A Portrait of Firms that Invest in R&D

by

Lucia Foster
U.S. Census Bureau

Cheryl Grim
U.S. Census Bureau

Nikolas Zolas
U.S. Census Bureau

CES 16-41 September, 2016

The research program of the Center for Economic Studies (CES) produces a wide range of economic analyses to improve the statistical programs of the U.S. Census Bureau. Many of these analyses take the form of CES research papers. The papers have not undergone the review accorded Census Bureau publications and no endorsement should be inferred. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Republication in whole or part must be cleared with the authors.

To obtain information about the series, see www.census.gov/ces or contact J. David Brown, Editor, Discussion Papers, U.S. Census Bureau, Center for Economic Studies 5K034A, 4600 Silver Hill Road, Washington, DC 20233, CES.Papers.List@census.gov. To subscribe to the series, please click here.

Abstract

We focus on the evolution and behavior of firms that invest in research and development (R&D). We build upon the cross-sectional analysis in Foster and Grim (2010) that identified the characteristics of top R&D spending firms and follow up by charting the behavior of these firms over time. Our focus is dynamic in nature as we merge micro-level cross-sectional data from the Survey of Industrial Research and Development (SIRD) and the Business Research & Development and Innovation Survey (BRDIS) with the Longitudinal Business Database (LBD). The result is a panel firm-level data set from 1992 to 2011 that tracks firms' performances as they enter and exit the R&D surveys. Using R&D expenditures to proxy R&D performance, we find the top R&D performing firms in the U.S. across all years to be large, old, multinational enterprises. However, we also find that the composition of R&D performing firms is gradually shifting more towards smaller domestic firms with expenditures being less sensitive to scale effects. We find a high degree of persistence for these firms over time. We chart the history of R&D performing firms and compare them to all firms in the economy and find substantial differences in terms of age, size, firm structure and international activity; these differences persist when looking at future firm outcomes.

[*] Zolas is the corresponding author (Nikolas.J.Zolas@census.gov). Foster, Grim, and Zolas: Center for Economic Studies, Bureau of the Census. We thank Emin Dinlersoz, Rick Hough, Fariha Kamal, William Kerr, John Jankowski, Brian Sliker and participants at the Center for Economic Studies Seminar Series, Bureau of Economic Analysis-Census Research Workshop and 2013 Comparative Analysis of Enterprise Data Conference for helpful comments. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

1. Introduction

Understanding economic growth requires understanding the dynamics of innovation which is difficult given the elusive nature of innovation. The standard definition of "innovation" is "the act or process of introducing new ideas, devices, or methods."[1] Thus many types of expenditures can potentially qualify as investments in innovation, including training, scientific research and development (R&D), marketing, and management and organizational practices.[2] More technically, the "Oslo Manual" focuses on two types of firm-level innovations:

> *"A technological product innovation is the implementation/commercialisation of a product with improved performance characteristics such as to deliver objectively new or improved services to the consumer. A technological process innovation is the implementation/adoption of new or significantly improved production or delivery methods. It may involve changes in equipment, human resources, working methods or a combination of these. (OECD/Eurostat,2005, p.9)"*

It is only relatively recently that surveys have attempted to directly measure innovation. Since our timespan predates these direct attempts, we focus on one commonly used indicator for innovation: R&D activity.[3] Using Census micro-level data, we develop a detailed portrait of the evolution of R&D performing firms in the U.S., charting their existence, growth, expansion into international activity, and death. The goal of this descriptive paper is to provide some understanding of the nature of R&D performing firms and the path that these firms follow to become R&D performers.

Our main variable of interest in this paper is R&D activity, which we proxy by total R&D expenditures. Just as not all R&D expenditures will lead to R&D activity; not all R&D activity leads to innovations. We focus specifically on R&D expenditures rather than on specific innovation outcomes (which are included in later versions of the R&D survey which we use) due to the availability of this variable over time and the continuous nature of the variable (as opposed to the indicator variable which is used to measure innovations), which allows for measures of intensity and strict definition.

We use micro-level data on R&D expenditures to examine firms over a roughly twenty-year period from 1992 to 2011. While R&D data exists prior to this date, 1992 is the first year that the Survey of Industrial Research and Development (SIRD) was substantially revised to better collect R&D in the nonmanufacturing sector. [4] We study all R&D performing firms, but pay special attention to the top 200 R&D performing firms in any

---

[1] http://www.merriam-webster.com/dictionary/innovation, accessed February 10, 2015.

[2] Corrado et al. (2012) provides a discussion of approaches to measuring innovation or "intangible capital" at the country level.

[3] From the Oslo Manual: "One starting point for analysis of innovation activity could be R&D, which takes on a wide variety of functional forms related to problem-solving." (OECD/Eurostat, 2005, p.26)

[4] Foster and Grim (2010) document the availability of R&D data going back to the 1950's, however it was mainly focused on the manufacturing sector. Prior to 1992, there exists data on nonmanufacturing sectors, but only made up a relatively small portion of the survey.

given year. We focus on the top 200 because while the SIRD and BRDIS sample frame has evolved substantially over time, large R&D performers have consistently been included in the survey since its inception in 1953. Thus, this paper serves as dynamic companion to the cross-sectional descriptive paper by Foster and Grim (2010) which provides a detailed picture of the top 200 R&D performing firms in the year 2003.

Researchers have measured innovative activity in numerous ways, including company R&D expenditures, counts of innovations from trade journals, patent counts, and patent citations.[5] Each of these measures has strengths and weaknesses and some are simply not available for all of the firms in our sample. While none of these measures provides a perfect proxy for innovation, we expect that they are positively correlated and together provide a picture of innovation. For example, Shackelford (2013) finds that one in five U.S. companies with research and development activities applied for a patent from the U.S. Patent and Trademark Office in 2008.

This paper builds upon the work of Griliches (1980) and Bound et al. (1984) who also examine the determinants of R&D activity using the SIRD and Hirschey et al. (2012) who look at the firm characteristics for corporate R&D spending over time. Hirschey et al. (2012) is the largest study in the recent literature and uses Compustat data for publicly-traded companies. One drawback of these studies is the focus on the subset of firms known to conduct R&D (i.e., they suffer some selection bias). Our sample, when weighted, is representative of the U.S. economy where the majority of firms conduct little to no R&D.

Our work expands upon this existing literature by including a focus on the link between R&D activity and international trade and thus relates to Aw, Roberts, and Xu (2008, 2011). We hope to examine the deeper relationship between trade and R&D in future work. Analyzing the international activity of firms is helpful in understanding the types of firms that engage in R&D, as trade and multinational investment are methods for diversifying revenue, creating new markets, reducing costs and enhancing productivity, much as R&D is intended to do.

We find many of the patterns seen in the cross-sectional analysis of 2003 in Foster and Grim (2010) hold for our period of 1992-2011: top R&D firms are more likely to be large, old firms in the manufacturing sector. We also find that these same top-performing firms are largely multinationals who participate heavily in international trade and compete in less concentrated industries. In addition, we find that R&D performing firms in general have higher existence rates, are much more likely to be multi-unit and are much more likely to engage in international activity, specifically exporting. These patterns hold consistently over time as well, although a greater number of small, single-unit firms are beginning to participate in R&D activities. Thus we find the overall share of innovations by large firms is steadily declining (consistent with Block and Keller (2009)). In terms of future outcomes, R&D performers are more likely to exist and trade 10 years into the future.

The remainder of the paper is organized as follows. Section 2 describes the data. Firm characteristics of R&D performers are described in Section 3. The main results regarding

[5] As a robustness check, we look at company R&D (total R&D less federal R&D) as a measure of R&D activity and find only minor changes in the magnitude of our results.

the evolution and outcomes of top R&D performing firms are presented in Section 4. Section 5 provides conclusions and some potential future extensions to our research.

2. Data and Measurement

Our empirical exercises rely upon linking four micro-level datasets housed at the Census Bureau. We first describe the datasets that provide information on R&D activity and then the datasets that provide information on firm characteristics including trade. As with all Census restricted-use data, results reported from empirical exercises using these data must pass strict disclosure avoidance review; therefore, all identifying information is suppressed.

2.1. Research and Development Data

Our measure of total R&D expenditures comes from the Survey of Industrial Research and Development (SIRD) and its successor the Business Research and Development and Innovation Survey (BRDIS). These surveys are sponsored by the National Science Foundation (NSF) and collected by the Census Bureau. The SIRD was transformed into the BRDIS in 2008 in order to capture changes in how R&D was being performed. The largest changes to research operations cited by the NSF include the transition from a manufacturing-based economy to service-based one, the increased dispersal of R&D from former centralized labs, the rise of privately funded R&D, and the global focus of R&D. [6]

Many of these changes are shown in Figure 1, which highlights the growth of privately-funded R&D (panel a), increasing share of development expenditures in private domestic R&D (panel b), shift towards foreign R&D (panel c), and increased dispersion in terms of the firm composition of R&D performers where smaller sized firms make up a larger fraction of total private R&D expenditures (panel d). As a result, the transformation to the BRDIS included important increases in the scope of content and changes in some measurement methodologies. Hereafter, we refer to these surveys jointly as the BRDISX.

The purpose of the BRDISX is to provide "information on research and development performed or funded by businesses within the United States."[7] The BRDISX collects firm-level annual data on R&D expenditures at total levels by type (basic research, applied research, and development), source of funding (Federal R&D funds versus company R&D funds), industry, and the number of scientists and engineers.[8] There are 25,000-40,000 companies on average in the survey in each year.[9] We focus on the period 1992-2011 where the longitudinal linkages are of higher quality and the survey sample is

[6] The BRDIS website provides the reasons for developing a new R&D survey. See http://www.nsf.gov/statistics/srvyindustry/about/brdis/whybrdis.cfm (accessed October 4, 2015).

[7] For more information on the BRDIS see http://www.nsf.gov/statistics/srvyindustry/ (accessed October 4, 2015).

[8] The BRDISX also collects data on items such as net sales, total domestic employment, distribution of R&D by state, wages and salaries of R&D personnel, costs of materials and supplies consumed, and depreciation on R&D property and equipment.

[9] Using sample weights, the survey size is roughly 1-2 million in a year. The survey size expanded with the BRDIS, sampling more than 40,000 companies on average per year. Follow-up on non-responders was altered after 2001 so that while the total number of unweighted respondents stayed consistent, the total number of weighted respondents more closely matched the total number of weighted firms surveyed.

larger and has a more consistent scope over time.[10]

Identifying firms that do not have any R&D expenditures is key to our analysis. Within the BRDISX, we distinguish between all of the firms ("R&D sample") and those firms that report positive total R&D funds ("R&D performers"). About 20 percent of the firms in the sample report positive total R&D expenditures for 1992-2007 (SIRD collection).

Of the non-respondents, the NSF and Census Bureau follows up more closely with previously known R&D performers, so there will be a slight bias towards R&D performers in terms of observable data. However, we do not make any assumptions regarding missing data or non-respondents and simply drop them from our analysis.

Figure 2 shows the level of total R&D expenditures in constant dollars as measured by the BRDISX from 1992 to 2011 for the published data (solid line) and the underlying available microdata (dashed line). The microdata come extremely close to replicating the published totals in nearly all years, but there are some years where there is a slight divergence, namely in the first year of the BRDIS. We believe differences between the published data and the microdata reflect revisions to the BRDISX data. These revisions were applied to the published data but not carried back to the original micro-level source data. We do not incorporate these revisions into the micro-level data.

The BRDISX is firm-level, rather than establishment-level, data which introduces complications due to mergers and acquisitions. We attempt to mitigate the effect of these complications through our use of the Longitudinal Business Database (described below).

2.2. Firm-level Characteristics Data

We use microdata from the Longitudinal Business Database (LBD) to follow firms over time and construct firm characteristics such as firm age, existence, and death. The LBD links establishments found in the Census Bureau's Business Register (BR) over time, starting with the year 1976 (Jarmin and Miranda, 2002). The BR contains nearly all non-farm private business establishments in the U.S. economy and includes information on employment, payroll, industry, and location.[11] The LBD links establishments in the BR starting with the year 1976 with a focus on creating high quality links over time. Using the LBD, we create a measure of firm age (left-censored in 1976), an indicator for whether the firms are single unit or multi-unit, as well as the nature of their underlying establishments.

We use the LBD to generate measures of firm size, firm age, firm structure, and industry competitiveness. Matching to the LBD allows us to see the full life of the firm from 1976 to 2011, not just years in which the firm is in the BRDISX. For example, if a firm found in the BRDISX from 1993 until 1997 is found to exist prior to 1993 and after 1997, those

[10] The surveyunderwent significant changes in the sample, scope, and treatment of zero values, over earlier periods which can complicate comparability over time (see Foster and Grim (2010) for a detailed explanation).

[11] The primary source of information on the BR comes from the Business Master File and the Business Income Tax and Payroll Tax forms from the Internal Revenue Service. Information from the Census Bureau's Economic Census and Report of Organization Survey are also included on the BR. Additional information concerning industry and location comes from the Bureau of Labor Statistics and the Social Security Administration.

new firm-year observations are included in our dataset in order to generate a full panel by firm-year.

In keeping with much of the empirical literature using Census microdata, we measure firm size by the number of employees at the firm.[12] Following Jarmin, Haltiwanger, and Miranda (2013), we define firm age as the age of the oldest establishment associated with the firm in the first year the firm identifier appears in the LBD.[13] The LBD also provides us with a measure of firm structure – whether the firm is a single establishment firm or a multi-unit firm. We create an indicator variable for multi-unit firms and then create a multi-unit intensity variable by calculating the number of establishments for each multi-unit firm.

We generate a measure of industry competitiveness using the LBD. We calculate a Herfindahl-Hirschman index (HHI) measure of competitiveness for each industry using the annual receipts data collected from the BR.[14] We first calculate the share of sales of the firm relative to the total sales of all firms classified by the lead 4-digit NAICS code if after 1997 or 4-digit SIC code if prior to 1997.[15] We then sum up the squares of this value for all firms within each industry code. Firms in less competitive industries will have a higher HHI.

Finally, we use microdata from the Longitudinal Firm Trade Transactions Database (LFTTD). Linking to the LFTTD allows us to examine the international activity of R&D performing firms. The LFTTD contains transaction-level data on firm imports and exports from 1992 to 2011. For each transaction, the LFTTD contains information on the product class (Harmonized System (HS) industry code), value, quantity shipped, shipment date, destination/origin, and whether the transaction was conducted between related parties.[16] We create a measure of the volume of trade the firm engages in by aggregating the transaction-level import and export data to the firm level. We also use the LFTTD to determine if a firm is a multinational: if a firm has at least one "related"

---

[12] Researchers have also measured size using output or sales. Griliches (1980) uses value-added output. Acs and Audretsch (1987) and Akcigit and Kerr (2010) use sales as the denominators when calculating their innovation rates. Akcigit and Kerr (2010) also categorize firms by the number of their establishments.

[13] Establishment age is defined as the first appearance of an establishment in the LBD with positive employment. We require positive employment because we measure firm size by employment. Note because employment is defined in the LBD as of March 12th, it is possible for an establishment to have zero employment and positive payroll in a given year.

[14] Annual receipts data are collected in the Business Register. We clean the data by using the prior year's receipts from the "year+1" observation when available since revenue and receipt data often get corrected in the following year. We also discard observations with negative receipts, receipts per employee of less than $100 and more than $20,000,000, and the bottom and top 0.1% of receipts per employee. In addition, we drop observations from firms whose receipts increased by more than 100 times in a single year. Also, as a robustness check, we conduct the same analysis using HHI based on number of employees and payroll of the firm and get very similar results. The raw correlations between HHI (Sales) and HHI (Employment) and HHI (Payroll) are 0.74 and 0.69 respectively.

[15] The "lead" industry code for multi-unit firms was taken as the industry code for the largest establishment payroll within the multi-unit firm.

[16] Exporting parties are related when either owns 10 percent or more of the other party, whereas for imports, the threshold for related party is 5 percent under United States Code 19, §1401a(g). However, note that 19 U.S.C. §1401a(g) outlines seven different ways in which parties may be related in a U.S. import transaction, of which only one is ownership based.

party transaction in a given year we classify the firm as a multinational for that year.[17] In addition to a multinational indicator variable, we construct an intensive margin measure for multinational firms by the number of different countries the firm conducts related party transactions.

As a final step, we link all of the datasets described above.[18] Once this process is complete, we have every possible firm-year combination for a firm in the BRDISX. The match rate between the BRDISX and the LBD is quite high – over 99 percent for our combined set of BRDISX firm-year observations, with only slight variation by year. Finally, we match the combined R&D and LBD data to the LFTTD by firm identifier for 1992, the first year LFTTD data is available, to 2011. Approximately 10-15 percent of LBD firms in any given year conduct at least one international transaction.

3. Who are R&D Performing Firms?

In this section, we examine some of the distinguishing features of R&D performing firms and the relationship between R&D performance and firm characteristics. The firm characteristics are: industry, size, age, structure (multi-unit and multinational status), and trade activity. We start with some descriptive statistics before examining these characteristics in a regression framework.

We compare five groups of firms in our descriptive discussion: (1) top performing R&D firms, (2) all R&D performing firms, (3) all firms in the economy, (4) BRDISX respondents, and (5) all multi-unit firms (the last two partly control for sample design issues).[19] We define our top performing R&D firms as the 200 firms with the most R&D expenditures in any given year. As Figure 3 demonstrates, Top200 firms make up a significant share of overall R&D expenditures relative to the total economy (but this share has decreased over time from roughly three quarters to two-thirds).

Summary statistics for our five groups of firms are shown in Table 1. The sample is reasonably representative of the whole economy for the characteristics we examine. One difference is that respondents appear to be slightly more likely to be multi-unit firms and multi-nationals. Therefore, we include the group of multi-unit firms for the whole economy as another group for comparison. BRDISX respondent firms are also slightly younger than firms in the economy as a whole.

R&D performers are quite different from other firms in the economy and the Top200 tend to show even more pronounced differences. Generally speaking, R&D performers and Top200 firms are more likely to be larger, multi-unit multinationals who trade and operate in less competitive industries. Of R&D performing firms, 21 percent are multi-unit firms, while only 7 percent of firms in the whole economy are multi-units.

---

[17] A similar definition of multinationals is used in Bernard et al. (2009).

[18] We start by stacking the SIRD and the BRDIS firm-level datasets for 1972 to 2011. These data provide information on R&D expenditures, number of scientists and engineers, and firm-level industry code. Next, we match the stacked R&D data to the LBD. We first aggregate the establishment-level LBD to the firm level using available firm identifiers. We then use the firm identifiers to link the BRDISX data to the LBD.

[19] R&D performers are firms reporting positive total R&D expenditures. All statistics for R&D performers in this paper are weighted using BRDISX sample weights. Whole economy data are limited to firms with at least five employees to be consistent with the BRDISX sample frames.

Top200firms are almost exclusively multi-units.

Mean employment at R&D performing firms is eight times the mean employment at firms in the whole economy and nearly 14 percent larger than mean employment in multi-unit firms. Median employment is more than 12,500 employees per Top200 firm, while the median R&D performing firm has 26 employees.[20] Firms with more than ten thousand employees comprise less than 1 percent of all firms reporting R&D, yet these firms contribute more than 50 percent of total R&D expenditures (see Figure 1d). Figure 4 shows a non-parametric representation of the relationship between size (proxied by sales) and R&D expenditures (log sales on the x-axis and log R&D on the y-axis) with a size elasticity of approximately 0.563, indicating that a 1 percent increase in size corresponds to a 0.56 percent increase in R&D expenditures. This elasticity has been steadily declining over time from 0.64 in 1992/93 to 0.50 in 2010/11, suggesting that size is becoming less important a determinant of R&D expenditures. This elasticity also differs significantly from the almost unit elasticity found in Bound et al. (1984) who looked at publicly-traded firms in Compustat, which tend to be larger and may have indadvertantly led to the higher elasticity. Further, although R&D performers represent fewer than 2 percent of the total number of firms in our sample, their employment share is approximately 15 percent.

R&D performers are younger than multi-unit firms but Top200 are older than multi-unit firms (and R&D performers). The third panel of Table 1 shows the median age of multi-units is 18, 11 for R&D performers, and 22 for Top200 firms. Figure 5 highlights the allocation of firms by birth cohort (recall the LBD starts in 1976), where the birth cohort is broken up in 5-year increments. The birth year of R&D performing firms does not differ greatly from whole economy firms and non R&D performing firms. However, when we turn to the Top200 firms, the age composition differs markedly, with older firms comprising the majority of Top200 firms in all years.

Another distinguishing characteristic of R&D performers is their substantial amount of trade (imports plus exports) and their multinational status. In any given year, one out of ten firms that perform R&D are multinationals and a large majority of Top200 firms are multinationals (83 percent). These multinational R&D performing firms appear to be substantively different from the typical R&D performing firm in the BRDISX sample. For one, they are much larger than average. The average multinational R&D performing firm has 5,217 employees (versus the 2,297 employees found in the average R&D firm and the 908 employees found in the average non-multinational R&D firm) and conducts nearly 2.5 times as much R&D as the average R&D performing firm and more than 8 times as much R&D as non-multinational firms ($69M versus $28M versus $8M). Firms in our sample who are multinationals are more likely to have positive R&D activity. Of the firms in our sample, more than 77 percent of multinational firms have positive R&D. These figures imply there is potentially a significant relationship between R&D activity and multinational activity, which is one of the key findings in Slaughter (2009).

Trade also appears to be linked with R&D as the cumulative trade (imports plus exports) of R&D performers accounts for nearly 60 percent of U.S. trade, and the Top200 account

[20] To ensure no confidential information is disclosed, reported medians are "fuzzy" medians, where the "fuzzy" median is the mean of all observations between the 45th and 55th percentiles.

for nearly a third of all U.S. trade on average in any given year. Besides importing and exporting at higher levels, R&D performers are also more likely to trade in general. In terms of exports, this is consistent with the results in Aw, Roberts, and Xu (2011) albeit for an entirely different sample (Taiwanese electronic manufacturing plants). On average, between 50-60 percent of all R&D performers executed at least one international transaction, compared with 7-8 percent of all SIRD respondents. The R&D performers who had at least one international trade transaction also appear to be different from the average R&D firm. They perform nearly 1.5 times as much R&D as the average R&D firm, and more than 5 times as much R&D as firms that do not trade ($36M versus $6.5M). They are also on average 9 times larger than R&D performing firms that do not conduct trade.

Finally, the last panel of Table 1 shows R&D performers tend to be in less competitive (more concentrated) industries than firms as a whole based on the higher HHI level. Top200 firms are in even less competitive industries than R&D performers.

Table 2 highlights additional differences between the Top200 and R&D performers. The Top200 account for on average approximately 65 percent of total R&D expenditures. These firms are also slightly more R&D intensive. Further, Top200 firms make up approximately one third of each of total domestic sales and total domestic employment. Top200 firms employ more than half of all scientists and engineers within R&D performing firms.

3.1. The Relationship Between R&D Performance and Firm Characteristics

We now use regression analysis to examine the relationship between R&D expenditures and firm characteristics for R&D performers. We focus on both R&D expenditure levels and intensity (as measured by R&D-to-sales ratio). To motivate the structure and components of our estimating equation, we incorporate some preliminary findings (such as the log-log relationship between firm size and expenditures) and a discussion of the explanatory variables, which mainly include the measurable components drawn from Census microdata. Our main estimating equation is of the form:

$$\log RD_{it} = \beta_1 \log SIZE_{it} + \sum \delta AGE_{ik} + \beta_2 \log TRADE_{it} + \gamma MU_{it} + \mu MNC_{it} + \beta_3 \log HHI_{jt} + \theta_j + \varepsilon$$

Where RD_{it} is firm R&D expenditures or R&D-to-sales ratio, $SIZE_{it}$ is firm employment, AGE_{ik} are various firm-age dummies, $TRADE_{it}$ is total trade dummy/intensity by firm (imports and exports are broken out separately), MU_{it} is a dummy/intensity measure for multi-unit firms, MNC_{it} is a dummy/intensity measure for multinational firms, HHI_{jt} is the Herfindahl-Hirschman index measure of industry *j*'s competitiveness by firm revenues, and θ_j are industry effects. Unfortunately, industry is not collected on the

SIRD, but is instead derived from establishment-level data rolled up to the firm level using payroll as the activity variable. Thus, all of the R&D expenditures in a firm are assigned to a single industry potentially masking diversity in R&D activity. We keep this caveat in mind in all of our analysis using industry. All regression results are weighted by the BRDISX sample weight and are run on the 1992-2011 data only.

Our explanatory variables are chosen based on a combination of data availability and preexisting studies. Due to the panel structure of the data, it is important to have consistent variables throughout the time period (1992-2011) in any given year for all firms. Additional details found in other Census microdata sets, such as the Economic Census could not be used due to their limited availability (once every five years) and differences across sectors (R&D firms can include establishments in manufacturing and non-manufacturing sectors). In addition, annual surveys, such as the Annual Survey of Manufactures only encompass a subset of establishments of R&D performing firms and would introduce additional selection issues into the study.

Nevertheless, each of the included variables have been linked with R&D performance in the literature. Firm size has been among the most studied determinants of R&D, being linked with innovation starting with Schumpeter (1942). Follow-up studies by Griliches (1980) and Bound et al. (1984) continued to document the positive relationship between firm size and R&D expenditures. More recently, Akcigit and Kerr (2010), use the same underlying datasets used in this study and find that R&D expenditures rise, while R&D intensity falls as firms get bigger. Looking at the interaction between firm size and market structure, Acs and Audretsch (1987, 1988) find that large firms have an innovation advantage in relatively concentrated markets, while small firms have an innovation advantage when markets are more competitive. Our study uses employment as a measure of size due to a combination of greater availability and higher validity. Revenue measures compiled by firms are subject to greater change on a year-to-year basis.

Firm age has also been linked with innovation and R&D expenditures. Akcigit and Kerr's (2010) model predicts that younger firms are more innovative than older firms. Firm structure also plays an important role as greater diversification in revenue streams allows firms to engage in more R&D activity due to the ability to spread risk more evenly (Doi, 1985). This diversification can take the form of additional plants and production facilities which can produce more/different items and/or additional markets found abroad. Finally, industry competitiveness plays an important role as well. Firms that have greater monopolistic power have greater incentives to innovate because they are able to better appropriate the returns from their R&D investment (Gilbert and Newbery, 1982; Blundell et al., 1999). On the other hand, in highly competitive industries, R&D may equate to the survival of the firm so that more firms engage in R&D activity. Aghion et al. (2005) hypothesize that the relationship between competition and innovation is an inverted U-shape. In their model, competition has the effect of simultaneously increasing the profit that comes from innovation while also reducing the incentives for innovation.

There are other factors associated with R&D performance in the literature, but not included in this study. R&D persistence has been noted extensively in the literature (Garcia-Quevado et al., 2011; Guellec and de la Potterie, 2003; Myriam and Etienne,

2007) as a key determinant for R&D expenditures[21]. Our analysis focuses predominantly on the across-firm determinants of R&D expenditures, so we exclude prior R&D expenditures due to selection issues where non-performers are dropped from the survey after 1-2 years. Other factors that we would like to consider if we had the data include firm ownership – public versus private and institutional (Bushee (1998) and Aghion et al. (2013)), management and previous R&D experiences (Garcia-Quevado et al. (2011), Guellec and de la Potterie, (2003) and Myriam and Michel (2007)), human capital and skill diversification (Kaiser (2002)), and productivity (Aw, Roberts, and Xu (2008, 2011)).

Since most firms in the sample do not conduct R&D, it is important to account for the selection bias of excluding non-R&D performers and the sampling methodology in our estimation. As mentioned in the previous section, the sampling frame of the BRDISX has undergone extensive revisions over time, with special emphasis on known R&D performers (for further discussion on these sampling issues, see Foster and Grim (2010)). Hence, we run a Heckman selection model, which corrects for this bias by including a Heckman correction term (i.e., inverse Mill's ratio) into the regression. In the first stage of the regression ("selection"), we use dummy variables for trade status (importing or exporting status), multi-unit status and multinational status, while including industry and year fixed effects. In the second stage, after accounting for the selection bias, we include trade intensity (imports and exports), multi-unit intensity (number of establishments per firm) and multinational intensity (number of countries of related party trade transactions) in order to assess whether increased diversification impacts the levels of R&D expenditures.[22]

Regression results on R&D expenditure levels for this specification are shown in Columns 1, 2 and 3 in Table 3. The initial selection stage (Column 1) reveals that size, as expected, is positively associated with whether or not a firm performs R&D. Age is negatively associated with whether or not a firm performs R&D. As firms become older, they are increasingly less likely to perform R&D and their expenditures tend to be lower

[21] We ran an additional specification including lagged measures of R&D and whether previous R&D experience and levels influenced current R&D performance using the Arellano-Bond dynamic panel data model. The results suggest that previous R&D experience is a strongly positive and significant determinant for whether a firm conducts R&D in the current period. Including this regressor significantly reduces the sample size (due to non-performers being dropped out of the survey) and reduces the magnitudes of the coefficients. However, the signs and significance of the coefficients mostly stays the same. Prior R&D spending has a near unit elasticity (0.870).

[22] In addition to the selection bias, there may be issues with bias in our log-linear specification due to heteroskedasticity of the sample data. With a log-linear specification, our results may actually underpredict the real effect of these variables (see Haworth and Vincent (1979) and Santos-Silva and Tenreyro (2006) for discussion) due to the concavity of the log function. This disadvantage also applies to the two-stage Heckman selection model which relies heavily on homoscedastic data. Therefore, as an alternative specification, we also run the Poisson psuedo-maximum likelihood estimation (PPML), which is consistent with heteroskedastic data and also has the benefit of being able to handle zeroes in the dependent variable. However, this does not alter the results. Unfortunately, we have to exclude industry fixed-effects due to non-convergence, which is a common problem with PPML regressions when using a large number of fixed effects. We also run negative binomial and zero-inflated Poisson (ZIP) regressions in order to account for the over-dispersion in the data and excess zeroes. The results are qualitatively similar. However, due to the difficulty of interpretation of the coefficients, we elect to stick with the Heckman selection model as our primary estimation model.

than young firms. Firm structure plays an important role in determining which firms conduct R&D as firms who trade, multinational firms and multi-unit firms are all more likely to perform R&D. Interestingly, the exporting coefficient is nearly twice as large as the importing coefficient, highlighting that exporting firms are a stronger predictor of R&D performing firms. Finally, industries with higher concentration also lead to a greater proportion of R&D performers.

After controlling for the selection, we compare the "corrected" specification (3) with the uncorrected specification (2) and find only minor differences. The corrected specification tends to magnify the coefficients slightly. We find that size is a key determinant of R&D expenditure levels in both the corrected and uncorrected specification. Note, we run a similar specification using "Sales" as the proxy for the size (not shown here) and get mostly similar results with a slight difference in the magnitude of the coefficient. Age has a significantly negative impact on R&D expenditures, suggesting that as firms get older, they perform less R&D. This may be due to the fact that size and age are correlated (correlation of 0.152). However, running the corrected specification using only the age dummies and the industry fixed effects gives positive, significant results for the oldest firms (aged 25+) and insignificant effects for all other groups, suggesting that even without size and other international factors, age does not play a major role in determining R&D expenditures.

On the other hand, international economic activity has a positive and significant effect on R&D expenditures, with both import and export levels leading to higher levels of R&D expenditures. The coefficient for exports is more than twice the size the coefficient for imports, suggesting that risk diversification may be an important incentive for R&D since exports and R&D performance appear to be highly associated. In addition, multinational activity also leads to higher levels of overall expenditures. On the other hand, having a greater number of establishments and plants is associated with lower levels of R&D. This may be due to the fact that higher decentralization leads to increased costs of R&D. Finally, we find that industry concentration has virtually no impact on R&D levels.

Turning now to the R&D intensity specification, we find that many of the signs and patterns still hold, with the exception of size. R&D intensity remains very persistent, but size has no impact on R&D intensity. We find that the oldest set of firms continues to have a negative association with R&D activity. We also find that international activity such as exporting and multinational status have a significant and positive impact on R&D intensity. Other factors, such as importing, multi-unit intensity and industry competitiveness have no impact on R&D intensity.

Embedded within the pooled results are also dynamic changes happening to the covariates that may hint at potential changes to the R&D process by firm types over this time period. Table 4 breaks up the panel used in Table 3 into three separate five-year time periods and runs a similar specification paying special attention to how the values of the coefficients change throughout. Many of the differences we find across cohorts are mostly modest, but we do find lower responsiveness in the levels of R&D that are driven by scale effects (employment), and international engagement (trade and multinational status). Table 4 shows a small but noticeable downward shift in the elasticity over the time period as the elasticity declines from 0.587 in 1995-2000 period to 0.531 in the 2005-2010 period.

Overall, the results thus far largely confirm earlier studies on R&D performance, which have been done with much smaller sample sizes and exclude non-R&D performers. We find that large, multinational firms have higher overall levels of R&D expenditures and are more likely to conduct R&D than smaller domestic firms. However, we find that size is also associated with lower levels of R&D intensity. It is possible the results are influenced by the inclusion of retail and wholesale firms, which may have large sale measures with minimal R&D. However, we run the same specification on a subset of manufacturing firms and find the results are very similar. [23] Other possibilities include the fact that we are measuring domestic R&D expenditures, which may miss large amounts of R&D spending for multinationals and the fact that larger firms are more likely to perform R&D via acquisition than in-house R&D.

Having identified a number of firm characteristics associated with R&D expenditures we now turn to the evolution of these firms and how these firms differ from the national sample of firms.

4. How Have R&D Performing Firms Changed over Time?

We next examine the history and evolution of R&D performing firms, paying special attention to the top 200 R&D performing firms in each year ("Top200").[24] Foster and Grim (2010) characterizes the Top200 firms for a specific year (2003) and charts their evolution over time. Their findings suggest that the majority of Top200 firms existed since the beginning of the LBD (1976) and that a significant share of these firms were large R&D producers from the start of that time period. Post-1976 entrants that ended up as a Top200 firm in 2003 grew at an extraordinary pace. Here, we expand their analysis of the evolution of Top200 firms to all R&D performing firms and track the evolution of each of these firms over time.

4.1. R&D Performing Firms

Figure 6 shows the existence of firms in the whole economy (LBD firms with 5 or more employees) and R&D performing firms for 1992 to 2011. Each line represents the x-axis year (*t*) minus the 5, 10, 15, 20, 25 outcomes. For example, the year 2000 values of the *t*-5 line in the top panel of Figure 6 shows the percent of R&D performing firms in 2000 that existed in 1995. Note the *t*-20 and *t*-25 lines are left-censored since the LBD starts in 1976.

R&D performing firms are persistent relative to firms in the whole economy– they are significantly more likely to have existed in prior years. Approximately 90 percent of R&D performing firms exist 5 years earlier, while only roughly between 45 and 60 percent of firms in the whole economy exist 5 years prior. The same pattern holds going back to existence 25 years prior. Roughly 30 percent of R&D performing firms exist 25 years earlier while less than 20 percent of firms in the whole economy exist 25 years earlier. It is also illuminating to note the increase in the percent of firms in existence in

[23] We designate whether a firm is a manufacturing firm by the firm having at least one establishment conducting a manufacturing activity (NAICS 31-33).

[24] We also looked at the top firms who make up 50, 60 and 75 percent of total R&D expenditures and get qualitatively similar results. However, we stay with the top 200 definition for consistency with Foster and Grim (2010).

prior years over time in the whole economy (top panel of Figure 6). This implies firms in the whole economy are older on average in 2011 than they were in 1992. Interestingly, this pattern does not hold for R&D performing firms, as the proportion of firms in existence over each 5-year period stays more or less constant and actually declines a bit. Due to changes in the BRDISX sample over time, it is difficult to say anything concrete about the time series pattern seen in the R&D performing firms panel. Changes over time in that panel may reflect changes in the composition of the sample.

In addition to looking at the backward history of R&D performing firms relative to firms in the whole economy, we can also compare outcomes and their future performance/characteristics. We find similar patterns of existence/survival in the future for firms in the whole economy where approximately 35 percent of firms in any given year exist in *t*+10 years. We see much higher rates of future existence among R&D performers, with approximately 80 percent of R&D performers are still in existence 10 years out.

Turning now to the likelihood of conducting international activity, we map both the backward and forward-looking proportion of firms who trade or are multinationals (as defined through related-party transactions). For whole economy firms, we have between 7-10% of all firms that either import or export in the years prior or future, with that figure staying constant throughout the time period (1992-2011). Similarly, we have roughly 2-3% of the whole economy firms being multinationals throughout this time period and the years going forwards and backwards. By comparison, Figure 7 shows that between 65-80% of R&D performing firms will/have imported or exported in the years prior and future, with this figure staying mostly constant throughout the time-period. What is interesting however in Figure 7a is the reduced likelihood of R&D performing firms in the latter years to trade (compare the shift in the lines between 1995 R&D performing firms to 2010 R&D performing firms). This would again support the notion that smaller, domestic-only firms are becoming more involved in R&D. Also, of importance in this figure is the upward shift in the proportion of 1995 and 2000 R&D performing firms that trade in the future occurring in the early 2000's, which coincided with China's entry into the World Trade Organization.

Figure 7b looks at the proportion of R&D performing firms who are multinationals. We find that compared to the whole economy firms, R&D performing firms are between 10 and 20 times as likely to be multinationals. We also see a relatively steep decline in the proportion of multinationals occurring prior to 1995, which may potentially hint at some sampling biases as the sample was expanded. After 1995, we find that the proportion of multinational firms by cohort stays relatively constant in the years before and after. We also find a similar downward shift in the lines occurring across cohorts, suggesting that the latter years contain a higher proportion of domestic, non-multinational firms.

To conclude, the backward and forward-looking figures on international activity suggest that R&D performing firms are considerably more likely to either import or export or be multinationals than other firms in the whole economy. We see that in the latter years, fewer R&D performing firms trade or are multinationals. We also find that international activity seems to be unrelated to R&D since the rates of international activity stay mostly constant in the years before and after a firm conducts R&D. We now turn to the backward and forward-looking attributes for Top200 firms.

4.2. Top200 Firms

As first highlighted in Foster and Grim (2010), Top200 R&D performing firms differ substantially from other firms in that they tend to be significantly larger, older and persistent. In this section, we explore that persistence and look at the Top200 R&D performing firms in all of the years between 1992 and 2011 and their evolution before and after. We start by focusing on existence and survival.

In Figure 8, we find that more than 60 percent of the Top200 firms in 2001 exist 25 years ago and more than 95 percent of Top200 firms exist 5 years ago, both of which are significantly higher than the survival rates of whole economy firms and R&D performing firms. In addition, a multi-unit graph for Top200 (not shown) shows nearly identical rates across all years where more than 95 percent of Top200 firms that exist in any of the previous years are multi-unit firms as well. Similar rates of activity hold for Trade (figure not shown, but is consistently above 90%) and multinational status (figure not shown, but is consistently high). Going forward, we find similarly high survival rates of Top200 R&D performing firms. While these differ significantly from whole economy firms, the forward-looking survival rates do not differ much from the forward-looking survival rates of R&D performing firms.

In addition to increased persistence of survival and international activity, we also find that the share of Top200 firms that are in the Top200 in earlier years is remarkably high. Between 30 and 35 percent of Top200 firms were also among the Top200 twenty-five years ago and more than half of the Top200 firms are in the Top200 ten years ago. Looking forward, we find similar rates of persistence with more than half of the Top200 firms remaining in the Top200 ten years into the future.

5. Conclusions and Future Research

We use firm-level microdata to create a detailed portrait of R&D performing firms in the U.S. We look at the relationship between R&D performance and a set of measurable firm characteristics and then examine the history and evolution of R&D performing firms. Critical to our analysis is the ability to combine survey data on R&D expenditures with data from the LBD. Linking the BRDISX data to the LBD enables us to examine important dynamics of R&D performing firms that are not possible to examine using a survey with a rotating panel alone. The use of the LBD also allows us to compare R&D performing firms to firms in the economy as a whole. Incorporating the LFTTD panel in addition to the LBD allows us to look at the relationship between R&D performance and multinational status and trade volume.

We find top R&D performing firms are generally large, multi-unit, multinational old firms that conduct trade transactions. Many of the top R&D performing firms are also among the top performing firms in years past and continue to stay on top into the future. We also find evidence that more smaller domestic-only firms are becoming engaged with R&D with R&D expenditures being less sensitive to scale effects over time. This may be

the result of sampling issues with the survey, but the results hold even after controlling for selection.

While we look at many interesting characteristics of the top R&D performing firms in this paper, many areas remain for future work. In terms of the evolution of these firms, we could investigate the roles of merger and acquisition activity and the transformation from a non-employer firm to an employer firm. For the latter, we could examine the transition using either the non-employer universe or perhaps the linked employer-employee data from the Longitudinal Employer-Household Dynamics program (to track employees as they move to different firms and perhaps start up their own new firms). It would be interesting to take the flip side of our evolution analysis and examine how many of the fastest growing firms in the LBD had R&D expenditures. This work would complement the work in Akcigit and Kerr (2010). We are also interested in a more detailed study of how the survival rate of R&D performing firms differs from other non-performing firms after including a variety of firm and industry-level controls similar to Fernandes and Paunov (2012).

Finally, as noted earlier, we would like to enhance our analysis to include a measure of productivity. Clearly not all R&D expenditures result in innovations and not all innovations result in productivity growth. We could measure the productivity growth at top R&D performing firms and compare to the productivity growth of other firms, similar to Acemoglu et al. (2013). We could also look at the relationship between R&D expenditures, trade, and productivity as in Aw, Roberts, and Xu (2011) who find self-selection into R&D investment (and exporting) by high productivity plants which is reinforced by the direct impact of R&D investments (and exports) on productivity. It would be interesting to see if these findings hold for U.S. firms and how the two activities are intertwined. A newly available research dataset at the Census Bureau will enable us to now take up this line of inquiry.In summary, many avenues could be explored using these rich data to further our understanding of the innovative processes in the U.S. economy.

References

Acemoglu, Daron, Ufuk Akcigit, Nicholas Bloom and William Kerr. 2013. "Innovation, Reallocation and Growth." NBER Working Paper, No. 18993..

Acs, Zoltan J. and David B. Audretsch. 1987. "Innovation Market Structure and Firm Size." *Review of Economics and Statistics*, 69(4): 567-74.

Acs, Zoltan J. and David B. Audretsch. 1988. "Innovation in Large and Small Firms: An Empirical Analysis." *American Economic Review*, 78(4): 678-90.

Aghion, Phillipe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. 2005. "Competition and Innovation: an Inverted-U Relationship." *Quarterly Journal of Economics*, 120 (2): 701-28.

Aghion, Philippe, John Van Reenen and Luigi Zingales. 2013. "Innovation and Institutional Ownership." *American Economic Review*, 103(1): 277-304.

Akcigit, Ufuk and William R. Kerr. 2010. "Growth Through Heterogeneous Innovations." NBER Working Paper, No. 16443.

Aw, Bee Yan, Mark Roberts, and Daniel Yi Xu. 2011. "R&D Investment, Exporting, and Productivity Dynamics." *American Economic Review*, 101 (4): 1312-1344.

Aw, Bee Yan, Mark Roberts, and Daniel Yi Xu. 2008. "R&D Investments, Exporting, and the Evolution of Firm Productivity." *American Economic Review:Papers and Proceedings*, 98 (2): 451-6..

Bernard, Andrew, J. Bradford Jensen and Peter Schott. 2009. "Importers, Exporters and Multinationals: A Portrait of Firms in the U.S. that Trade Goods." In *Producer Dynamics: New Evidence from Microdata*, eds. Timothy Dunne, J. Bradford Jensen, and Mark Roberts, 513-52. Chicago and London: NBER/The University of Chicago Press, 2009.

Block, Fred, and Matthew R. Keller. 2009. "Where do innovations come from? Transformations in the US economy, 1970–2006." *Socio-Economic Review*, 7(3): 459-483.

Blundell, Richard, Rachel Griffith, and John Van Reenen. 1999. "Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms." *Review of Economic Studies*, 66(3): 529-54.

Bound, John, Clint Cummins, Zvi Griliches, Bronwyn H. Hall, and Adam Jaffe. 1984. "Who Does R&D and Who Patents?" In *R&D, Patents and Productivity*, ed. Zvi Griliches, 21-54. Chicago and London: NBER/The University of Chicago Press.

Bushee, Brian. 1998. "The Influence of Institutional Investors on Myopic R&D Investment Behavior." *The Accounting Review*, 73 (3): 305-333.

Corrado, Carol, Jonathan Haskel, Cecilia Jona-Lasinio, and Massimiliano Iommi. 2012. "Intangible Capital and Growth in Advance Economies: Measurement Methods and Comparative Results." IZA Discussion Paper Series, No. 6733.

Doi, Noriyuki. 1985. "Diversification and R&D Activity in Japanese Manufacturing Firms" *Managerial and Decision Economics*, 6(3): 147-52.

Fernandes, Ana and Caroline Paunov, 2012. "The Risks of Innovation: Are Innovating Firms Less Likely to Die?" Policy Research Working Paper, 6013.

Foster, Lucia and Cheryl Grim. 2010. "Characteristics of the Top R&D Performing Firms in the U.S.: Evidence from the Survey of Industrial R&D." Center for Economic Studies Discussion Paper, CES-WP-10-33.

Garcia-Quevado, Jose, Gabrielle Pellegrino and Marco Vivarelli. 2011. "The Determinants of YICs' R&D Activity. Institute d'Economia de Barcelona Working Paper, 2011/31.

Gilbert, Richard and David Newbery. 1982. "Preemptive Patenting and the Persistence of Monopoly." *American Economic Review*, 72 (3): 516-26.

Guellec, Dominique and Bruno Van Pottelsberghe De La Potterie. 2003. "The Impact of Public R&D Expenditure on Business R&D." *Economics of Innovation and New Technology*, 12 (3): 225-43.

Griliches, Zvi. 1980. "Returns to Research and Development Expenditures in the Private Sector." In *New Developments in Productivity Measurement*, ed. J.W. Kendrick and B. Vaccara, 419-61. Chicago and London: NBER/The University of Chicago Press.

Haworth, J.M. and P.J. Vincent. 1979. "The Stochastic Disturbance Specification and its Implications for Log-Linear Regression." *Environment and Planning*, 11(7): 781-90.

Hirschey, Mark, Hilla Skiba and M. Babajide Wintoki. 2012. "The Size, Concentration and Evolution of Corporate R&D Spending in US Firms from 1976 to 2010: Evidence and Implications." *Journal of Corporate Finance*, 18: 496-518.

Jarmin, Ron S. and Javier Miranda. 2002. "The Longitudinal Business Database." Center for Economic Studies Discussion Paper CES-WP-02-17.

Jarmin, Ron S. John Haltiwanger, and Javier Miranda. 2013. "Who Creates Jobs? Small versus Large versus Young." *Review of Economics and Statistics*. 95(2): 347-61.

Kaiser, Ulrich. 2002. "An Empirical Test of Models Explaining Research Expenditures and Research Cooperation: Evidence for the German Service Sector." *International Journal of Industrial Organization*, 20: 747-774.

Myriam, Abdelmoula and Jean Michel Etienne. 2010. "Determination of R&D Investment in French Firms: A Two-Part Hierarchical Model with Correlated Random Effects." *Economics of Innovation and New Technology*, 19(1): 53-70 .

National Science Foundation. 2014. "Business Research and Development and Innovation: 2011." NSF 15-307.

National Science Foundation. 2013. "Business Research and Development and Innovation: 2008-10." NSF 13-332.

National Science Foundation. 2011. "Research and Development in Industry: 2006-07." NSF 11-301.

OECD/Eurostat. 2005. *Oslo Manual: Guidelines for Collecting and Interpreting Technological Innovation Data, 3rd Edition.* The Measurement of Scientific and Technological Activities Proposed, OECD Publishing, Paris.

Santos Silva, J.M.C. and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economic and Statistics*, 88(4): 644-58.

Shackelford, Brandon. 2013. "One in Five U.S. Businesses with R&D Applied for a U.S. Patent in 2008." *InfoBrief*, 13-307, National Science Foundation.

Schumpeter, Joseph A. 1942. *Capitalism, Socialism and Democracy*. New York: Harper.

Slaughter, Matthew. 2009. "How U.S. Multinational Companies Strengthen the U.S. Economy." *Business Roundtable*. Spring (2).

Table 1. Summary Statistics on Firm Characteristics, Pooled by Year and Firm-Year

	1992 – 2011				
	Whole Economy[1]	**Multi-Units[1]**	**BRDISX Respondents[2]**	**R&D Performers[2]**	**TOP200[2,4]**
Firm Structure					
Multi-Units (%)	7	100	9	21	100
Employment at Multi-Units (%)	52	100	66	92	100
Multinationals (%)	1	4	1	10	83
Employment at Multinationals (%)	18	34	27	66	87
Employment (LBD)					
Mean (Year, Millions)	122	64	83	18	5
Mean (Firm-Year)	53	367	54	434	26,835
Median[3]	10	44	10	26	12,581
Standard Deviation	1,217	4,389	1,339	4,875	44,258
Firm Age					
Mean	13	17	12	13	22
Median[3]	11	18	10	11	22
Standard Deviation	9	9	8	9	9
Trade (Imports + Exports)					
Mean (Year, Billions $)	1,651	1,450	1,122	939	516
Mean (Firm-Year, Million $)	1	8	1	24	2,589
Median[3]	-	-	-	0	700
Standard Deviation	84	305	99	565	6,289
Industry Competitiveness (HHI)					
Mean	0.047	0.053	0.052	0.077	0.097
Median[3]	0.009	0.011	0.007	0.034	0.052
N (Mean-Year, 000's)	2,325	174	1,557	42	0.2

Source: Authors' calculations on the 1992-2011 LBD, SIRD and BRDIS.
Notes:
[1] LBD data. Whole economy data are limited to firms with at least five employees to be consistent with the SIRD and BRDIS sample frame, which does not sample firms with less than five employees.
[2] Matched BRDISX-LBD data. All statistics for R&D Performers and Top200 firms (including *N*) are weighted by BRDISX sample weights.
[3] Medians shown are "fuzzy" medians calculated as the mean of all observations between the 45th and 55th percentiles.
[4] To avoid potential disclosure of confidential data, some numbers shown in this column have been rounded.

Table 2. Summary Statistics on R&D

	1992 - 2011[1]	
	R&D Performers	**Top200**
R&D Expenditures ($M)		
Mean (Year)	200,900	133,100
Mean (Firm-Year)	5	640
Median[2]	0.2	280
Standard Deviation	80	1,020
R&D Intensity (R&D Expenditures/Sales)		
Mean	0.11	0.17
Median[2]	0.03	0.11
Domestic Net Sales ($M)		
Mean (Year)	8,104,000	2,800,000
Mean (Firm-Year)	200	14,000
Median[2]	4	3,400
Standard Deviation	20,500	281,000
Domestic Employment		
Mean (Year)	18,300,000	5,400,000
Mean (Firm-Year)	430	26,300
Median[2]	25	12,000
Standard Deviation	4,900	43,900
Number of Scientists and Engineers		
Mean (Year)	978,300	525,100
Mean (Firm-Year)	23	2,500
Median[2]	1	1,200
Standard Deviation	340	4,100
N (Mean-Year)	42,000	200

Source: Authors' calculations on the 1992-2007 SIRD and 2008-2011 BRDIS
Notes: Numbers have been rounded for disclosure purposes

[1] All statistics (including *N*) are weighted by SIRD sample weights and are rounded for disclosure purposes.
[2] Medians shown are "fuzzy" medians calculated as the mean of all observations between the 45th and 55th percentiles.

Table 3. Regression Results - R&D Expenditures and Firm Characteristics, 1992-2011

	Probit R&D= Yes/No (1)	Uncorrected Log R&D (2)	Corrected Log R&D (3)	Corrected Log R&D/Sales (5)
Employment (log)	0.203***	0.727***	0.781***	-0.205***
	(0.00729)	(0.0182)	(0.0193)	(0.0229)
Firm Age 6-10	-0.129***	-0.272***	-0.308***	-0.282***
	(0.0243)	(0.0646)	(0.0652)	(0.0784)
Firm Age 11-15	-0.213***	-0.391***	-0.447***	-0.447***
	(0.0286)	(0.0667)	(0.0679)	(0.0804)
Firm Age 16-20	-0.292***	-0.486***	-0.567***	-0.471***
	(0.0292)	(0.0733)	(0.0751)	(0.0823)
Firm Age 20-25	-0.358***	-0.590***	-0.687***	-0.663***
	(0.0269)	(0.0708)	(0.0722)	(0.0850)
Firm Age 25+	-0.449***	-0.620***	-0.745***	-0.618***
	(0.0269)	(0.0650)	(0.0670)	(0.0768)
Imports (log) [1]	0.367***	0.0107*	0.0186***	0.000867
	(0.0293)	(0.00418)	(0.00426)	(0.00494)
Exports (log) [1]	0.646***	0.0230***	0.0386***	0.0140**
	(0.0213)	(0.00392)	(0.00452)	(0.00539)
Multinational[1]	0.154***	0.503***	0.474***	0.369***
	(0.0364)	(0.0303)	(0.0305)	(0.0343)
Multi-Unit[1]	-0.137***	-0.0110	-0.0286	-0.0217
	(0.0245)	(0.0227)	(0.0227)	(0.0257)
Herfindahl Index (log)	0.0781***	-0.0336	-0.0100	-0.0168
	(0.00510)	(0.0195)	(0.0200)	(0.0253)
Industry Effects	Yes	Yes	Yes	Yes
Year Effects	Yes	Yes	Yes	Yes
R-square	-	0.518	-	-
N (weighted, 000's)	22,571	682	22,571	22,571

Notes: Regressions are run for firms with non-missing R&D expenditures between 1992 and 2011 that can be matched to the LBD. Regressions are weighted by BRDISX sample weight. The omitted firm age category is 0-5. The industry fixed effects are based on 4-digit NAICS industry codes.

[1]The "Trade", "Multinational" and "Multi-Unit" variables are indicator variables (1="Yes", 0="No") in the first-stage analysis and continuous (as measured by number of countries and number of units, respectively) in the second-stage analysis.

Robust standard errors are shown in parentheses. * denotes significance at the 5% level, ** significance at the 1% level, *** significance at the 0.1% level.

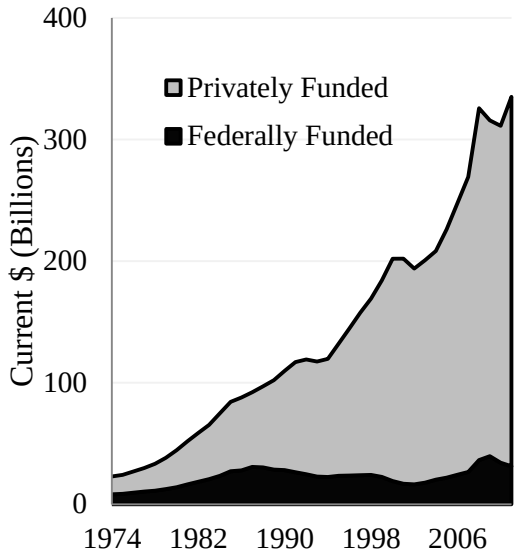
Table 4. Regression Results - R&D Expenditures and Firm Characteristics, 1995-2010

	1995-2000		2000-2005		2005-2010	
	Probit R&D= Yes/No (1)	Corrected Log R&D (2)	Probit R&D= Yes/No (1)	Corrected Log R&D (2)	Probit R&D= Yes/No (1)	Corrected Log R&D (2)
Employment (log)	0.183***	0.834***	0.169***	0.829***	0.214***	0.808***
	(0.0151)	(0.0490)	(0.0119)	(0.0346)	(0.0118)	(0.0308)
Firm Age 6-10	-0.110*	-0.458***	-0.102*	-0.150	-0.164***	-0.183
	(0.0492)	(0.122)	(0.0462)	(0.129)	(0.0381)	(0.102)
Firm Age 11-15	-0.272***	-0.714***	-0.164**	-0.253*	-0.230***	-0.235*
	(0.0513)	(0.138)	(0.0509)	(0.127)	(0.0423)	(0.107)
Firm Age 16-20	-0.285***	-0.859***	-0.181***	-0.365*	-0.286***	-0.326**
	(0.0591)	(0.146)	(0.0536)	(0.142)	(0.0459)	(0.111)
Firm Age 20-25	-0.374***	-0.987***	-0.275***	-0.745***	-0.384***	-0.339**
	(0.0516)	(0.134)	(0.0528)	(0.138)	(0.0474)	(0.127)
Firm Age 25+	N.A.	N.A.	-0.393***	-0.670***	-0.438***	-0.544***
			(0.0493)	(0.121)	(0.0401)	(0.0952)
Imports (log) [1]	0.448***	0.0367***	0.415***	0.0226***	0.310***	0.0250***
	(0.0584)	(0.0105)	(0.0403)	(0.00656)	(0.0435)	(0.00627)
Exports (log) [1]	0.788***	0.0610***	0.742***	0.0540***	0.569***	0.0166*
	(0.0464)	(0.0135)	(0.0360)	(0.00772)	(0.0334)	(0.00732)
Multinational[1]	0.0859	0.486***	0.135*	0.435***	0.189***	0.434***
	(0.0749)	(0.0706)	(0.0589)	(0.0486)	(0.0505)	(0.0441)
Multi-Unit[1]	-0.0647	-0.217***	-0.0397	-0.0460	-0.151***	0.0127
	(0.0499)	(0.0535)	(0.0371)	(0.0424)	(0.0394)	(0.0373)
Herfindahl Index (log)	0.0430***	0.0130	0.0665***	0.00832	0.138***	-0.0375
	(0.00839)	(0.0308)	(0.00889)	(0.0477)	(0.0100)	(0.0980)
Industry Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Effects	Yes	Yes	Yes	Yes	Yes	Yes
N (weighted, 000's)	7,444		7,878		7,223	

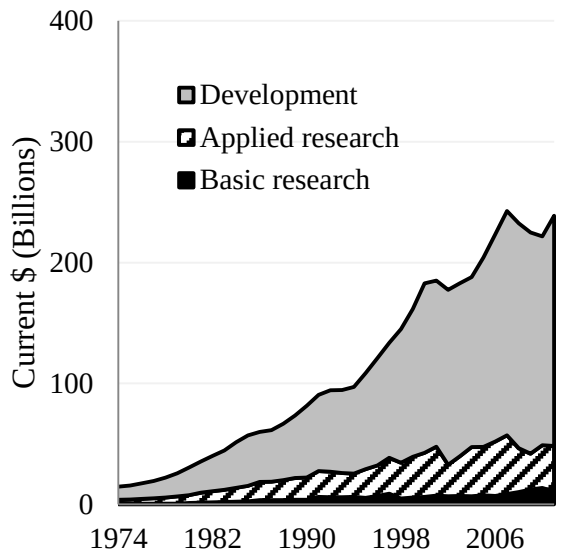
Notes: Regressions are run for firms with non-missing R&D expenditures between 1995 and 2010 that can be matched to the LBD. Regressions are weighted by BRDISX sample weight. The omitted firm age category is 0-5. The industry fixed effects are based on 4-digit NAICS industry codes.

[1]The "Trade", "Multinational" and "Multi-Unit" variables are indicator variables (1="Yes", 0="No") in the first-stage analysis and continuous (as measured by number of countries and number of units, respectively) in the second-stage analysis.

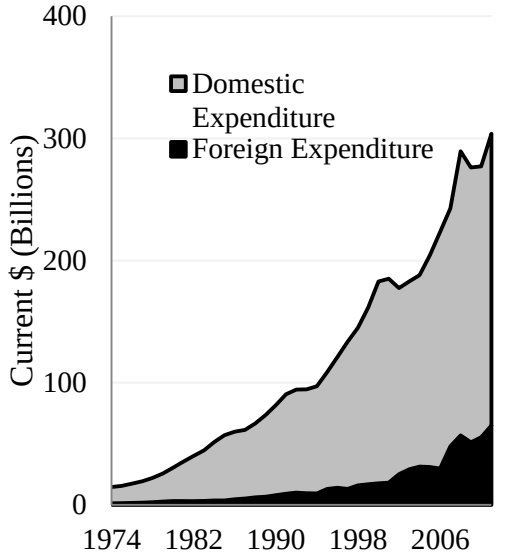
Robust standard errors are shown in parentheses. * denotes significance at the 5% level, ** significance at the 1% level, *** significance at the 0.1% level.



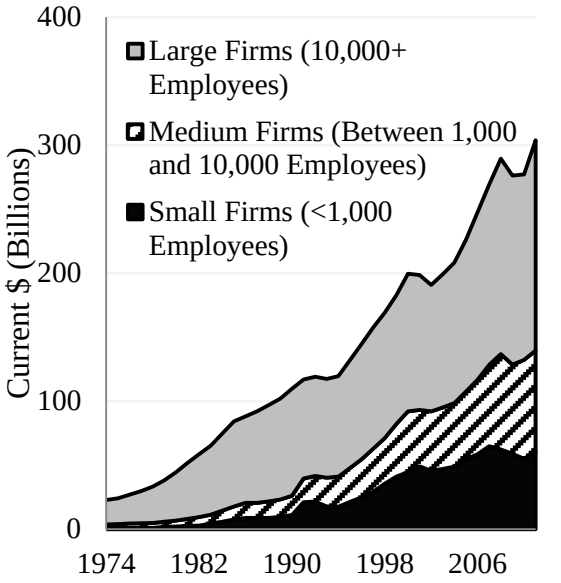
a. Total R&D by Funding Type, 1974-2011



b. Private Domestic R&D by Research Type, 1974-2011



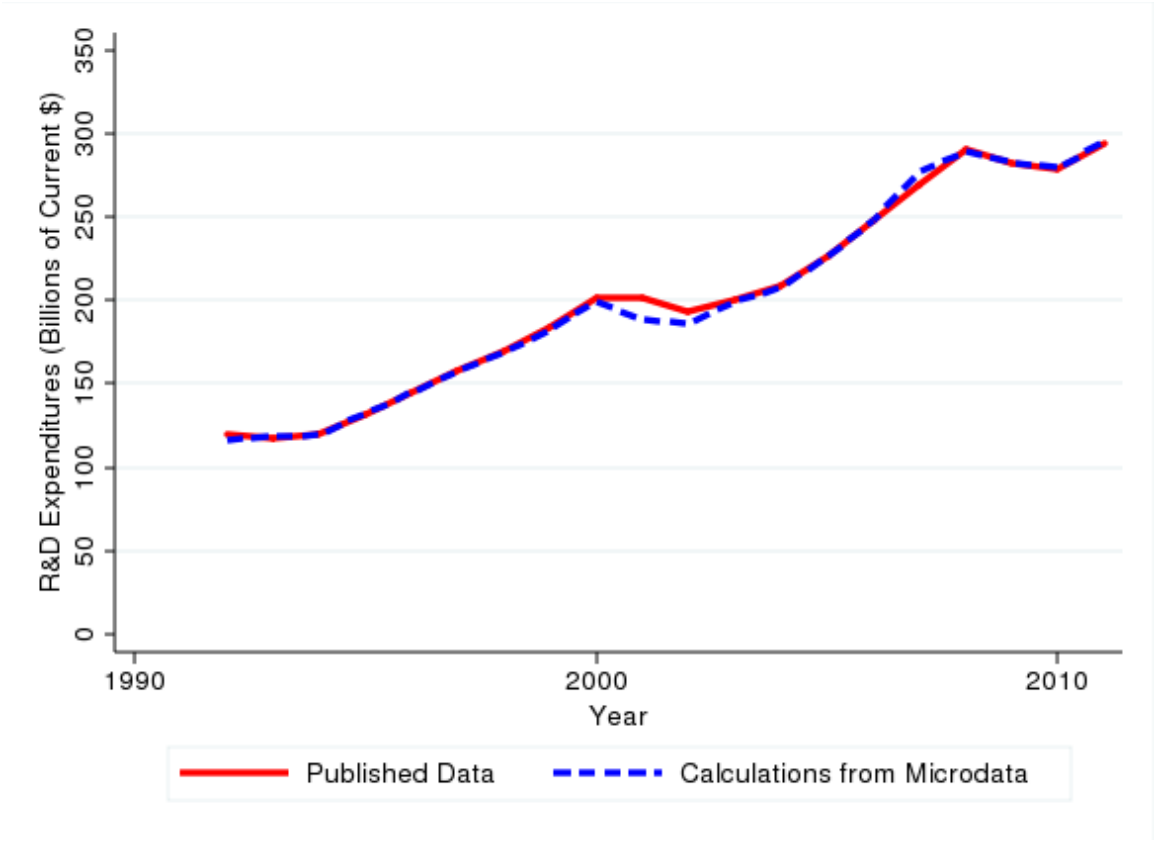
c. Private R&D by ExpenditureLocation,1974-2011



d. Private R&D by Firm Size, 1974-2011

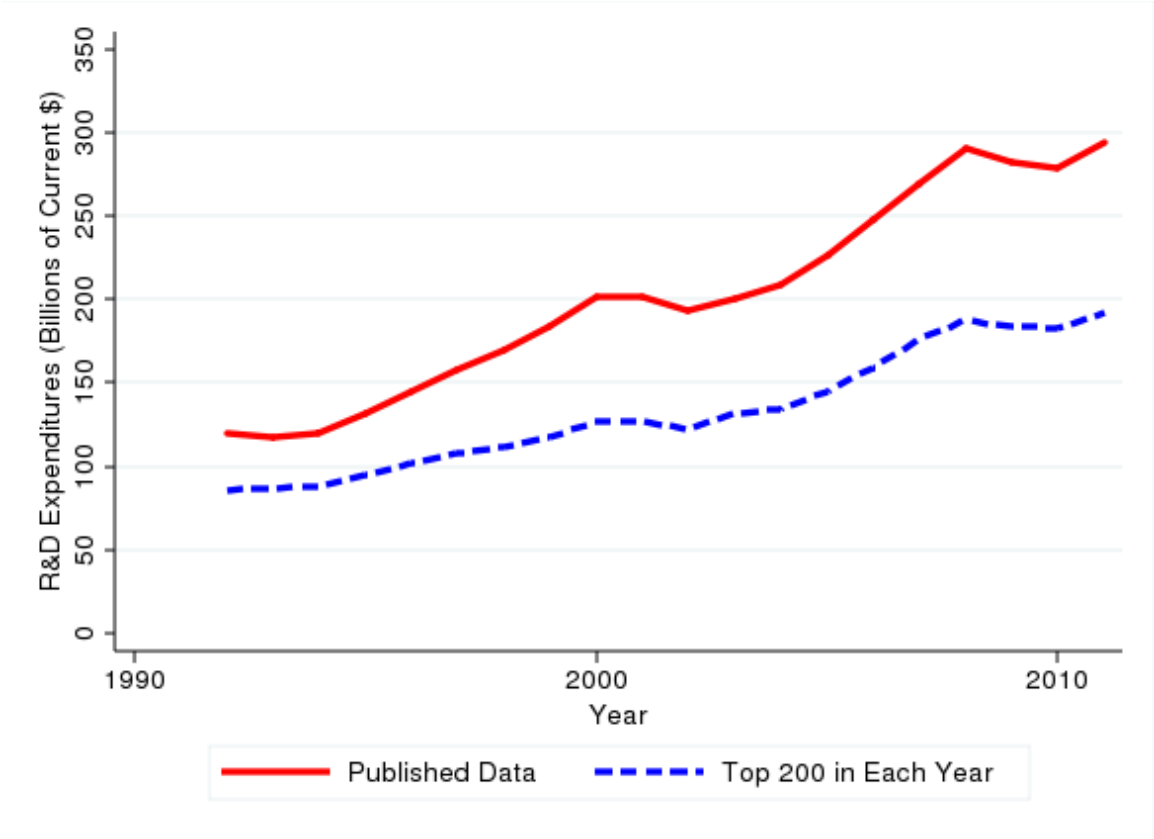
Sources: Published data are from NSF (2011, 2013, 2014).

Figure 1. Evolution of R&D Expenditures



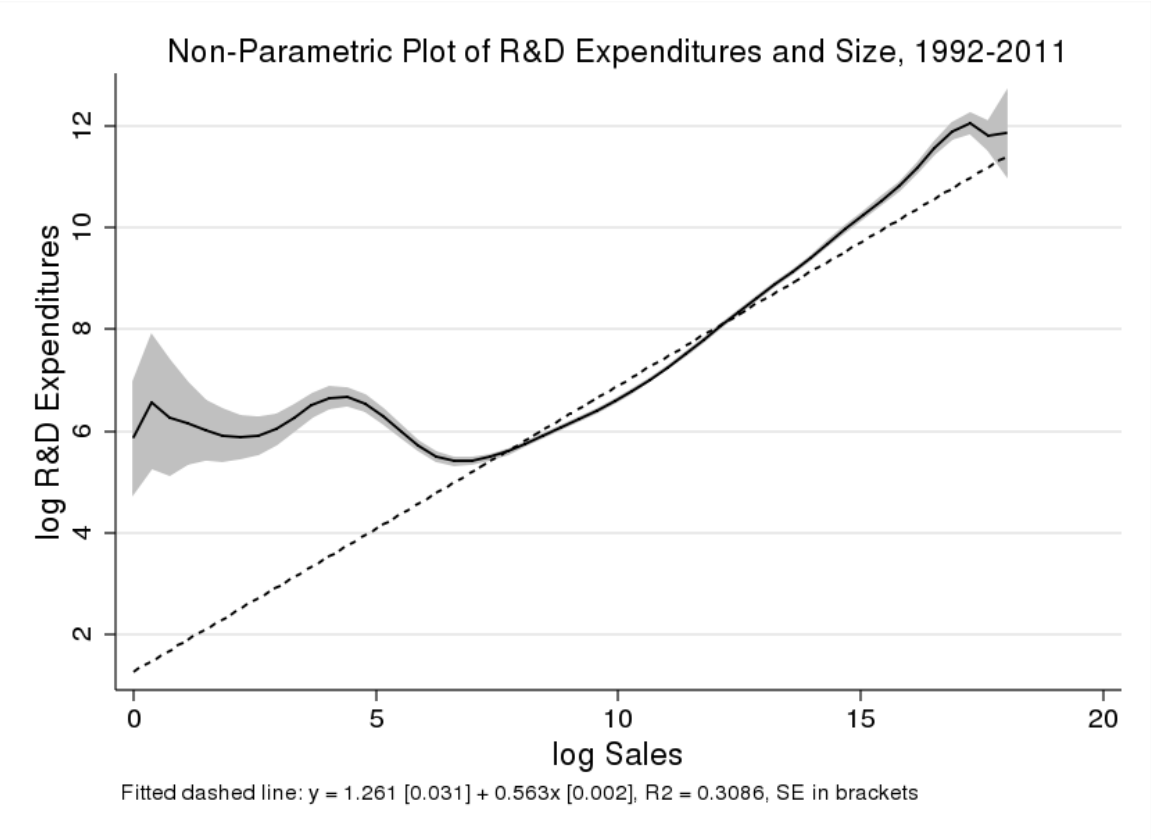
Sources: The 1953-2007 published data are from NSF (2011), Table 32. The 2008-2010 data are from NSF (2013), Table 1. The 2011 published data are from NSF (2014), Table 1. Calculations from microdata are the authors' calculations on the 1972-2007 SIRD and 2008-2011 BRDIS microdata files.

Figure 2. Total Domestic R&D Expenditures, 1992-2011



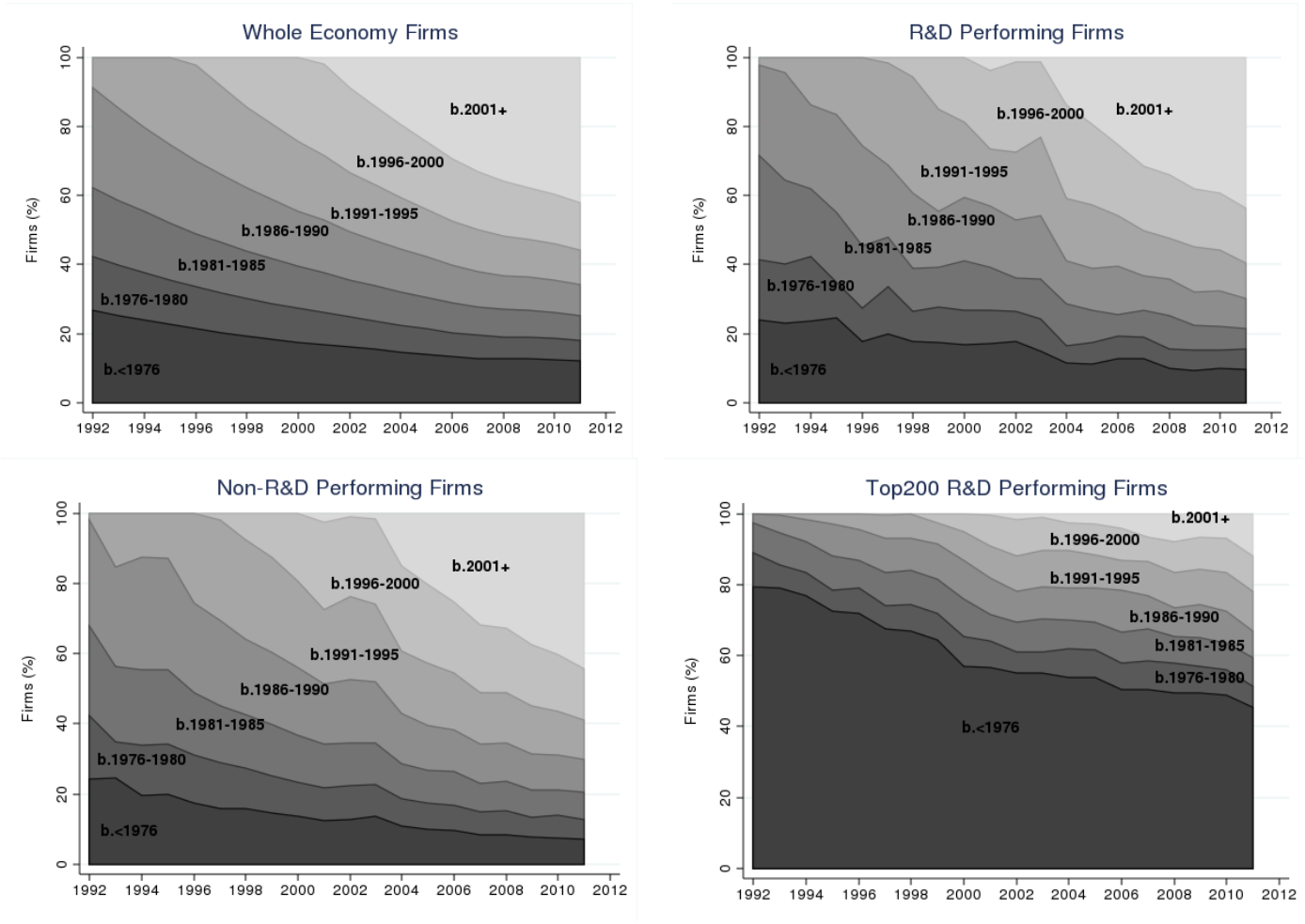
Sources: The 1953-2007 published data are from NSF (2011), Table 32. The 2008-2010 data are from NSF (2013), Table 1. The 2011 published data are from NSF (2014), Table 1. Calculations from microdata are from the authors' calculations on the 1972-2007 SIRD and 2008-2011 BRDIS microdata files.

Figure 3. Total Domestic R&D Expenditures, All Firms and Top 200 R&D Performing Firms in Each Year, 1992-2011



Source: Authors' calculations on the linked SIRD, BRDIS, and LBD microdata files.

Figure 4. Relationship between Sales and R&D Expenditures, 1992-2011

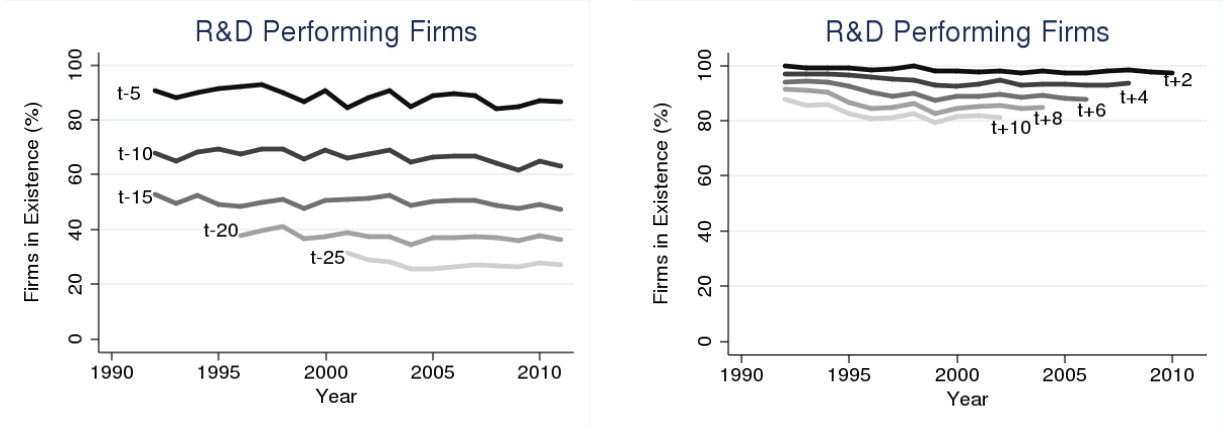


Source: Authors' calculations on the linked SIRD, BRDIS, and LBD microdata files.

Notes: The whole economy excludes firms with less than five employees for consistency with the SIRD sample frame. Firm birth year is determined as the birth year of the oldest establishment the first time the firm is seen in the LBD.

Figure 5. Allocation of Firms by Birth Cohort and Firm Type, 1992-2011

6a. Whole Economy Firms Existence Rates

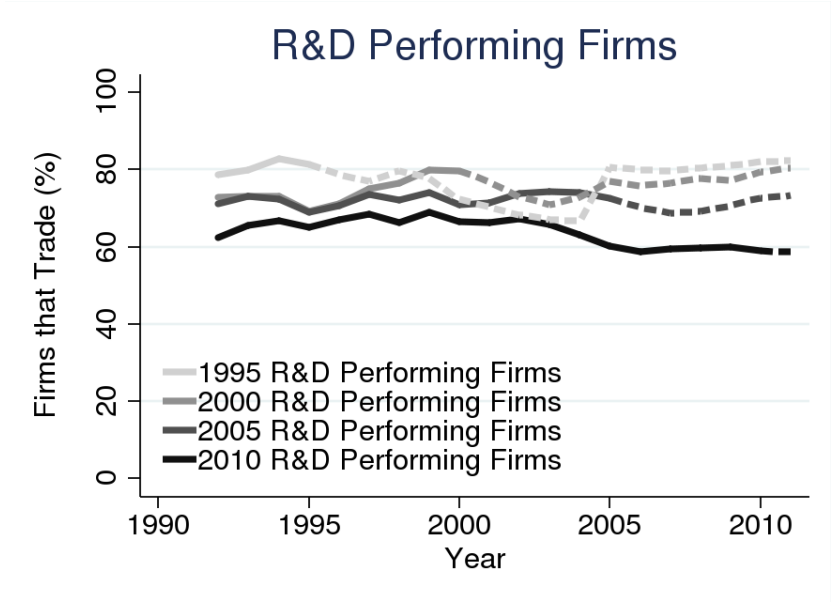


6b. R&D Performing Firms Existence Rates

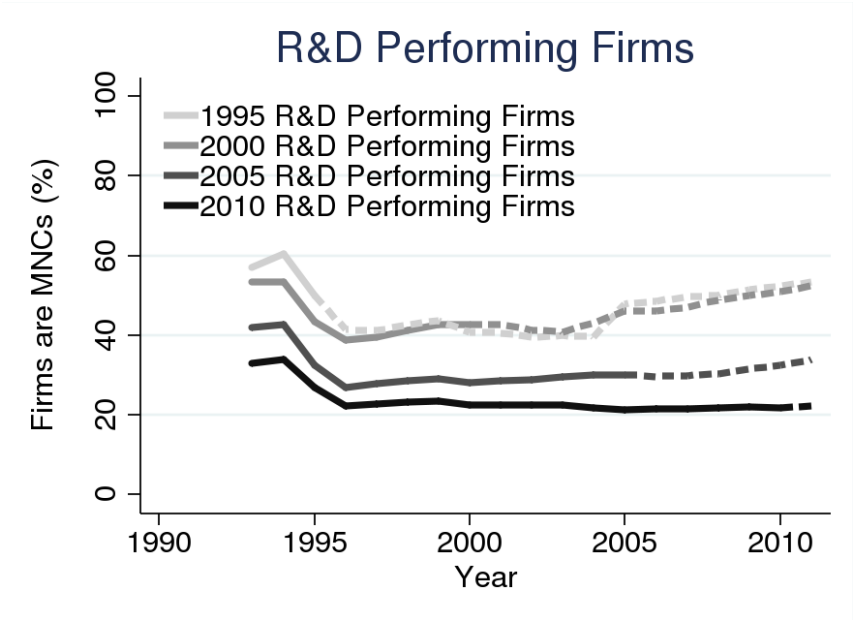
Source: Authors' calculations on the linked SIRD, BRDIS, and LBD microdata files.

Note: The *t*-20 and *t*-25 series are left-censored because LBD data starts in 1976.

Figure 6. Forward and Backward Existence by Firm Type, 1992-2011



7a. Backward and Forward-Looking Trade

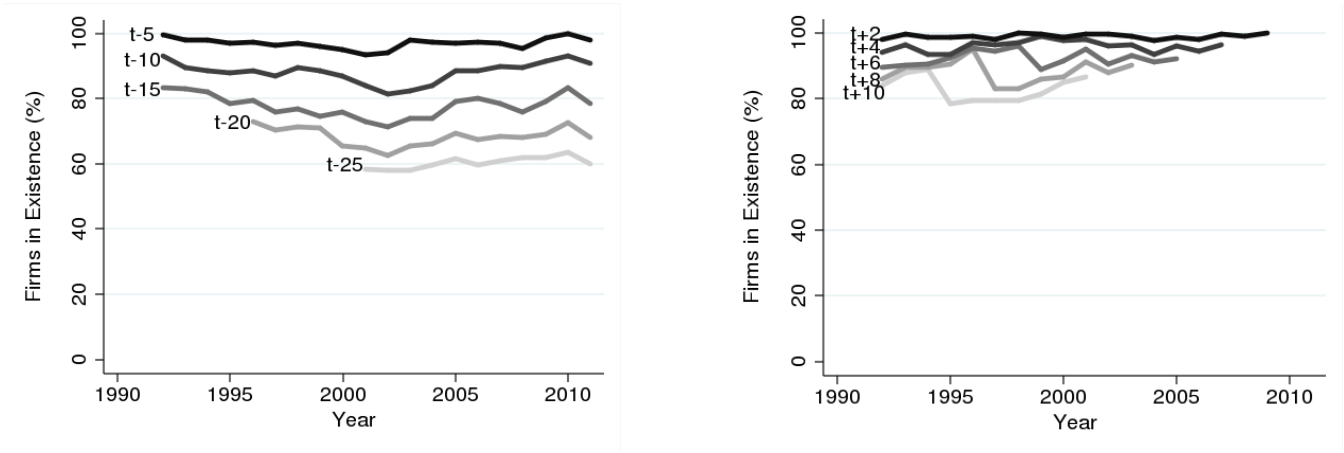


7b. Backward and Forward-Looking MNC Status

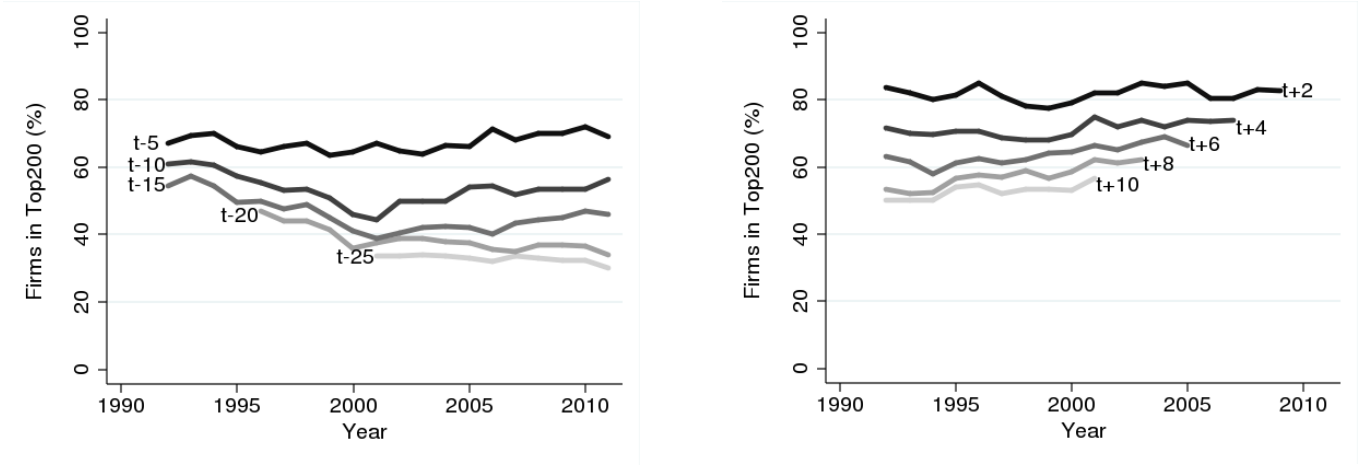
Source: Authors' calculations on the linked SIRD, BRDIS, LBD, and LFTTD microdata files.

Note: Dotted-line indicates forward-looking, while solid line indicates backward looking

Figure 7. Backward and Forward-Looking Trade and MNC Patterns by Firm Type, 1995-2010



8a. Top 200 Existence Rates, Backwards and Forwards



8b. Top200 likelihood of remaining in the Top200, Backwards and Forwards

Source: Authors' calculations on the linked SIRD, BRDIS, LBD, and LFTTD microdata files.

Notes: In panels a, b, and d, the *t*-20 and *t*-25 series are left-censored because the LBD starts in 1976, In panel c, there are no *t*-20 and *t*-25 series and the *t*-10 and *t*-15 series are left-censored because the LFTTD data start in 1992.

Figure 8. Historical Persistence of Top200 Firms, 1992-2011