}^*) \geq F(\hat{q}_{old})$ and $\phi^* = \phi_{old}$. This implies $\hat{q}^* + \hat{\lambda}^* < \hat{q}_{old} + \hat{\lambda}_{old}$.(Draw a graph) (no leapfrogging condition) Then

$$\begin{aligned} \frac{\beta\phi^*}{1+g} v'\left(\frac{\hat{q}^* + \hat{\lambda}^*}{1+g}\right) &> \frac{\beta\phi_{old}}{1+g} v'\left(\frac{\hat{q}_{old} + \hat{\lambda}_{old}}{1+g}\right) \\ &= \tilde{w}h_{\hat{\lambda}}\left(\hat{\lambda}_{old}, \phi_{old}\right) \\ &> \tilde{w}h_{\hat{\lambda}}\left(\hat{\lambda}^*, \phi_{old}\right) \\ &= \tilde{w}h_{\hat{\lambda}}\left(\hat{\lambda}^*, \phi^*\right) \end{aligned}$$

which is a contradiction.

Case 3. Finally assume that $\hat{\lambda}(\hat{q})$ and $\phi(\hat{q})$ both are increasing functions. Concavity and boundednes of v implies $\lim_{\hat{q}\to\infty} v'(\hat{q}) = 0$ and accordingly $\lim_{\hat{q}\to\infty} \frac{\beta\phi(\hat{q}^{**})}{1+g} v'\left(\frac{\hat{q}^{**}+\hat{\lambda}(\hat{q}^{**})}{1+g}\right) = 0$. Take $\hat{q}^* \in Q$ and let $\xi \equiv \tilde{w}h_{\hat{\lambda}}\left(\hat{\lambda}(\hat{q}^*), \phi(\hat{q}^*)\right)$. Then $\exists \hat{q}^{**} > \hat{q}^* \in Q$ such that $\frac{\beta\phi(\hat{q}^{**})}{1+g} v'\left(\frac{\hat{q}^{**}+\hat{\lambda}(\hat{q}^{**})}{1+g}\right) < \xi \leq \tilde{w}h_{\hat{\lambda}}\left(\hat{\lambda}(\hat{q}^{**}), \phi(\hat{q}^{**})\right)$ where the first inequality uses the limit condition and the second inequality uses the monotonicity of the R&D decisions. The result is a contradiction to the fact that $\left(\hat{\lambda}(\hat{q}^{**}), \phi(\hat{q}^{**})\right)$ is an equilibrium. Then it must be the case that $\hat{\lambda}(\hat{q})$ or $\phi(\hat{q})$ is strictly decreasing for some $\hat{q}^* < \hat{q}^{**}$. Since it was proven under Case1-2 that $\hat{\lambda}(\hat{q})$ and $\phi(\hat{q})$ cannot move in the opposite directons ever, the only possibility is that both $\hat{\lambda}(\hat{q})$ are $\phi(\hat{q})$ strictly decreasing functions.

QED.

■

Proof of Theorem (2)**.**
R&D spending can be expressed as

$$\begin{aligned} R\&D\ Intensity &= \frac{wh_i}{y_i} \\ &= \frac{\tilde{w}h\left(\hat{\lambda}_i, \phi_i\right)}{\frac{Y}{2\gamma}\left(\frac{\alpha y_0 - \delta D}{Y} - c_i\right)^2} \end{aligned}$$

Proposition (**??**) showed that both $\hat{\lambda}_i$ and ϕ_i are decreasing in $\hat{q}_i$. On the other hand, the denominator is decreasing for $c_i \in C$ and $c_i'(\hat{q}_i) < 0$ which implies teh denominator is increasing in $\hat{q}_i$.Therefore the R&D intensity is decreasing in firm size. ■

Proof of Proposition (4)**.**

$$y_i = \frac{Y}{\gamma}\left(\frac{\alpha y_0 - \delta D}{2Y} - \frac{w}{q}\right)$$

Then the expected firm growth

$$\mathbb{E}g(y_i) = g + \mathbb{E}g(f_i)$$

where $f_i \equiv \left(\kappa - \frac{w}{q}\right)$ and $\kappa \equiv \frac{\alpha y_0 - \delta D}{2Y}$. We can express $\mathbb{E}g\left(f_i\right)$ as

$$
\begin{aligned}
g\left(f_i\right) &= \frac{\kappa - \left[\phi_i \frac{w(1+g)}{q_i+\lambda_i} + \left(1-\phi_i\right)\frac{w(1+g)}{q_i}\right]}{\kappa - \frac{w}{q}} - 1 \\
&= \frac{-\left[\phi_i \frac{w(1+g)}{q_i+\lambda_i} + \left(1-\phi_i\right)\frac{w(1+g)}{q_i}\right] + \frac{w}{q}}{\kappa - \frac{w}{q}} \\
&= \frac{\frac{\lambda_i \phi_i (1+g)}{q_i+\lambda_i} - g}{\frac{q}{w}\kappa - 1} \\
&= \frac{\frac{\hat{\lambda}_i \phi_i (1+g)}{\hat{q}_i+\hat{\lambda}_i} - g}{\kappa \hat{q}_i - 1} \qquad (35)
\end{aligned}
$$

■

Proof of Theorem (3)**.**

Now for the second part, take the derivative wrt $\hat{q}_i$. First let us look at the derivative of $\Upsilon\left(\hat{q}_i\right) \equiv \frac{\hat{\lambda}(\hat{q}_i)\phi(\hat{q}_i)}{\hat{q}_i+\hat{\lambda}(\hat{q}_i)}$

$$
\begin{aligned}
\Upsilon'\left(\hat{q}_i\right) &= \frac{\left[\hat{\lambda}'\left(\hat{q}_i\right)\phi\left(\hat{q}_i\right) + \hat{\lambda}\left(\hat{q}_i\right)\phi'\left(\hat{q}_i\right)\right]\left[\hat{q}_i + \hat{\lambda}\left(\hat{q}_i\right)\right] - \left[1 + \hat{\lambda}'\left(\hat{q}_i\right)\right]\hat{\lambda}\left(\hat{q}_i\right)\phi\left(\hat{q}_i\right)}{\left[\hat{q}_i + \hat{\lambda}\left(\hat{q}_i\right)\right]^2} \\
&= \frac{\hat{\lambda}'\left(\hat{q}_i\right)\phi\left(\hat{q}_i\right)\hat{q}_i + \hat{\lambda}\left(\hat{q}_i\right)\phi'\left(\hat{q}_i\right)\hat{q}_i + \hat{\lambda}^2\left(\hat{q}_i\right)\phi'\left(\hat{q}_i\right) - \hat{\lambda}\left(\hat{q}_i\right)\phi\left(\hat{q}_i\right)}{\left[\hat{q}_i + \hat{\lambda}\left(\hat{q}_i\right)\right]^2} \\
&< 0
\end{aligned}
$$

where the second equality uses Proposition 1.

$$
\frac{dg\left(f_i\right)}{d\hat{q}_i} = \frac{\Upsilon'\left(\hat{q}_i\right)\left(\kappa\hat{q}_i - 1\right) - \kappa\left[\Upsilon\left(\hat{q}_i\right) - g\right]}{\left(\kappa\hat{q}_i - 1\right)^2}.
$$

Since $g_i \geq g$, from equation (35) we have $\Upsilon\left(\hat{q}_i\right) \geq g$. Using this result in the above ratio

$$
\frac{dg\left(f_i\right)}{d\hat{q}_i} < 0.
$$

■

9 References

- **Acemoglu, Daron and Akcigit, Ufuk (2008)** "State Dependent Intellectual Property Rights Policy" *MIT mimeo.*

- **Adda, Jerome and Russell Cooper (2003)** *Dynamic Economics: Quantitative Methods and Applications*, MIT Press, Cambridge MA.

- **Aghion, Philippe, Christopher Harris and John Vickers (1997)** "Competition and Growth with Step-by-Step Innovation: An Example" *European Economic Review*, 41, 771-782.

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

- **Aghion, Philippe and Peter Howitt (2002)** "A Model of Growth Through Creative Destruction" *Econometrica*, 110, 323-351.

- **Amirkhalkhali, Saleh and Arun K. Mukhopadhyay (1993)** "The Influence of Size and R&D on the Growth of Firms in the U.S" *Eastern Economic Journal*, 19(2), 223-233.

- **Anselin, Luc, Atilla Varga, and Zoltan Acs (1997)** "Local Geographic Spillovers between University Research and High Technology Innovations" *Journal of Urban Economics* 42(3), 422-448

- **Anselin, Luc, Atilla Varga, and Zoltan Acs (2000)** "Geographical spillovers and university research: A spatial econometric perspective." *Growth and Change* 31, 501-15.

- **Blair, John M. (1972)** *Economic Concentration: Structure, Behaviour and Public Policy*, Harcourt Brace Jovanovich, New York.

- **Bottazzi, Giulio and Angelo Secchi (2003)** "Sectoral Specifities in the Dynamics of U.S. Manufacturing Firms" *LEM Papers Series*, 2003/18.

- **Bound, John, Clint Cummins, Zvi Griliches, Bronwyn H. Hall and Adam B. Jaffe (1984)** "Who Does R&D and Who Patents?" in *R&D, Patents and Productivity.* Ed. Z. Griliches, University of Chicago Press, Chicago.

- **Campbell, Jeffrey R. and Hugo A. Hopenhayn (2002)** "Market Size Matters" *NBER Working Papers*, No 9113.

- **Caves, Richard E. (1998)** "Industrial Organization and New Findings on the Turnover and Mobility of Firms" *Journal of Economie Literature*, 36, 1947-1982.

- **Cohen, Wesley M.,Richard C. Levin and David C. Mowery (1987)** "Firm Size and R&D Intensity: A Re-Examination" *Journal of Industrial Economics*, 4, 543-565.

- **Cohen, Wesley M.(1995)** "Empirical Studies of Innovative Activity", in Stoneman, P. (ed.):*Handbook of the Economics of Innovation and Technological Change*, Oxford, Blackwell.

- **Cohen, Wesley M. and Steven Klepper (1996) "A reprise of size and R&D"** ***The Economic Journal*****, 106 (437), 925-951.**

- **Comanor, William S.(1965)** "Research and Technical Change in the Pharmaceutical Industry" *The Review of Economics and Statistics*, 47(2), 182-190.

- **Comanor, William S.(1967)** "Market Structure, Product Differentiation and Industrial Research" *Quarterly Journal of Economics*, 81, 639-51.

- **Cooper, Arnold.C. (1964)** "R&D is More Efficient in Small Companies" *Harvard Business Review*, 42(3), 75-83.

- C**orcos, Gregory, Massimo Del Gatto, Giordano Mion and Gianmarco I.P. Ottaviano (2007)** "Productivity and Firm Selection: Intra- vs International Trade" *CORE Discussion Paper* No. 2007/60.

- **Draho, Jason (2004)** *The IPO Decision: Why And How Companies Go Public*, Edward Elgar Publishing, Northampton, MA

- **Eckbo, Espen B. (2007)** *Handbook of Corporate Finance: Empirical Corporate Finance*, Elsevier, North-Holland.

- **Evans, David S.(1987a)** "The Relationship between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries" *Journal of Industrial Economics*, 35(4), 567-581.

- **Evans, David S.(1987b)** "Tests of Alternative Theories of Firm Growth" *Journal of Political Economy*, 95(4), 657-674.

- **Francesco Caselli (2005)** "Accounting for Cross-country Income Differences" in. P. Aghion and S. Durlauf, Editors, *Handbook of Economic Growth*, 1, Elsevier.

- **Gabe, Todd M. and David S. Kraybill (2002)** "The Effect of State Economic Development Incentives on Employment Growth of Establishments" *Journal of Regional Science*, 42, 703-730.

- **Gibrat, Robert (1931)** *Les Inégalités Économiques*, Librairie du Recueil Sirey, Paris.

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

- **Griliches, Zvi (1992)** "The Search for R&D Spillovers" *Scandinavian Journal of Economics*, 94, 29–47.

- **Griliches, Zvi (1995)** "R&D and Productivity: Econometric Results and Measurement Issues" in P. Stoneman, ed. *Handbook of the Economics of Innovation and Technological Change*, Oxford, Blackwell, 52-89.

- **Guner, Nezih, Gustavo Ventura and Yi Xu (2008)** "Macroeconomic Implications of Size-dependent Policies" *Review of Economic Dynamics*, 11(4), 721-744.

- **Hall, Bronwyn H.(1987)** "The Relationship between Firm Size and Firm Growth in the U.S. Manufacturing Sector," *Journal of Industrial Economics*, 35, 583-606.

- **Hall, Bronwyn H.(1990)** "The Impact of Corporate Restructuring on Industrial Research and Development" *Brooking Papers on Economic Activity*,1 , 85-136.

- **Hall, Bronwyn H., Adam B. Jaffe and Manuel Trajtenberg (2000)** "Market Value and Patent Citations: A First Look" *NBER Working Paper*, No. 7741.

- **Hall, Bronwyn H., Adam B. Jaffe and Manuel Trajtenberg (2001)** "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools" *NBER Working Paper*, No. 8498. Reprinted as Chapter 13 in Jaffe and Trajtenberg (2002).

- **Hall, Robert E. and Charles Jones (1999)** "Why Do Some Countries Produce So Much More Output per Worker Than Others?" *Quarterly Journal of Economics*, 114 (1), 83–116.

- **Hamberg, Daniel (1964)** "Size of Firm, Oligopoly and Research: The Evidence" *Canadian Journal of Economics and Political Science*, 30, 62-75.

- **Hansen, John A. (1992)** "Innovation, Firm Size, and Firm Age" *Small Business Economics*, 4(1), 37-44.

- **Harhoff, Dietmar, Francis Narin, F. M. Scherer and Katrin Vopel (1999)** "Citation Frequency and the Value of Patented Inventions" *The Reviewof Economics and Statistics,* 81(3), 511-515.

- **Horowitz, Irving L. (1962)** " Firm Size and Research Activity" *Southern Economic Journal*, 28, 298-301.

- **Jaffe, Adam B., Manuel Trajtenberg and Rebecca Henderson (1993)** "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations" *The Quarterly Journal of Economics*, 108(3), 577-598.

- **Jaffe, Adam B., Manuel Trajtenberg and Michael S. Fogarty (2000)** "The Meaning of Patent Citations: Report of the NBER/Case Western Reserve Survey of Patentees" *NBER Working Paper,* No. 7631. Reprinted as Chapter 12 in Jaffe and Trajtenberg (2002).

- **Jaffe, Adam B. and Manuel Trajtenberg (2002)** *Patents, Citations and Innovations: A Window on the Knowledge Economy*, Cambridge: MIT Press.

- **Jaffe,Adam B. (2003)** "Real Effects of Academic Research." *The American Economic Review* 79, 957-70.

- **Jones, Charles I. and John C. Williams (1998)** "Measuring the Social Return to R&D" *Quarterly Journal of Economics*, 113(4), 1119-1135.

- **Judd, Kenneth L. (1998)** *Numerical Methods in Economics*, MIT Press, Cambridge MA.

- **Kamien,Morton I. and Nancy Lou Schwartz (1975)** "Market Structure and Innovation: A Survey" *Journal of Economic Literature*, 13(1), 1-37.

- **Kamien,Morton I. and Nancy Lou Schwartz (1982)** *Market Structure and Innovation*, Cambridge University Press, Cambridge.

- **Keklik, Mümtaz (2003)** *Schumpeter Innovation and Growth*, Ashgate Publishing, Hempshire.

- **Klenow, Pete and Andres Rodriguez-Clare (1997)** "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" in B. Bernanke and J. Rotemberg, Editors, *NBER Macro Annual*, MIT Press.

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

- **Klette, Tor J. and Zvi Griliches (2000)** "Empirical Patterns of Firm Growth and R&D Investment: A QuUality LadderModel Interpretation" *NBER Working Paper,* No. 6753.

- **Lanjouw, Jean O., Ariel Pakes and Jonathan Putnam (1998)** 'How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data" *The Journal of Industrial Economics*, 46(4), 405-432.

- **Lanjouw, Jean O. and Mark A. Schankerman (1999)** "The Quality of Ideas: Measuring Innovation with Multiple Indicators" *NBER Working Paper*, No. W7345.

- **Lanjouw, Jean O. and Mark A. Schankerman (2001)** "Characteristics of Patent Litigation: A Window on Competition" *RAND Journal of Economics*, 129-151.

- **Lanjouw, Jean O. and Mark A. Schankerman (1997)**, "Stylized Facts of Patent Litigation: Value, Scope and Ownership" *NBER Working Paper*, No. 6297.

- **Lerner, Joshua (1994)** "The Importance of Patent Scope: An Empirical Analysis" *RAND Journal of Economics*, 25 (2), 319-333.

- **Mansfield, Edwin (1962)** " Entry, Gibrat's Law, Innovation and the Growth of Firms" *American Economic Review*, 52(5), 1023-1051.

- **Mansfield, Edwin (1968)** *Industrial Research and Technological Innovation: an Economic Analysis*, Norton, New York.

- **Mansfield, Edwin, R. John, S. Jerome, W. Samuel and H. Michael (1971)** *Research and Innovation in the Modern Corporation,* W.W. Norton & Company, New York.

- **Meisel, John B. and Steven A. Y. Lin (1983)** "The Impact of Market. Structure on the Firm's Allocation of Resources to. Research and Development" *Quarterly Review of Economics and Business*, 23, 28-43.

- **Melitz, Marc J. and Gianmarco I.P. Ottaviano (2005)** "Market Size, Trade, and Productivity" *Review of Economic Studies*, 75(1), 295-316.
- **Miranda, Mario J. and Paul L. Fackler (2002)** *Applied Computational Economics and Finance MIT Press*, Cambridge MA.
- **Nelson, Richard R. and Sidney G. Winter (1982)** *An Evolutionary Theory of Economic Change,* Cambridge, MA and London: Harvard University Press.
- **Mowery, David C. (1983)** "Industrial Research and Firm Size, Survival, and Growth in American Manufacturing, 1921-1946: An assessment" *Journal of Economic History*, 43, 953-980.
- **Ottaviano, Gianmarco, Takatoshi Tabuchi and Jacques-François Thisse (2002)** "Agglomeration and Trade Revisited" *International Economic Review*, 43, 409-436.
- **Pakes, Ariel, Bronwyn H. Hall and Zvi Griliches (1988)** "The Value of Patents as Indicators of Inventive Activity" *NBER Working Papers*, No W2083.
- **Pakes, Ariel and Margaret Simpson (1989)** "Patent Renewal Data" *NBER Preprints*, No 1265.
- **Pavit, Keith and S Wald (1971)** *Conditions for Success in Technological Innovations*, OECD, Paris.
- **Restuccia, Diego and Richard Rogerson (2007)** "Policy Distortions and Aggregate Productivity with Heterogeneous Plants" *NBER Working Paper*, No. 13018.
- **Sampat, Bhaven N.(1998)** *Cite-Seeing: Patent Citations and Economic Value*, Colombia University, New York.
- **Sampat, Bhaven N. and Arvids A. Ziedonis (2004)** "Patent Citations and the Economic Value of Patents" in Schmoch, Ulrich et al. eds., *The Handbook of Quantitative Science and Technology*, 277-298, Springer.
- **Schankerman, Mark and Jean Lanjouw(2001)** "Characteristics of Patent Litigation: A Window on Competition" *RAND Journal of Economics*, 129-151.
- **Scherer, Frederic M.(1965a)** "Firm Size, Market Structure, Opportunity and the Output of Patented Inventions" *American Economic Review*, 55, 1097-1125.
- **Scherer, Frederic M.(1965b)** "Size of Firm, Oligopoly and Research: A Comment" *Canadian Journal of Economics and Political Science*, 31, 256-266.
- **Scherer, Frederic M.(1980)** *Industrial Market Structure and Economic Performance*, Rand McNally, Canada.
- **Schumpeter, Joseph A. (1934)** *The Theory of Economic Development: An inquiry into profits, capital, credit, interest and the business cycle*, Transaction Publishers, New Jersey.

- **Schumpeter, Joseph A. (1942)** *Capitalism, Socialism and Democracy*, Routledge, London.
- **Shane, Hilary and Mark Klock (1997)** "The Relation Between Patent Citations and Tobin's q in the Semiconductor Industry" *Review of Quantitative Finance and Accounting*, 9(2).
- **Shane, Scott (1999a)** "Selling University Technology" mimeo.
- **Shane, Scott (1999b)** "Technological Opportunities and New Firm Creation" mimeo.
- **Spence, Michael (1984)** "Cost Reduction, Competition, and Industry Performance" *Econometrica*, 52(1), 101-21.
- **Syverson, Chad (2004)** "Market Structure and Productivity: A Concrete Example" *Journal of Political Economy*, 112 (6), 1181-1222.
- **Syverson, Chad (2007) "**Prices, Spatial Competition, and Heterogeneous Producers: An Empirical Test" *Journal of Industrial Economics*, 55(2), 197-222.
- **Trajtenberg, Manuel (1990)** "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *RAND Journal of Economics*, 21, 172-187.
- **Wooldridge, Jeffrey M. (2002)** *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge MA.