HARVARD | BUSINESS | SCHOOL

Scale and Innovation During Two U.S. Breakthrough Eras

Tom Nicholas

Working Paper

15-038

October 14, 2015

Scale and Innovation During Two U.S. Breakthrough Eras

Tom Nicholas*
Harvard Business School

October, 2015

Abstract

The relationship between scale and innovation is central to R&D-based growth. This paper uncovers new empirical evidence using comprehensive data on U.S. R&D firms active during the interwar and post-WWII eras. Variability in the nature of innovation is shown to be a primary determinant of the scale effect with novel innovation scaling at approximately half the rate of normal technological discoveries. This result holds across time and for different firm types (public, private and external finance dependent). The findings help to explain why novel innovations tend to be developed in such unpredictable ways.

Keywords: Scale, Innovation, R&D,

* Soldiers Field Road, Boston, MA 02163, USA. Email: tnicholas@hbs.edu. I am very grateful to Ramana Nanda and Ufuk Akcigit for comments and to the Division of Research and Faculty Development at Harvard Business School for provided funding.

I. INTRODUCTION

The interwar and post-WWII eras represent two fundamental periods in the development of U.S. R&D (Mowery and Rosenberg, 1991; Nelson and Wright, 1992; Field, 2003). Research activity became institutionalized in corporate laboratories such as General Electric's pioneering Schenectady facility and Xerox PARC, which opened in Palo Alto in 1970. Major technological advances were developed by laboratory scientists and researchers including neoprene and nylon at DuPont in the 1920s and 1930s, and the transistor at Bell Labs in 1947. The scale of resources devoted to innovation increased substantially over time, setting a foundation for the rise of U.S. industrial leadership. While Bell Labs employed around 3,000 workers when it first opened in 1925, it employed around 17,000 workers four decades later (Temin, 1989, p.4).

This paper examines the extent to which scale at the firm-level led to the production of more and novel ideas during the interwar and post-WWII eras. The empirics are based on data compiled principally from *Industrial Research Laboratories of the United States* (IRLUS), which represents a key source of information on R&D for time periods prior to that covered by the U.S. Census Bureau's longitudinal business data (which starts in 1976). Although several works have utilized R&D data from some of the IRLUS editions (e.g., Mowery, 1995, 2005; Mowery and Rosenberg, 1989, 1998; MacGarvie and Furman, 2007, 2009; Nanda and Nicholas, 2014), this paper is the first to organize information from *all* IRLUS editions from the early to the late twentieth century and to add complementary data on patents, citations and firm types.

The relationship between scale and innovation is central to traditional approaches to R&D-based growth. Early endogenous theories specified models where innovation increases with the amount of human capital or labor in the economy (Romer, 1990; Grossman and Helpman 1991; Aghion and Howitt 1992), although inconsistencies between theory and what could be identified in the data led to concerns about the "scale effect" assumption. For example, Jones (1995) showed that there was no increase in the rate of U.S. economic growth between 1950 and 1987 despite a more than five-fold increase in the number of scientists and engineers employed in R&D. At the more micro level empirical work in industrial organization has long considered the effect of firm-size on innovation to be crucial, especially in relation to the Schumpeterian idea concerning whether large firms are more innovative than smaller ones (e.g., Cohen, 2010).

Furthermore, the scale-innovation relationship has continued to be important in recent contributions to these literatures. New endogenous theories have integrated scale by focusing on variation within the firm size distribution (Klette and Kortum, 2004). For example, Akcigit and Kerr (2010) develop an endogenous growth framework motivated by late twentieth century empirical data showing that exploitation (or incremental) innovation scales much stronger with firm size than exploration (or novel) innovation. Their emphasis on the composition of firms and innovations links to a large micro literature indicating significant sources of firm-level heterogeneity in R&D. Public firms may develop different technologies to private firms for a variety of reasons including capital availability and incentives (Ferreira, Manso and Silva, 2012; Bernstein, 2012; Aghion, Van Reenen and Zingales, 2013; Acharya and Xu, 2014).

My analysis of the IRLUS data begins by testing for representativeness using benchmarks for the number of firms surveyed, research employment and patents. Prior to the release of the U.S. Census Bureau data, empirical studies of U.S. R&D exploited what Griliches (1990, p.1675) describes as "opportunity samples." Findings from these studies are prone to selection bias due to the focus on larger publicly traded firms. As Audretsch and Acs (1991, p. 739) point out, "virtually every study examining the relationship between firm size and technical change has had to use a truncated distribution of firm sizes." Data from the IRLUS volumes represent a major advance because they cover the full distribution of firm sizes. While some imperfections and inconsistencies in the survey collection methods used can be detected over time, the benchmarking exercise suggests that the IRLUS data closely track main trends in U.S. R&D.

Next, I estimate the relationship between scale, measured by research employment, and innovation using the standard knowledge production function approach (e.g., Griliches, 1979). I use raw patent counts and citations to capture the level and quality of innovation respectively and I also examine the type of innovations being developed using a patent *generality* measure (Hall, Jaffe and Trajtenberg, 2001). This acts as a proxy for the most novel and creative technological discoveries. In baseline specifications, I assume a contemporaneous linear relationship between R&D inputs and outputs, but I also relax these assumptions exploring lagged and curvilinear relationships in robustness checks. Finally, I examine firm-level heterogeneity by estimating the baseline specifications across a number of groupings of public and private firms, and firms that were active in high, or low, external finance dependent industries.

The main results indicate that scale was positively related to the level and quality of innovation, but the novelty of technological discoveries was considerably more invariant to scale over the firm size distribution. To the extent that larger firms tend to focus on incremental invention with respect to existing product lines (e.g., Baumol, 2002), and account for the largest share of R&D investment in aggregate, the results offer an explanation for why the arrival rate of novel types of innovations tends to be so unpredictable. The absence of a strong scale effect may simply amplify the inherently random nature of radical technological change.

These findings are strikingly similar across the interwar and post-WWII eras, despite large differences in the historical context, suggesting they represent stylized facts about underlying firm and innovation dynamics. The effects I find vary more by the type of technology being developed, than over time as the U.S. R&D sector scaled. There is some evidence of firm-level heterogeneity in the regression specifications but the results generally indicate that variability in the nature of technological discovery was a fundamental determinant of the relationship between scale and innovation. Given the representativeness of the IRLUS data the findings are generalizable. They help to establish the sources of variation in innovation and the areas in which firms were more homogenous for two significant breakthrough eras in U.S. R&D history.

The remainder of the paper is organized as follows. Section II provides a brief historical background to the development of R&D in the United States during the early-to-late twentieth century. Section III describes the data, Section IV examines the empirical setup, Section V presents the main results and robustness checks and Section VI presents results from testing for firm-level heterogeneity. Section VII concludes.

II. BREAKTHROUGH ERAS IN U.S. R&D

The scale of U.S. R&D increased dramatically during the twentieth century. Early examples of corporate labs include those founded by Thomas Edison in Menlo Park, and Alexander Graham Bell in Boston, both in 1876. Edison's New Jersey lab was a reasonably sophisticated workshop with about 40 employees at its peak, but Bell's lab amounted to no more than a garret in a boardinghouse (Hounshell, 1988, p.3). By approximately 1900 the corporate sector had developed on such a scale that innovation was brought increasingly within the boundaries of firms. R&D facilities became larger and the pace of growth was rapid. The number of scientists and engineers employed in industrial research labs more than doubled between 1921 and 1927 from 2,775 to 6,274 and more than quadrupled between 1933 and 1946 from 10,918 to 45,941 (Mowery and Rosenberg, 1998, pp.21-22). Even during the Great Depression, U.S. R&D expanded creating significant productivity benefits (Field, 2003).

Important consequences followed from a change in the organizational structure of R&D. Boundaries between basic and applied science became increasingly blurred and research directors debated the appropriate mix of research activity (Rosenbloom and Spencer, 1996). Some labs like Eastman Kodak's in Rochester, NY and DuPont's Experimental Station in Wilmington, DE invested heavily in basic research to develop radical innovations. This focus was considered to be essential to attract star scientists, but it also needed to foment commercial applications to be viable in the long run (Hounshell and Smith, 1988). Several pure scientific advances such polymer chemistry occurred in academia and a nexus of universities and corporate labs began to emerge, creating a foundation for modern science-based industry. MacGarvie and Furman (2005) show that from 1927 to 1946 university research programs had a strong causal impact on the direction of technological change in the pharmaceutical industry.

The Second World War marked an important turning point. Wartime research was responsible for some of the major technological innovations of the twentieth century in fields such as electronics and communications, and it also induced a fundamental change in how scientific knowledge was pursued. Following the publication of the seminal *Science, the Endless Frontier* by Vannevar Bush, who was director of the wartime *Office of Scientific Research and Development* and an advisor to the Roosevelt Administration, the government became more actively involved in funding research (Stephan, 2014). But, although the growth of federal funding transformed the pathway of U.S. R&D and science, it is important to note that firms also invested heavily in their own right with, "a vast expansion in the number of American companies doing R&D and in the size of their R&D programs". By the mid-1960s private funds accounted for about half of corporate R&D spending (Nelson and Wright, 1992, pp.1951-53).

Despite this growth in the scale of resources devoted to R&D we know very little beyond aggregate R&D statistics, and dynamics at the firm-level remain a largely unexplored area. This represents an important gap in our understanding. Growth theorists frequently lament the lack of good data for informing theory (e.g., Klette and Kortum, 2004) and this is even more relevant historically because it impedes the study of long run changes. For example, while Baumol (2002) argues that large firms in the post-WWII years focused more on incremental innovations designed to enhance markups on existing product lines, as opposed to radically new advances, this claim is impressionistic given gaps in the availability of evidence. The next section introduces new firm-level data to shed light on the relationship between scale and innovation during these important U.S. breakthrough eras.

III. THE FIRM-LEVEL DATA

R&D Firms from IRLUS

The main source for the R&D data is IRLUS, which was first published by the *National Research Council* (NRC). The NRC was established in 1916 to advise the U.S. government on science and industrial research. It began collecting data on corporate R&D soon after WWI and it subsequently started to publish the results of its direct correspondence surveys with firms. The IRLUS surveys are analogous to the extensively used modern "Yale Survey" and "Carnegie Mellon Survey" of R&D summarized in Levin et al., (1987) and Cohen et al., (1992). I use all firms listed in 12 editions of IRLUS, published in 1921, 1927, 1931, 1933, 1938, 1940, 1946, 1950, 1956, 1960, 1965 and 1970.[1] The IRLUS data exclude laboratories managed by federal, state or local agencies, so the coverage relates specifically to the corporate R&D sector.

The final dataset consists of 11,514 firms. The NRC surveyed across the firm size distribution. This aspect of the data is important from the standpoint of identifying the sources of variation in technological change given that several studies have shown that smaller firms are more innovative than larger firms (e.g., Acs and Audretsch, 1990). I partition the data by the interwar years including WWII and the post-WWII years as these seem natural break points in the data based on the historical context described in section II. Figure 1 shows the geographic location of laboratories for these major time periods. The concentration of facilities in the east coast and mid-western manufacturing belts is clearly illustrated in the data. This spatial distribution of facilities is consistent with what is known about the economic geography of manufacturing during the early-to-late twentieth century (Glaeser, 2012).

Descriptive statistics are provided in Table 1. Systematic information is included in IRLUS volumes on the total number of research workers employed in R&D, which I use to proxy for scale. This is a standard firm size measure going back to the work of Scherer (1965). Some editions of IRLUS also document the number of scientists and engineers and other technical personnel, which I also use in empirical specifications as a robustness check. Unfortunately research expenditure data are not available and neither are Standard Industrial Classification

[1] The first edition was published in 1920, but the 1921 volume has a fuller coverage of firms, hence I use that edition as the starting point. Beginning in 1965 IRLUS was published by the R. R. Bowker Company.

codes are not included. To address the latter, firms were sorted into 16 industries based on descriptions in the volumes of their main areas of R&D activity.[2]

Patent, Citation and Innovation Novelty Metrics

The IRLUS volumes provide no data on R&D outputs. I therefore used complementary sources. I hand-matched the database of NRC firms with assignees of U.S. patents between 1920 and 1970 using a ten year event window around the year that a firm was observed in the IRLUS volumes. A wide event window increases the accuracy of the patent-match, but it also creates an "excess zeros" problem. In the empirics I mainly use only firms patenting at least once during this window, but I also include specifications run on all firms to test for any bias. Although the surveys were conducted for snapshot years, U.S. patent data are reported annually, so the matching process created a year-on-year count of patents that were assigned to each firm. I use patents by their application date to ensure proximity between the timing of R&D inputs and outputs. Firms in the IRLUS volumes were assigned 730,026 patents between 1920 and 1970. Two-thirds of the 11,514 firms in the dataset patented at least once during this time period.

To examine the types and quality of technologies being developed, I merged the patent data against a dataset containing all citations by U.S. patents to patents granted between 1947, the first year citations were systematically recorded on patent documents, to September 2008. This resulted in 4.3 million citations being matched to the 730,026 patents.

To correct for any biases associated with variable citation lags over time, I followed standard practice and calculate *scaled citations* by measuring the number of citations relative to matching patents in the same 3-digit USPTO technology class in the same year. Thus, if C is the number of citations to a patent in technology class f in year t and c_{ft} is the average number of citations to patents in that class, the ratio C_{ft}/c_{ft} is the scaled citations measure. I take the annual mean of this ratio for each firms' portfolio of patents to create firm-year observations.

Additionally, I use the citations data to construct a *generality* metric to identify novel innovations. Generality identifies patents that have an influence on a broad range of future inventions. For example, William Shockley's 1950 patent (2,502,488) for a semiconductor amplifier in USPTO class 330 is cited 25 times across 9 distinct 3-digit USPTO technology

[2] The industry categories are as follows: Auto; Chemicals; Communications, TV, Radio; Electrical Equipment & Electronics; Food & Tobacco; Machinery & Machine Tools; Medical Equipment; Metals; Mineral Products; Miscellaneous; Paper & Products; Petroleum & Coal; Rubber & Plastics; Scientific Instruments & Related; Stone, Clay & Glass; Textiles & Leather.

classes. Generality is defined as one minus the Herfindahl index of patent citations by 3-digit class. I use the bias-adjusted measure described in Hall (2005), which addresses bias when the number of citations in a technology class is small. I also scale the generality measure relative to matching patents in the same 3-digit USPTO technology class in the same year and create mean firm-year observations. The *scaled generality* measure is less likely to be confounded by spurious variation over time or in the cross section of technologies.

Distinguishing Firm Types

Firm-level heterogeneity may affect the relationship between scale and innovation, some of which can be observed by grouping firms into categories. The most common groupings in the literature are public versus private firms and firms that relied more or less on external finance. Firms may innovate more, or less, conditional on capital availability.

I used the CRSP files to obtain the names of firms whose stocks were traded on the New York Stock Exchange from 1925, and I matched these to the IRLUS firm names. For years prior to 1925, public listing was identified from the weekly finance publication, the *Commercial and Financial Chronicle*. These data reveal the changing nature of the relationship between scale and innovation. Figure 2 illustrates that the R&D sector accounted for less than 20 percent of U.S. patents in 1920 but over 50 percent of U.S. patents by 1970. In 1920 private and publicly-traded firms accounted for closely comparable shares of U.S. patents. Yet, by 1970 the share accounted for by publicly-traded firms was over four times larger.

I used Rajan and Zingales' (1998) measure to categorize firms according to their relative reliance on external finance. Rajan and Zingales calculate the level of dependence for U.S. firms in the 1980s using capital expenditures minus cash flow from operations over capital expenditures. As a robustness check of the Rajan-Zingales categorization of finance dependence for historical periods, I use the measure provided by Nanda and Nicholas (2014) for the interwar years. They calculate bank notes payable over fixed assets for all firms listed in the IRLUS volumes that were also covered in the financial publication *Moody's Manual of Industrials*.

Testing for Representativeness

An important issue is representativeness. While the early IRLUS editions cover close to the universe of U.S. R&D firms because the sector in aggregate was much smaller, the later editions are best thought of as samples of R&D firms, which may be subject to some potential biases. For example, the 1940 edition of IRLUS includes 2,264 companies compared to 1,769 for the 1938 edition, which the editorial introduction attributes to "the fact that the questionnaire was sent to a larger number of companies than has been done heretofore" (IRLUS, 1940, preface). Furthermore, direct correspondence surveys are notoriously prone to non-response bias. This kind of bias cannot be corrected retrospectively, but some indication of representativeness can be determined by benchmarking the main variables from the IRLUS data.

Figure 3 shows data for total research employment in the IRLUS surveys and breakdowns by scientists and engineers employed for specific years. Reassuringly the employment levels from my data collection coincide for overlapping years with the number of scientists and engineers employed reported in Mowery's (1981) study, which also uses the IRLUS volumes.

More generally, the Bureau of Labor Statistics reports a series for scientists and engineers employed in R&D. This tracks closely the total research worker series from the IRLUS for most of the time period but the BLS series exhibits a higher growth rate than the series for scientists and engineers compiled from IRLUS, and the trajectories are somewhat different for the final observation year, 1970. One explanation for the discrepancy between the BLS and IRLUS data is that the definitions of what constitutes a "scientist" or "engineer" or "research and development" could vary. Also, the BLS data are known to cover around four-fifths of all scientists and engineers employed, whereas the IRLUS volumes may cover a smaller share.

To get a sense of this latter potential source of bias, Figure 4 relates the IRLUS data to the number of firms reported in the NSF's *Survey of Industry Research and Development*. This comparison shows the IRLUS data includes around 37 percent of firms over the snapshot years, ranging from a high of 44 percent of firms in 1956 to a low of 31 percent in 1970. Relatively less coverage of firms in 1970 may help to explain the divergence in the series in Figure 3 for total research worker employment when comparing the IRLUS firms with BLS data for that year.

Finally, Figure 4 benchmarks patents and shows an extremely close correspondence between the time series for the number of patents assigned to publicly traded firms from the IRLUS surveys and the patent series for all publicly traded firms reported by Kogan, et al., (2012).

In sum, while the IRLUS data do not cover all firms due to the nature of the direct correspondence methodology and the possibility of non-response, they still provide a good coverage of corporate research activity across time and the firm size distribution. Moreover, the data appear to be reasonably reflective of the main trends in R&D activity and patenting.

IV. EMPIRICS

The Knowledge Production Function

A useful starting point for the empirical analysis is the knowledge production function first developed by Griliches (1979). Briefly stated this model assumes that knowledge is a key factor of production, from which specifications can be derived expressing R&D output in terms of R&D inputs. Typically output is measured by patent counts and R&D employment or expenditures proxy for inputs (e.g., Scherer, 1965). Several studies have used this methodology to investigate the link between innovation, scale and R&D as summarized in Cohen (2010). With firm-level data, the estimating equation usually takes the following form:

$$PATENTS_{it} = \beta_1 R\&D_{t-n} + \phi_i + \gamma_t + \varepsilon_{it} \qquad (1)$$

Patents of firm i at time t are related to R&D contemporaneously, or with a lag. The parameters ϕ controls for unobservable firm attributes that are fixed over time and γ represents year fixed effects to control for annual shocks, such as demand conditions or changes in the propensity to patent. Firm-level characteristics tend to be particularly important to explaining variability in R&D (Cohen, 2010). As such it is important to note that equation (1) is not an explicit empirical test of traditional type growth models where the unit of analysis would be a country not a firm and the left hand side variable would be a growth rate in a covariate like TFP. But equation (1) is an explicit input to newer generations of endogenous growth. Klette and Kortum (2004) use estimates of β to infer the relationship between R&D and innovation. Akcigit and Kerr (2010) consider the scaling properties of different technologies at the firm-level.

When elasticities are recovered from this type of regression in the literature, β is larger in the cross-section than it is in the within-firm dimension, but both estimates suggest diminishing returns to scale with β in the range of 0.4 to 0.7 (Gurmu and Pérez-Sebastián, 2008). In traditional endogenous growth models R&D generates spillovers of non-excludable knowledge implying increasing returns. These are more likely to be observed in at higher (industry or country) levels of aggregation (Audretsch and Feldman, 2004).

In the literature, knowledge is assumed to be homogenous in equation (1) and captured by patents, yet the characteristics of innovation are highly heterogeneous. Distinguishing the type of knowledge produced is central to newer growth models with heterogeneous technological change (Akcigit and Kerr, 2010). Accordingly, I use three outcome measures to pick up varieties of innovation: patents, scaled citations, and scaled generality. These metrics measure the level, quality, and novelty of technological discovery respectively. With input use measured by R&D employment, which should produce slightly larger estimates of the elasticity than other scaling variables like assets or sales (Scherer, 1965, p.258), estimates of β across these outcome variables reflect the extent to which different types of technology scaled with firm size.

Measurement and Econometric Issues

Three estimation issues stand out with respect to equation (1): specifying the lag, establishing the correct functional form, and identifying β given the endogeneity of R&D. The first issue is straightforward to address. Because the U.S. patents I use are measured as of their application date there should be a close association between R&D inputs and patent-based outputs. Firms typically patent their ideas close to the point of invention in order to mitigate expropriation risk. Griliches (1990) survey suggests temporal alignment should be close, as do several studies that include evidence on innovation gestation periods (e.g., Scherer, 1965; Hanlon, 2014). To verify this assumption in the data I estimate using distributed lags of R&D employment.

Scherer (1965) was among the first to think seriously about the most appropriate functional form of patent-R&D regressions. Using a sample of the largest firms active during the 1950s he established an approximately linear relationship between R&D employment and patents and in fact found a coefficient slightly above unity in a logarithmic regression of patents on R&D employment. Notwithstanding linearity appears to be a reasonable overall assumption, I also test the robustness of the results to assuming a curvilinear relationship between the innovation measures and R&D. Specifically, I use squared and cubic terms in R&D employment and estimate marginal effects on the outcome measures across the range of firm sizes in the data.

Finally, identification is a significant challenge in the estimation of patent-R&D regressions. Market conditions and technological opportunity will influence the direction of causation between R&D inputs and outputs. If firms change R&D in response to their anticipation of future shocks to the production function, the parameters of the knowledge-based specification can be biased. While I am unable to make progress in this area due to a lack of plausibly exogenous variation in R&D investment, using firm fixed effects and distributed lags should mitigate some of the bias. Moreover, an interesting first step is to uncover the basic correlations in the new data and then to interpret these effects with respect to the long run historical context.

V. MAIN RESULTS AND ROBUSTNESS

Baseline Estimates

Table 2 reports parameter estimates from knowledge production function specifications for different dimensions of technological knowledge (patents, citations and generality) using the logarithm of contemporaneous research workers to proxy for the scale of R&D. Because the number of research workers is only available for the snapshot IRLUS volumes, I linearly interpolate this variable to provide year-on-year estimates for the full panel of data. R&D employment is a slow moving variable and the gaps between the survey years are reasonably small (an average of 4.5 years with a minimum of 2 years and a maximum of 6 years). All of the specifications are in "log-log" form so the parameters can be interpreted as elasticities. Although elasticities are unit-free measures, I also report standardized coefficients to facilitate more accurate comparisons of economic magnitude across the specifications.

Panel A reveals a positive scale effect during the interwar and WWII era when using patents as a dependent variable, but with diminishing returns, especially in the within-firm dimension. According to the elasticity estimates in columns 1 and 2 (without and with industry fixed effects) a 10 percent increase in research workers is associated with a 5 percent increase in patents. Using firm fixed effects (column 3) produces much smaller estimates in the region of a 1.5 percent increase. Including zero observations for the dependent variable (i.e., firms that did not patent) in columns 4 and 5 with industry and firm fixed effects does not change the substantive results. I find little bias in the scale effect from sorting on the dependent variable in columns 1 to 3.

Columns 6 to 10 reveal that the scale effect is remarkably stable over time. The parameters are close in size in the post-WWII years to those estimated in columns 1 to 5 for the interwar years and WWII. If technological change was becoming progressively harder (e.g., Kortum, 1997) we might expect to observe diminishing returns to scale over time. The data suggest a long run relationship between scale and innovation that is inconsistent with this hypothesis.

It is well known that patents vary considerably by their quality. Yet, Panel B shows that there are no confounding effects due to variation in the quality of technological discoveries. In fact, the pattern in the coefficients across the patent citation specifications is quite consistent with what is shown in Panel A with raw patents as a dependent variable. The elasticities are larger in the cross section (column 1) or with industry fixed effects (columns 2 and 4) than in the within-firm dimension (columns 3 and 5) and they are also highly stable across time when comparing the interwar years and WWII with the post-WWII era.

In Panel C, however, the results suggest strong variation in the size of the scale effect by the type of technology being developed. Like in Panels A and B, the coefficients are highly statistically significant across the specifications but their economic magnitude is much smaller in Panel C. Standardized coefficients measuring the effect of a change in research workers on scaled generality are approximately half the size of the coefficients for raw patents (Panel A) and scaled citations (Panel B). Interestingly, the effect for scaled generality is also consistently estimated when comparing the coefficients in columns 1 to 5 for the interwar years and WWII with the coefficients in columns 6 to 10 for the post-WWII period.

I obtain very similar substantive results when using *scientific* workers rather than *all* research workers to estimate these parameters though there is a slight convergence in terms of economic magnitude across Panels A, B and C. Standardized coefficients in Table A1 in the Appendix reveal that the effect of a change in scientific research workers on scaled generality is approximately two-thirds of the size of the coefficients for raw patents (Panel A) and scaled citations (Panel B). Overall, the results show that the level and quality of innovation scaled much stronger with respect to firm size than novel innovation during these breakthrough eras.

Testing for Lagged Effects

One possible source of bias is variability in the timing of patents as an output measure with respect to R&D employment as an input measure. According to Scherer (1965, p.1097) "[approximately] nine months pass between the conception of an industrial invention and the filing of a patent application." Furthermore, any variable gestation periods by type of technological development may be particularly confounding to the baseline results. Normal and novel technologies may be subject to different time dependencies. Estimates in Table 2 only assume a contemporaneous relationship between innovation and R&D.

Table 3 replicates the specifications from columns 2 and 3 and columns 7 and 8 in Tables 2 using distributed lags of research worker employment from time $t=0$ to time $t-4$. From the patent and citations specifications in Panels A and B it can be seen that the contemporaneous relationship accounts for most of the total effect (the sum of the coefficients), both with industry and firm fixed effects. The contemporaneous relationship is also statistically strongest in the scaled generality specifications (Panel C). Notably, however, the sum of the coefficients is insignificantly different from zero at the customary levels in the fixed effects estimates. In general the baseline results are robust to relaxing the assumption of a contemporaneous link between innovation outcomes and R&D, while the most demanding specifications in Table 3 imply a somewhat weaker relationship between scale and innovation novelty than between scale and both the level and quality of technological change.

Functional Form

As a further robustness check, I use quadratic and cubic terms in R&D employment. So far, the specifications have estimated an average effect over all values of R&D, yet as Scherer (1965) showed there may be a non-linearity in the effect over the range of the firm size distribution. Because including higher order terms of the R&D variable in equation (1) means running the regressions with continuous variable interactions, significance needs to be estimated at particular values to recover meaningful marginal effects and standard errors. Figures X and X plot average marginal effects and 95 percent confidence intervals for the effect of research workers on the outcome measures over the distribution of research workers observed in the data.

Figure 6 shows that there is an economically larger effect of R&D on patents in the patent and scaled citation specifications with industry fixed effects for the interwar and WWII era, with average marginal effects that are statistically indistinguishable from unity for the very largest R&D firms. This can also be observed in the post-WWII data although the effect begins to attenuate slightly in the upper end of the research worker distribution. Estimates for scaled generality exhibit diminishing returns to scale for larger firms across both eras. This is consistent with larger firms focusing on incremental innovation (captured by the patent and scaled citation measures) and small and medium-sized firms focusing more on novel technological discoveries.

The pattern of these average marginal effects is robust to the inclusion of firm fixed effects in Figure 7. Although the standard errors are larger and so the confidence intervals frequently overlap relative to estimating with industry fixed effects, formal Wald tests reject polynomials up to order three in only one of the fixed effects specifications (scaled generality for the post-WWII era). In terms of economic magnitude the estimated effects for patents and scaled citations are much larger than for scaled generality. While there is some evidence of non-linearity, the curvilinear estimates are broadly consistent with the baseline results from Table 2 concerning the varying relationship between scale and the types of innovations being developed.

VI. Firm-Level Heterogeneity

Different subgroups of enterprises may be more, or less, able to scale different kinds of technologies, an especially pertinent issue given the changing trends over time in the organization of firms engaged in R&D (Figure 2). While the estimates have already gone some way towards alleviating this concern using firm fixed effects, the fact that the size of the coefficients does change when controlling for time invariant firm-level unobservables merits further analysis. One explanation is measurement error, which can have a particularly strong impact on the fixed effects estimator, biasing the coefficients downwards. Another explanation is that the firm fixed effects reflect economically important variation.

The literature suggests several sources of firm-level heterogeneity, which can be explored in an estimation context. Ferreira, Manso and Silva (2013) develop a model where the "short-termism" of the financing environment biases publicly-traded firms towards more conventional technological developments. Alternatively, Aghion, Van Reenen and Zingales (2014) argue that career concerns from undertaking risky innovation projects are mitigated in publicly-traded firms with higher levels of institutional ownership. Acharya and Xu (2014) find that in financially dependent industries public firms perform more novel R&D than private firms, whereas public and private firms have similar innovation outcomes in industries where access to external capital matters less. Related contributions also using modern data include Seru (2013) who finds that conglomerates relying on internal capital markets develop less novel technologies and Bernstein (2014) who finds public ownership leads to a focus on more incremental innovations.

Given the richness of the data in the IRLUS volumes it is possible to examine the contemporaneous relationship between R&D and innovation at firm group levels. I estimate equation (1) using patents, scaled citations and scaled generality as dependent variables and plot the point estimates and 95 percent confidence intervals in Figures X and X for industry and firm fixed effects specifications respectively. Graphical presentation eases the task of comparing coefficients across the firm type categories. Public and private firms are distinguished by the date of their stock market listing. External finance dependence is determined discretely by above ("high") and below ("low") median values of the Rajan-Zingales measures for U.S. firms in the 1980s. Here, I assume the Rajan-Zingales measures are time invariant, but as an additional check I also report estimates in the Appendix Figure A1 for the alternative measure of finance dependence constructed by Nanda and Nicholas (2014) for the interwar years.

Three main findings emerge. First, the estimates are highly sensitive to the specification of the type of fixed effects. For example, with industry fixed effects (Figure 8) the elasticity of R&D with respect to patents and scaled citations suggests statistically significant differences between public and private firms during the post-WWII era, yet in Figure 9 with firm fixed effects the point estimates are much smaller and the confidence intervals also overlap. As in the baseline estimates, this could reflect attenuation bias due to measurement error across the IRLUS surveys. Furthermore, the larger standard errors would be expected given that with firm fixed effects, the coefficients are being identified from within-firm changes over time.

However, second, if the firm fixed effects estimates are controlling for genuine unobserved heterogeneity, then the sources identified in the literature using modern data have very little impact on explaining the relationship between R&D and innovation in the historical IRLUS data. There is no evidence in Figure 9 of any systematic differences in the relationship between R&D and patents, scaled citations or scaled generality by public versus private firms, by firms active in high versus low external finance dependent industries (see also Appendix Figure A1), or according to sub-groups of firms in these categories (see the bottom panels of Figure 9).

Third, the absence of statistically significant variation in the estimates by firm category suggests the baseline findings are stylized across all firms, and the results also support the hypothesis that different types of innovations scaled at different rates. Even if measurement error leads to attenuation bias in the fixed effects coefficients this should be symmetric across the patents, scaled citations and scaled generality regressions. When normalizing the estimates in Figure 9, there are economically large differences in the size of the coefficients across the outcome measures. The standardized coefficients (reported in Appendix Table A2) for scaled generality are about half the size of the patent and scaled citations coefficients for both public and private firms during both breakthrough eras. For external finance dependent firms the standardized coefficients for scaled generality are from a third to four-fifths the size of the patent and scaled citations coefficients. The findings imply novel technological discoveries as proxied by the scaled generality metric were generated in a much more scale-invariant way.

VII. Conclusion

This paper has attempted to provide new evidence on the relationship between scale and innovation during two central periods in the history of U.S. R&D with a view to understanding the nature of technological development during these periods more generally. Historically the interwar and post-WWII eras stand out because they helped define how corporate research laboratories functioned. Research activity moved increasingly within the boundaries of firms and the scale of resources devoted to innovation, both privately and by the federal government, increased significantly over time. R&D efforts helped to shape U.S. industrial leadership.

The first contribution of the paper has been to present the newly collected data. Macro trends in R&D are well-documented and understood but micro-level evidence is considerably more limited. According to Klette and Kortum (2004) R&D based growth theory should reflect empirical facts identified in firm-level studies, yet R&D data are particularly difficult to assemble historically. As such the IRLUS surveys present a major body of information, which can be linked to other sources of data such as patents and patent citations to create a repository of knowledge on firm-level R&D inputs and outputs. While there are some limitations to this source because it is based on direct correspondence surveys, it is broadly reflective of the firm size distribution covering the spectrum of large and small firms. It is the closest historical comparator to the U.S. Census Bureau's longitudinal business data for the modern era.

The second contribution of the paper has been to provide new evidence on the nature of firm-level technological change during these breakthrough eras. The period covered is an ideal setting for analyzing the relationship between scale and innovation because R&D firms expanded in size over time. This relationship has motivated a large empirical industrial organization literature in the Schumpeterian tradition and the scale effect property also helped to define a generation of endogenous growth theory models. Although the empirics presented here do not represent an explicit test of traditional endogenous growth theory, they are closely related to new endogenous theory approaches which have attempted to generate a much closer alignment between theory and micro-level empirics (e.g., Klette and Kortum, 2004; Akcigit and Kerr, 2010).

While much effort has gone into incorporating firm-level heterogeneity into endogenous growth frameworks, the current results emphasize a greater degree of heterogeneity at the technology level. Of course, one interpretation of this finding is that it may be specific to the historical context. Changes since the late twentieth century in the financing environment, or the proliferation of entrepreneurial startups may have a significant effect on the types of firms engaging in R&D (e.g., Acs and Audretsch 1987; Kortum and Lerner 2000; Baumol 2002). Such changes may imply a more central role for firm-level heterogeneity through entry and exit or reallocation across firms. It could also be that historically other fixed unobservables like the quality of research workers, or management talent may yield better insights into firm-level heterogeneity than the public versus private and external finance distinctions considered here.

Finally, the results have a broader implication. The emphasis on technological heterogeneity is consistent with several economic history based contributions that highlight its importance to growth. Mokyr's (1990) distinction between novel "macro" and incremental "micro" inventions, Gordon's (2012) analogous distinction between waves of discrete inventions followed by incremental improvements, and Squicciarini and Voigtländer's (2015) distinction between "average" and "upper-tail" knowledge underscore that growth is endogenous to the nature of technological discovery. For two of the most important epochs in the history of U.S. R&D the level and quality of innovation scaled strongly across the firm size distribution whereas novel innovation did not. Adding resources at the firm-level may have been associated with incremental improvements to existing product lines, but it did not lead to a proportional development in the production of radical ideas. Novel innovation scaled at about half the rate of normal discoveries. The absence of a strong scale effect helps to explain the inherent uncertainty typically associated with this form of technological change.

References

Acharya, Viral V. and Zhaoxia Xu. 2013. "Financial Dependence and Innovation: The Case of Public versus Private Firms." NBER Working Papers 19708 National Bureau of Economic Research.

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

Acs, Zoltan J. and David B. Audretsch. 1990. "Innovation and Small Firms." Cambridge, Mass.,: MIT Press.

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

Aghion, Philippe and Peter Howitt. 1992. "A Model of Growth through Creative Destruction." *Econometrica* 60(2):323–51.

Akcigit, Ufuk and William R. Kerr. 2010. "Growth Through Heterogeneous Innovations." Working Paper 16443 National Bureau of Economic Research.

Audretsch, David B. and Maryann P. Feldman. 2004. "Knowledge spillovers and the geography of innovation." In *Handbook of Regional and Urban Economics*, ed. J. V. Henderson and J. F. Thisse. 1 ed. Vol. 4 Elsevier chapter 61, pp. 2713–2739.

Baumol, William. 2002. *The Free-Market Innovation Machine: Analyzing the Growth. Miracle of Capitalism*. Princeton: Princeton University Press.

Cohen, Wesley M. 2010. Fifty Years of Empirical Studies of Innovative Activity and Performance. In *Handbook of the Economics of Innovation*, ed. Bronwyn Hall and Nathan Rosenberg. Amsterdam: Elsevier pp. 129–198.

Ferreira, Daniel, Gustavo Manso and Andre C. Silva. 2014. "Incentives to Innovate and the Decision to Go Public or Private." *Review of Financial Studies* 27(1):256–300.

Field, Alexander. 2003. "The Most Technologically Progressive Decade of the Twentieth Century." *American Economic Review* 93(4):1399–1413.

Furman, Jeffrey L. and Megan J. MacGarvie. 2007. "Academic science and the birth of industrial research laboratories in the U.S. pharmaceutical industry." *Journal of Economic Behavior and Organization* 63(4):756–776.

Furman, Jeffrey L. and Megan MacGarvie. 2009. "Academic collaboration and organizational innovation: the development of research capabilities in the US pharmaceutical industry, 1927–1946." *Industrial and Corporate Change* 18(5):929–961.

Gordon, Robert J., 2012. "Is U.S. Economic Growth Over? Faltering Innovation Confronts the Six Headwinds." Working Paper 18315 National Bureau of Economic Research.

Griliches, Zvi. 1979. "Issues in Assessing the Contribution of Research and Development to Productivity Growth." *Bell Journal of Economics* 10(1):92–116.

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

Grossman, G. M. and E. Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, Mass.: MIT Press.

Gurmu, Shiferaw and Fidel Prez-Sebastin. 2008. "Patents, R&D and Lag Effects: Evidence from Flexible Methods for Count Panel Data on Manufacturing Firms." *Empirical Economics* 35(3):507–526.

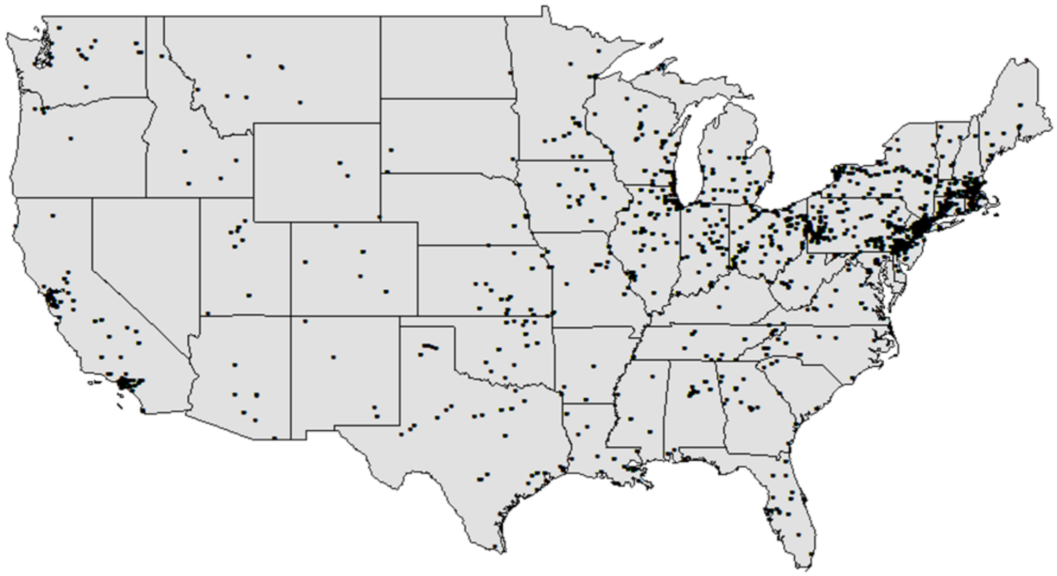
Hall, Bronwyn. 2005. "Note on the Bias in the Herfindahl Based on Count Data." *Revue d'Economie Industrielle* 110:149–156.

Hanlon, Walker. 2014. "Necessity is the Mother of Invention: Input Supplies and Directed

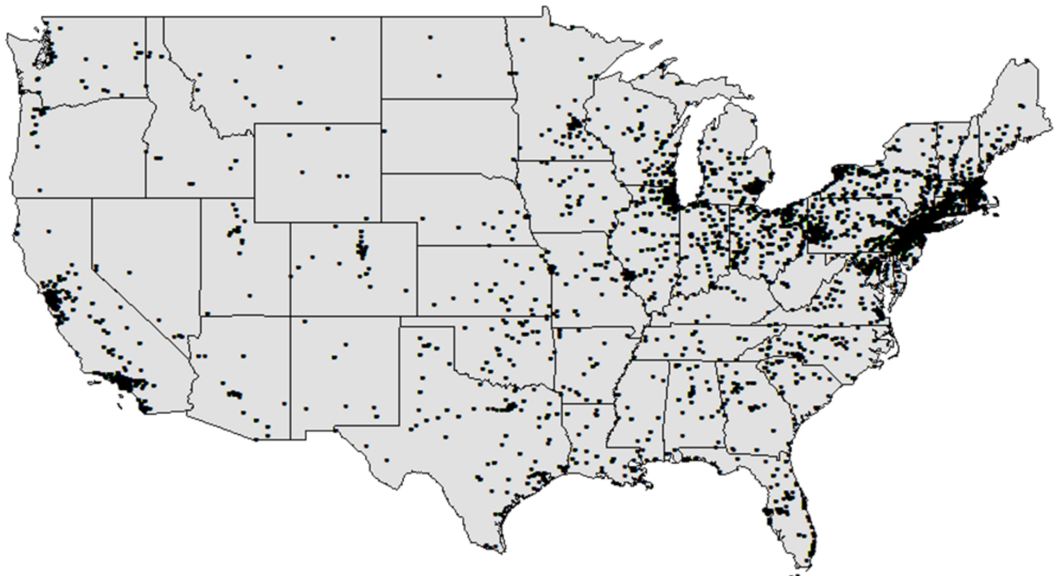
Technical Change." *Econometrica* forthcoming.
Hounshell, David and J. Smith. 1988. *Science and Corporate Strategy: Du Pont, 1902-1980*. Cambridge, Massachusetts: Cambridge University Press.
Jones, Charles I. 1995. "R&D-Based Models of Economic Growth." *Journal of Political Economy* 103(4):759–84.
Klette, Tor Jakob and Samuel Kortum. 2004. "Innovating Firms and Aggregate Innovation." *Journal of Political Economy* 112(5):986–1018.
Kogan, Leonid, Dimitris Papanikolaou, Amit Seru and Noah Stoffman. 2012. "Technological Innovation, Resource Allocation, and Growth." NBER Working Paper.
Kortum, Samuel S. 1997. "Research, Patenting, and Technological Change." *Econometrica* 65(6):1389–1420.
Kortum, Samuel and Josh Lerner. 2000. "Assessing the Contribution of Venture Capital to Innovation," *Rand Journal of Economics* 31(4): 674–692.
Mokyr, Joel. 1990. *The Lever of Riches*. Oxford: Oxford University Press.
Mowery, David. 1981. "The Emergence and Growth of Industrial Research in American Manufacturing, 1899-1946." Ph.D. Dissertation, Stanford University.
Mowery, David. 2005. "Development of Industrial Research in U.S. Manufacturing." *American Economic Review* 80(2):345–349.
Mowery, David and Nathan Rosenberg. 1989. *Technology and the Pursuit of Economic Growth*. Cambridge, Massachusetts: Cambridge University Press.
Nanda, Ramana and Tom Nicholas. 2014. "Did Bank Distress Stifle Innovation During the Great Depression?" *Journal of Financial Economics* 114(2):273–292.
National Research Council. 1921-1970. *Industrial Research Laboratories of the United States*. Washington D.C.: National Research Council.
Nelson, Richard and Gavin Wright. 1992. "The Rise and Fall of American Technological Leadership: The Postwar Era in Historical Perspective." *Journal of Economic Literature* 30(4):1931–1964.
Rajan, Raghuram G and Luigi Zingales. 1998. "Financial Dependence and Growth." *American Economic Review* 88(3):559–86.
Romer, Paul. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98(5):71–102.
Scherer, F. M. 1965. "Corporate Inventive Output, Profits, and Growth." *Journal of Political Economy* 73:290.
Scherer, F. M. 1979. "Size of Firm, Oligopoly, and Research: A Comment." *The Canadian Journal of Economics and Political Science* 31(2):256–266.
Squicciarini, Mara P. and Nico Voigtländer. 2015. "Human Capital and Industrialization: Evidence from the Age of Enlightenment," *Quarterly Journal of Economics, forthcoming*.
Stephan, Paula. 2014. "The Endless Frontier: Reaping What Bush Sowed?" In *The Changing Frontier: Rethinking Science and Innovation Policy*, ed. Adam Jaffe and Benjamin Jones. Chicago: University of Chicago Press pp. 321–370.
Temin, Peter. 1989. *The Fall of the Bell System A Study in Prices and Politics*. Cambridge, England: Cambridge University Press.

Figure 1. Geographic Location of R&D Labs

Interwar & WWII

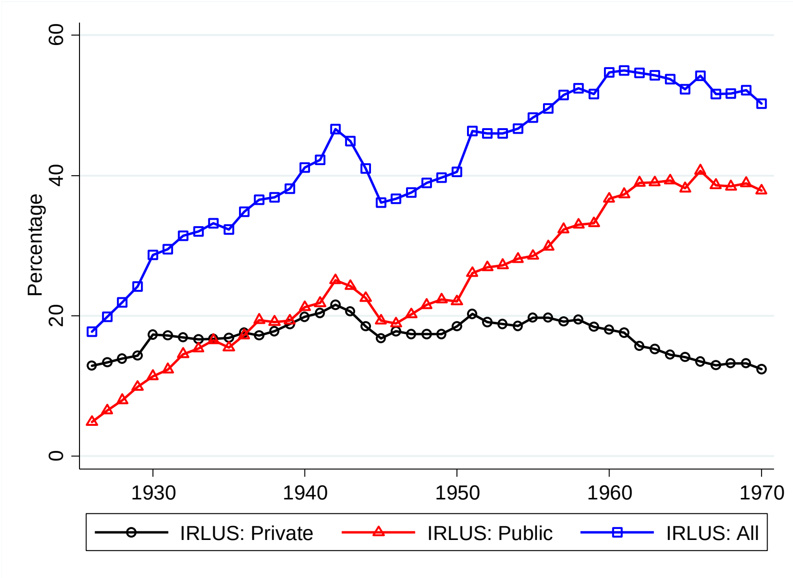


Post-WWII



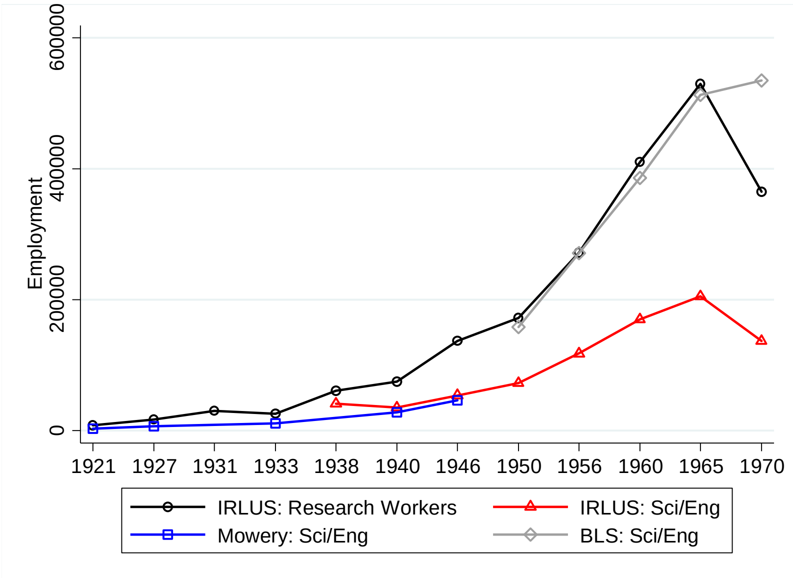
Notes: Each observation represents an R&D lab location for a firm active during the respective periods. Geocoding on the basis of addresses listed in the IRLUS volumes.

Figure 2. IRLUS Firm Share of Total U.S. Patents



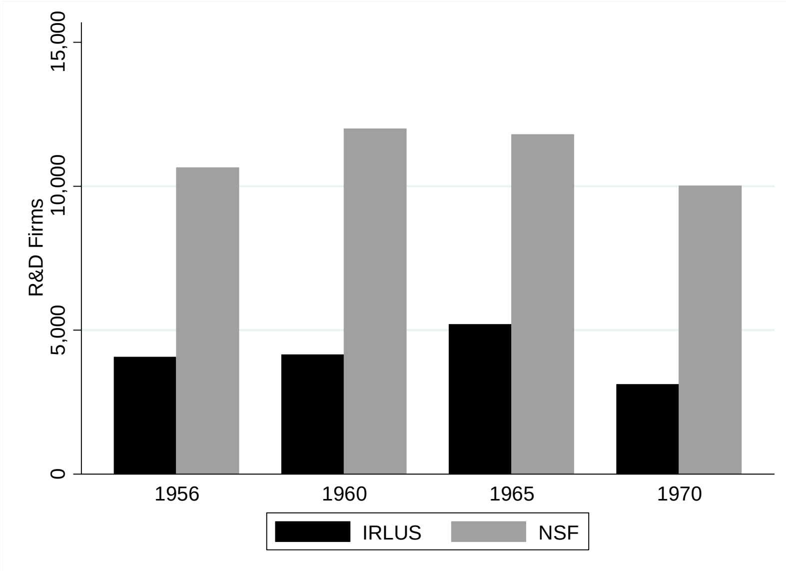
Notes: IRLUS data from a match of firms in IRLUS with U.S. patents. Figures expressed as a percentage of total U.S. patents by U.S. inventors (i.e., excluding patents granted in the U.S. to foreign domiciled inventors).

Figure 3. Benchmarking Research Employment in the IRLUS Surveys



Notes: Mowery data are also from IRLUS volumes and reported in Mowery (1981). The U.S. Bureau of Labor Statistics series is reported in *Statistical Abstract of the United States* series W178.

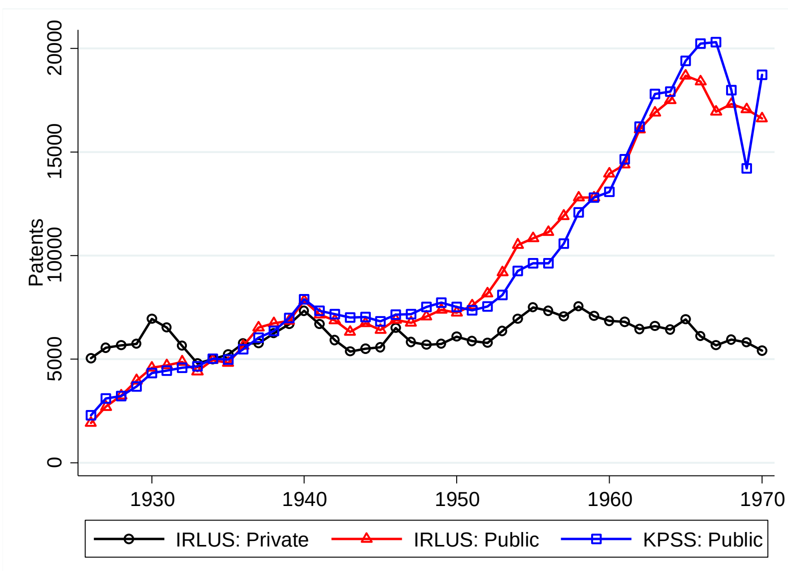
Figure 4. Benchmarking the Number of Firms in the IRLUS Surveys



Notes: The NSF data were taken from the *NSF Survey of Industry Research and Development*:

- The estimate for 1956 is from 1953/54 and is for all firms conducting R&D in their own facilities in industries matching to the ones I observe in the IRLUS volumes.
- The estimate for 1960 is from 1962 and is for all firms conducting R&D in manufacturing.
- The estimate for 1965 is from 1965 and is for all firms conducting R&D in manufacturing.
- The estimate for 1970 is from 1970 and is for all firms conducting R&D in manufacturing.

Figure 5. Benchmarking Patents in the IRLUS Surveys

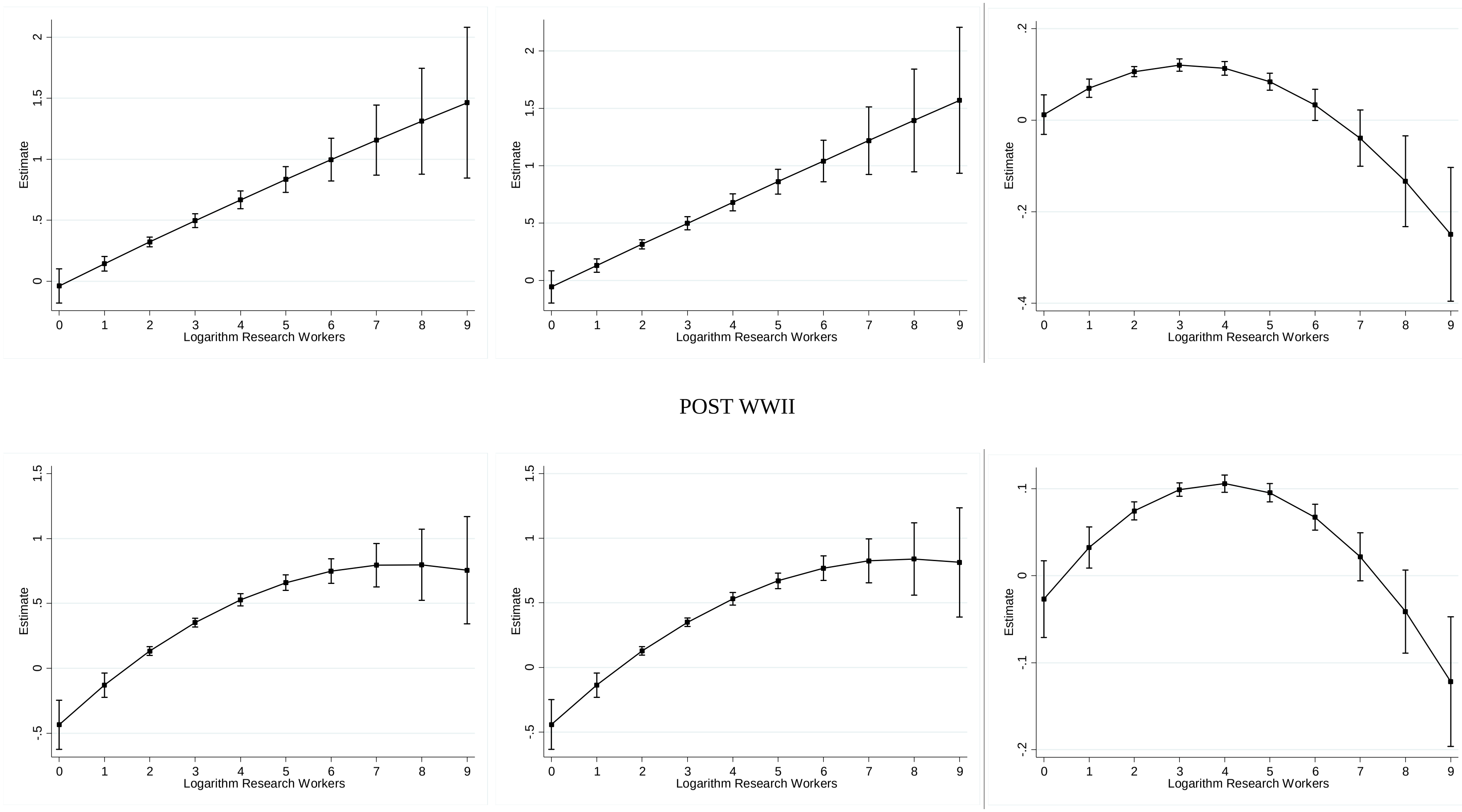


Notes: IRLUS data from a match of firms in IRLUS with U.S. patents. KPSS data are from the match of CRSP firms and U.S. patents reported in Kogan et. al., (2012).

Figure 6. Average Marginal Effects of Research Workers with Industry Fixed Effects

PATENTS SCALED CITATIONS SCALED GENERALITY

INTERWAR & WWII

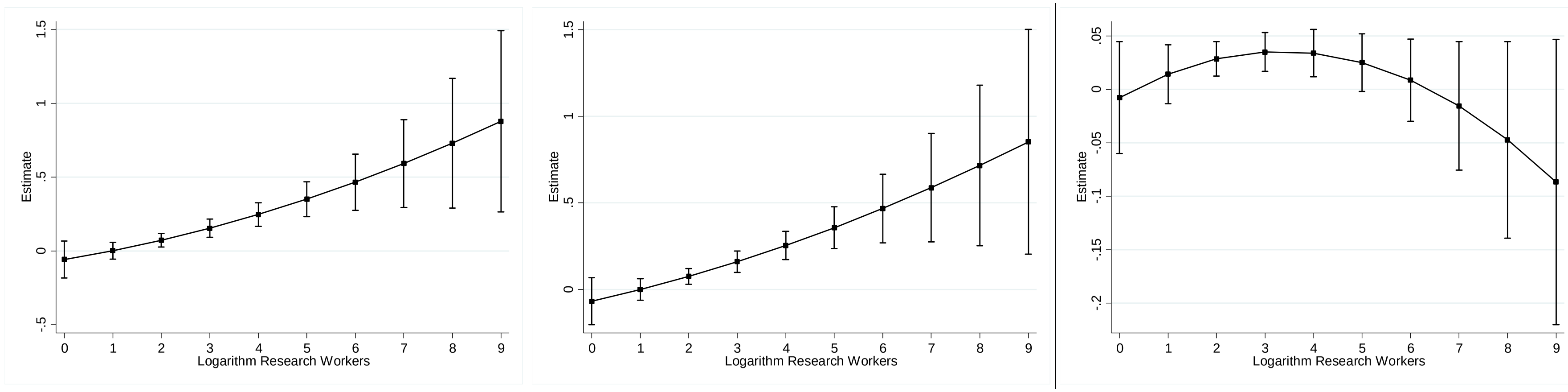


Notes: The point estimates and 95 percent confidence intervals are from specifications of equation (1) using linear, squared and cubic terms for the logarithm of research workers as the independent variables. One added to rescale zero values of the logarithmic dependent variables. The plots show the average marginal effect of research workers on the innovation outcome measures over a range of research workers values in the data. Estimates are for patenting firms that patented at least once during the time period.

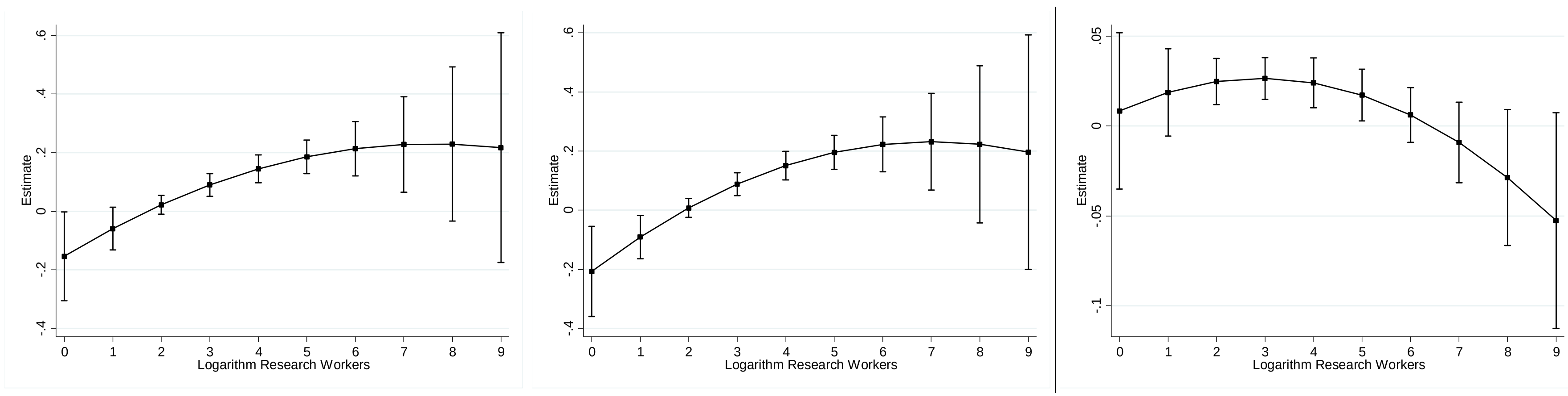
Figure 7. Average Marginal Effects of Research Workers with Firm Fixed Effects

PATENTS | SCALED CITATIONS | SCALED GENERALITY

INTERWAR & WWII



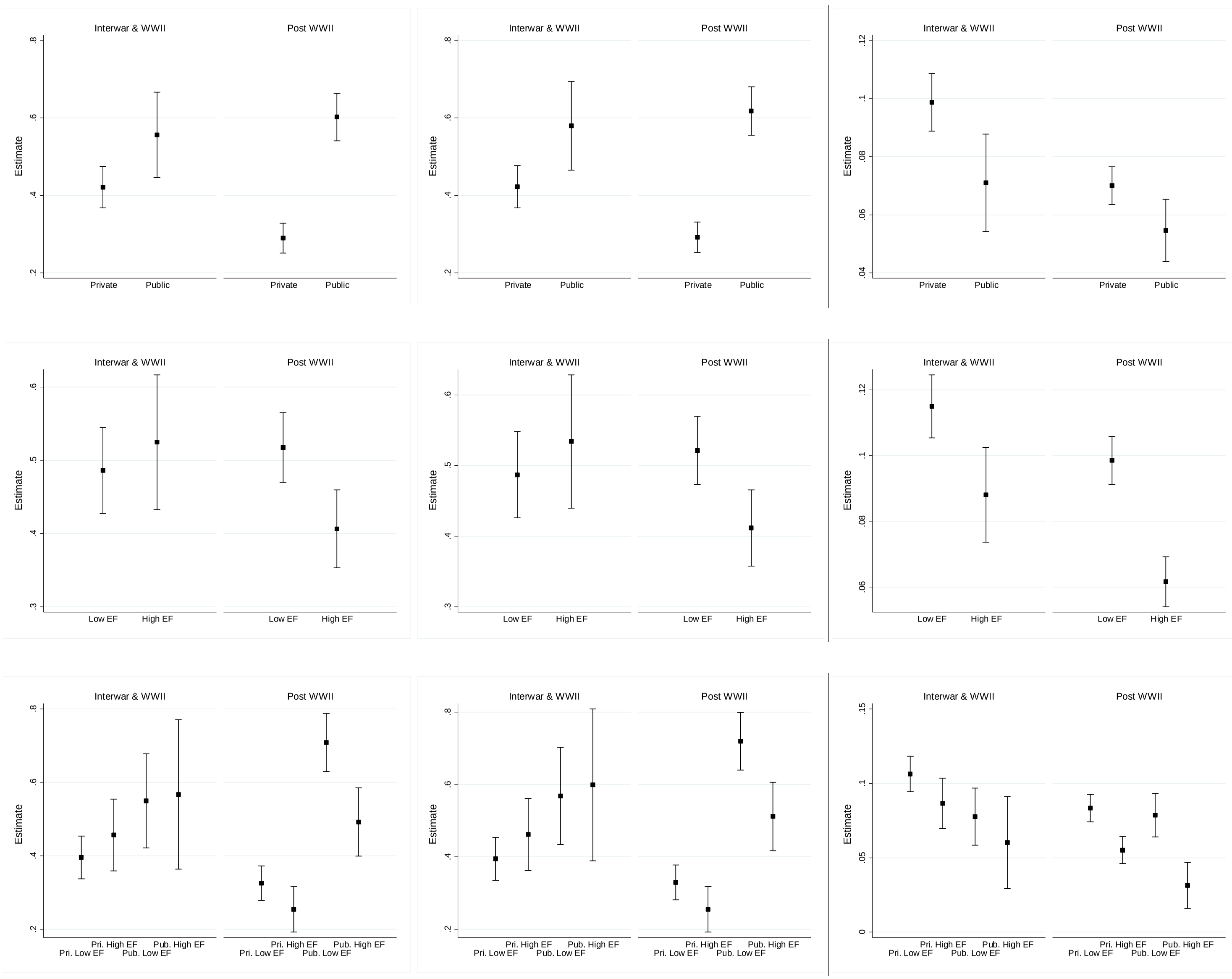
POST WWII



Notes: The point estimates and 95 percent confidence intervals are from specifications of equation (1) using linear, squared and cubic terms for the logarithm of research workers as the independent variables. One added to rescale zero values of the logarithmic dependent variables. The plots show the average marginal effect of research workers on the innovation outcome measures over a range of research workers values in the data. Estimates are for patenting firms that patented at least once during the time period.

Figure 8. Estimates for Public, Private and External Finance Dependence Firms: Industry Fixed Effects

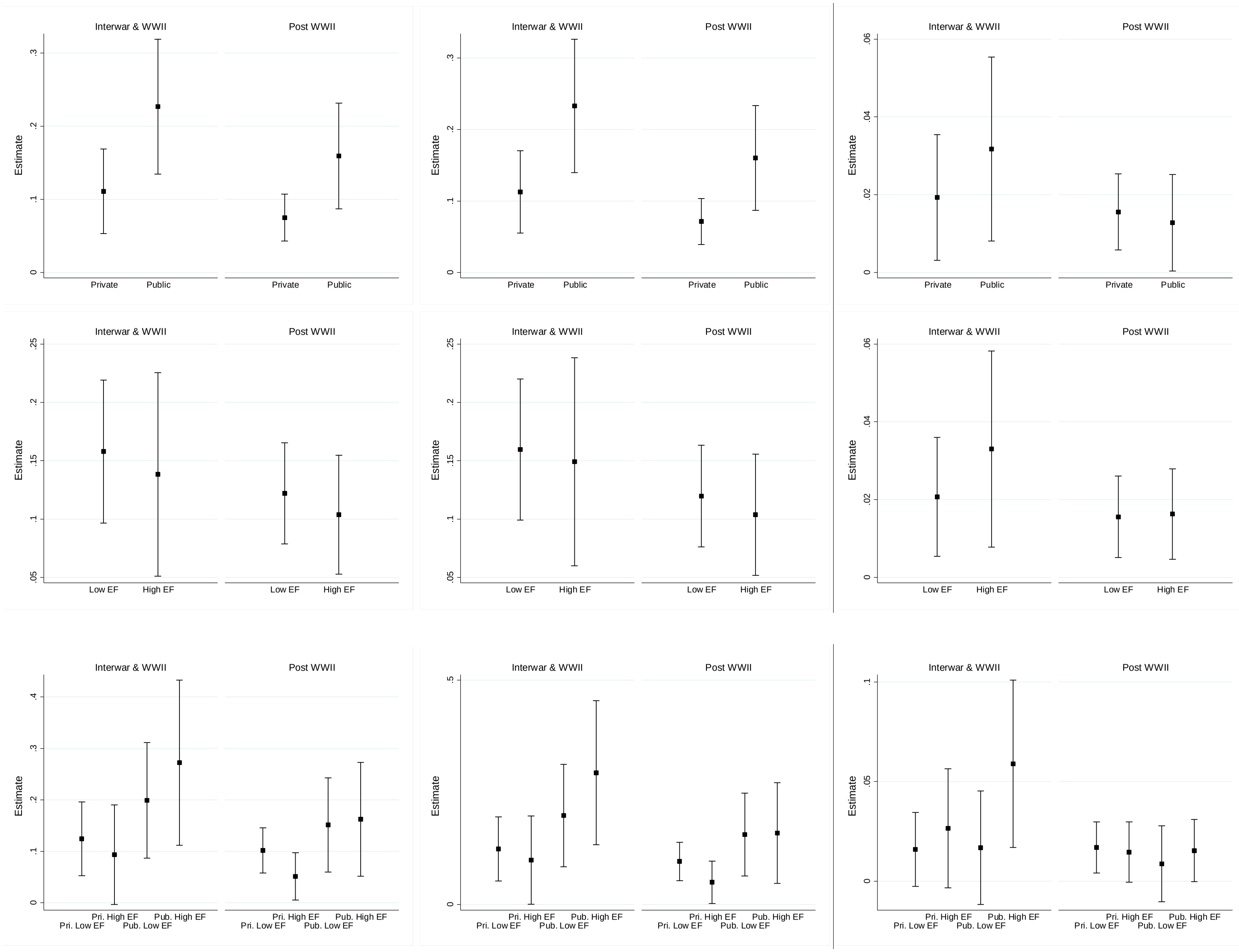
PATENTS SCALED CITATIONS SCALED GENERALITY



Notes: The point estimates and 95 percent confidence intervals are from specifications of equation (1) run on sub-samples of firms with the logarithm of research workers as the independent variable. One added to rescale zero values of the logarithmic dependent variables. Standard errors clustered by firm. Estimates are for patenting firms that patented at least once during the time period.

Figure 9. Estimates for Public, Private and External Finance Dependence Firms: Firm Fixed Effects

PATENTS SCALED CITATIONS SCALED GENERALITY



Notes: The point estimates and 95 percent confidence intervals are from specifications of equation (1) run on sub-samples of firms with the logarithm of research workers as the independent variable. One added to rescale zero values of the logarithmic dependent variables. Standard errors clustered by firm. Estimates are for patenting firms that patented at least once during the time period.

Table 1. Descriptive Statistics

	Interwar and WWII		Post-WWII	
	Patenting Firms	All Firms	Patenting Firms	All Firms
Patents	4.34	3.09	4.03	2.32
	(24.38)	(20.68)	(25.09)	(19.16)
Citations	19.53	13.93	27.36	15.78
	(133.24)	(112.85)	(180.93)	(138.07)
Scaled Citations	4.48	3.19	4.26	2.45
	(26.49)	(22.46)	(27.73)	(21.16)
Generality	0.15	0.11	0.15	0.09
	(0.25)	(0.23)	(0.26)	(0.21)
Scaled Generality	0.43	0.30	0.36	0.21
	(0.81)	(0.71)	(0.64)	(0.52)
Research Workers[a]	29.71	25.38	93.58	84.00
	(128.57)	(114.58)	(487.13)	(455.93)
Research Workers[b]	39.69	34.53	104.29	94.43
	(171.62)	(156.18)	(503.68)	(479.86)
Scientific Workers[a]	26.56	23.34	38.46	33.67
	(93.32)	(85.36)	(213.30)	(186.35)
Scientific Workers[b]	24.99	22.12	43.65	38.49
	(87.40)	(80.23)	(230.89)	(206.27)

Notes: Means with standard deviations in parentheses for firms in the IRLUS volumes published in 1921, 1927, 1931, 1933, 1938, 1940, 1946, 1950, 1956, 1960, 1965 and 1970. The superscripts “a” and “b” for the research and scientific worker variables refer to measured values as of the 12 survey years and linearly imputed values for all years respectively.

Table 2. The Relationship Between Innovation and Research Workers

	Panel A: Patents									
	Interwar & WWII					Post-WWII				
	Patenting Firms			All Firms		Patenting Firms			All Firms	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log (Research Workers)	0.515***	0.502***	0.150***	0.473***	0.130***	0.467***	0.461***	0.113***	0.373***	0.093***
	[0.027]	[0.026]	[0.026]	[0.024]	[0.023]	[0.018]	[0.018]	[0.017]	[0.016]	[0.014]
Standardized Coefficient	*0.532*	*0.518*	*0.154*	*0.513*	*0.141*	*0.536*	*0.529*	*0.129*	*0.485*	*0.121*
Observations	25,570	25,570	25,570	31,777	31,777	53,861	53,861	53,861	75,245	75,245
Clusters (firms)	2,340	2,340	2,340	3,261	3,261	5,561	5,561	5,561	10,254	10,254
R^2	0.27	0.30	0.77	0.30	0.79	0.29	0.29	0.78	0.24	0.80
Industry FE	N	Y	N	Y	N	N	Y	N	Y	N
Firm FE	N	N	Y	N	Y	N	N	Y	N	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

	Panel B: Scaled Citations									
	Interwar & WWII					Post-WWII				
	Patenting Firms			All Firms		Patenting Firms			All Firms	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log (Research Workers)	0.519***	0.506***	0.155***	0.474***	0.135***	0.472***	0.466***	0.111***	0.375***	0.092***
	[0.027]	[0.027]	[0.026]	[0.025]	[0.023]	[0.019]	[0.019]	[0.017]	[0.016]	[0.014]
Standardized Coefficient	*0.523*	*0.510*	*0.156*	*0.505*	*0.144*	*0.529*	*0.522*	*0.125*	*0.479*	*0.117*
Observations	25,570	25,570	25,570	31,777	31,777	53,861	53,861	53,861	75,245	75,245
Clusters (firms)	2,340	2,340	2,340	3,261	3,261	5,561	5,561	5,561	10,254	10,254
R^2	0.26	0.29	0.74	0.29	0.76	0.28	0.28	0.75	0.24	0.77
Industry FE	N	Y	N	Y	N	N	Y	N	Y	N
Firm FE	N	N	Y	N	Y	N	N	Y	N	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

	Panel C: Scaled Generality									
	Interwar & WWII					Post-WWII				
	Patenting Firms			All Firms		Patenting Firms			All Firms	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log (Research Workers)	0.102***	0.099***	0.028***	0.105***	0.024***	0.087***	0.087***	0.020***	0.081***	0.016***
	[0.004]	[0.005]	[0.008]	[0.004]	[0.007]	[0.003]	[0.003]	[0.005]	[0.003]	[0.004]
Standardized Coefficient	*0.296*	*0.287*	*0.081*	*0.318*	*0.074*	*0.288*	*0.289*	*0.066*	*0.300*	*0.059*
Observations	25,570	25,570	25,570	31,777	31,777	53,861	53,861	53,861	75,245	75,245
Clusters (firms)	2,340	2,340	2,340	3,261	3,261	5,561	5,561	5,561	10,254	10,254
R^2	0.08	0.10	0.35	0.13	0.41	0.08	0.09	0.36	0.10	0.43
Industry FE	N	Y	N	Y	N	N	Y	N	Y	N
Firm FE	N	N	Y	N	Y	N	N	Y	N	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The dependent variables are patents (panel A), scaled citations (panel B) and scaled generality (panel C) in logarithms with one added to rescale zero values. Patenting firms are those that patented at least once during the time period. All firms include those firms that did not patent. Standardized coefficients measure the effect of a one standard deviation change. Standard errors in parentheses are clustered by firm: ***p < 0.01, **p < 0.05, *p < 0.1.

Table 3. Testing for a Lagged Relationship

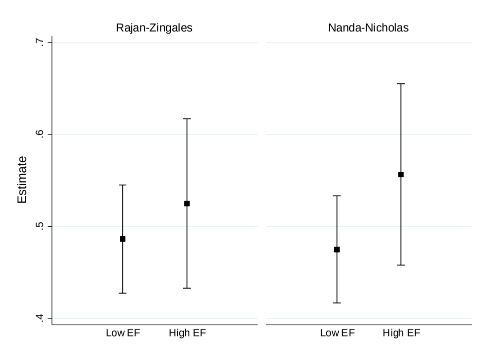
	Panel A: Patents				Panel B: Scaled Citations				Panel C: Scaled Generality			
	Interwar & WWII		Post-WWII		Interwar & WWII		Post-WWII		Interwar & WWII		Post-WWII	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log (Research Workers)	0.564***	0.152***	0.453***	0.115***	0.555***	0.140***	0.456***	0.112***	0.111***	0.009	0.105***	0.028**
	[0.054]	[0.041]	[0.041]	[0.023]	[0.056]	[0.045]	[0.042]	[0.024]	[0.018]	[0.019]	[0.013]	[0.011]
Log (Research Workers)$_{t-1}$	-0.228***	-0.040	-0.056	-0.013	-0.200***	-0.017	-0.059	-0.018	-0.014	0.029	-0.015	-0.009
	[0.058]	[0.048]	[0.036]	[0.023]	[0.062]	[0.055]	[0.038]	[0.025]	[0.028]	[0.030]	[0.015]	[0.015]
Log (Research Workers)$_{t-2}$	0.095**	0.035	-0.096***	-0.002	0.069	0.018	-0.095***	0.000	-0.011	-0.018	-0.019	-0.001
	[0.042]	[0.043]	[0.023]	[0.021]	[0.047]	[0.049]	[0.025]	[0.022]	[0.025]	[0.027]	[0.014]	[0.014]
Log (Research Workers)$_{t-3}$	-0.089*	-0.062	-0.061*	-0.006	-0.043	-0.022	-0.042	0.017	0.004	0.012	-0.008	0.004
	[0.053]	[0.040]	[0.033]	[0.023]	[0.057]	[0.043]	[0.035]	[0.027]	[0.022]	[0.022]	[0.013]	[0.014]
Log (Research Workers)$_{t-4}$	0.211***	0.061**	0.269***	0.020	0.176***	0.035	0.255***	0.004	0.011	-0.010	0.027**	-0.017*
	[0.050]	[0.030]	[0.039]	[0.025]	[0.052]	[0.030]	[0.040]	[0.027]	[0.015]	[0.014]	[0.010]	[0.010]
Sum of coefficents	*0.553*	*0.145*	*0.509*	*0.115*	*0.556*	*0.154*	*0.514*	*0.116*	*0.102*	*0.022*	*0.091*	*0.006*
	(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.050)*	*(0.000)*	*(0.381)*
Observations	17,185	17,185	42,942	42,942	17,185	17,185	42,942	42,942	17,185	17,185	42,942	42,942
Clusters (firms)	1,929	1,929	3,950	3,950	1,929	1,929	3,950	3,950	1,929	1,929	3,950	3,950
R^2	0.33	0.81	0.32	0.80	0.32	0.78	0.31	0.77	0.11	0.39	0.10	0.39
Industry FE	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
Firm FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

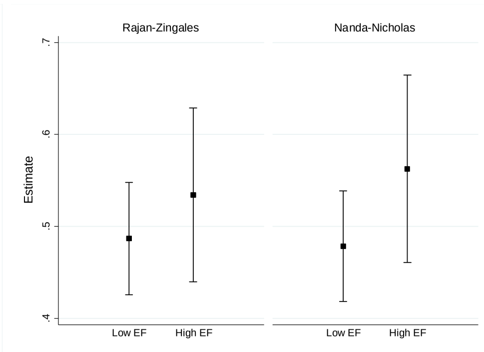
Notes: The dependent variables are patents, scaled citations and scaled generality in logarithms with one added to rescale zero values. Estimates are for patenting firms that patented at least once during the time period. Standard errors in parentheses are clustered by firm: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

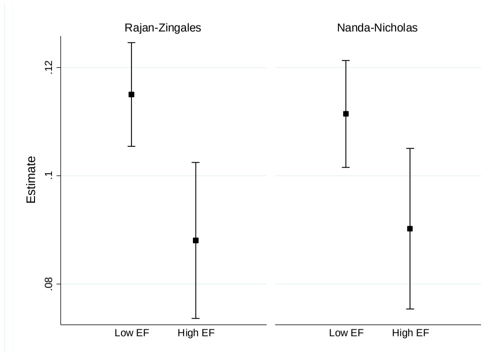
Appendix. Figure A1. Estimates for External Finance Dependence Firms for the Interwar & WWII Era

PATENTS | SCALED CITATIONS | SCALED GENERALITY

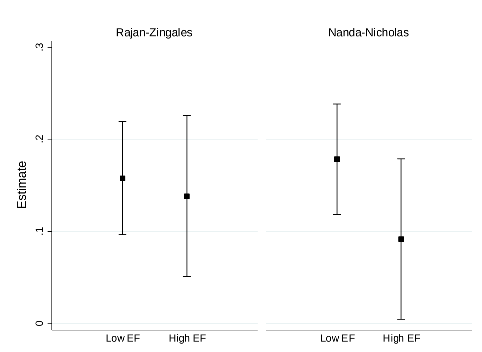
INDUSTRY FIXED EFFECTS

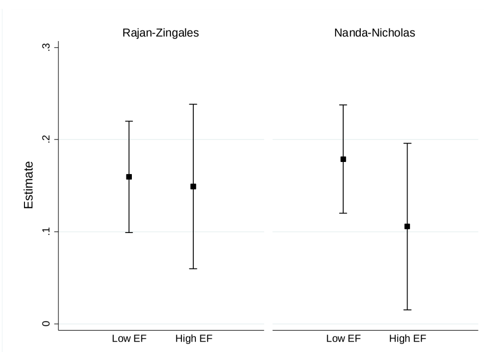


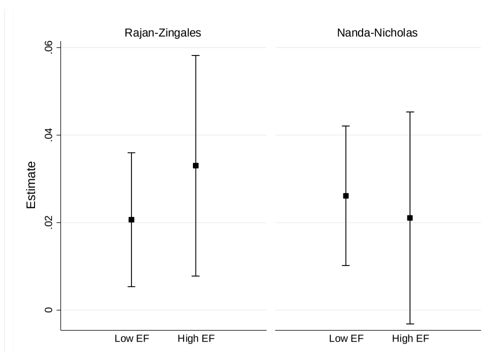




FIRM FIXED EFFECTS







Notes: These figures replicate the results in the middle panels of Figures 8 and 9 using a different measure of financial dependence. Rajan and Zingales (1998) calculate the level of dependence for U.S. firms in the 1980s using capital expenditures minus cash flow from operations over capital expenditures. Nanda and Nicholas (2014) calculate bank notes payable over fixed assets using financials for all interwar firms listed in the IRLUS volumes that were also listed in *Moody's Manual of Industrials*. One added to rescale zero values of the logarithmic dependent variables. Point estimates and 95 percent confidence intervals from specifications run on sub-samples of firms with the logarithm of research workers as the independent variable. Standard errors clustered by firm.

Appendix. Table A1. The Relationship Between Innovation and Scientific Research Workers

	Panel A: Patents									
	Interwar & WWII					Post-WWII				
	Patenting Firms			All Firms		Patenting Firms			All Firms	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log (Scientific Workers)	0.565***	0.555***	0.102***	0.522***	0.082***	0.506***	0.500***	0.115***	0.408***	0.095***
	[0.031]	[0.031]	[0.029]	[0.028]	[0.025]	[0.020]	[0.020]	[0.017]	[0.017]	[0.014]
Standardized Coefficient	*0.527*	*0.518*	*0.095*	*0.517*	*0.081*	*0.528*	*0.521*	*0.120*	*0.479*	*0.111*
Observations	10,894	10,894	10,894	13,632	13,632	53,215	53,215	53,215	74,491	74,491
Clusters (firms)	1,823	1,823	1,823	2,369	2,369	5,549	5,549	5,549	10,224	10,224
R^2	0.26	0.30	0.83	0.29	0.85	0.28	0.29	0.78	0.24	0.80
Industry FE	N	Y	N	Y	N	N	Y	N	Y	N
Firm FE	N	N	Y	N	Y	N	N	Y	N	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

	Panel B: Scaled Citations									
	Interwar & WWII					Post-WWII				
	Patenting Firms			All Firms		Patenting Firms			All Firms	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log (Scientific Workers)	0.567***	0.558***	0.107***	0.522***	0.087***	0.511***	0.505***	0.112***	0.410***	0.092***
	[0.032]	[0.032]	[0.033]	[0.029]	[0.028]	[0.020]	[0.020]	[0.017]	[0.017]	[0.015]
Standardized Coefficient	*0.518*	*0.510*	*0.098*	*0.508*	*0.085*	*0.520*	*0.514*	*0.114*	*0.473*	*0.107*
Observations	10,894	10,894	10,894	13,632	13,632	53,215	53,215	53,215	74,491	74,491
Clusters (firms)	1,823	1,823	1,823	2,369	2,369	5,549	5,549	5,549	10,224	10,224
R^2	0.25	0.28	0.80	0.28	0.82	0.27	0.28	0.75	0.24	0.77
Industry FE	N	Y	N	Y	N	N	Y	N	Y	N
Firm FE	N	N	Y	N	Y	N	N	Y	N	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

	Panel C: Scaled Generality									
	Interwar & WWII					Post-WWII				
	Patenting Firms			All Firms		Patenting Firms			All Firms	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log (Scientific Workers)	0.117***	0.117***	0.028**	0.122***	0.023**	0.095***	0.096***	0.021***	0.089***	0.017***
	[0.006]	[0.006]	[0.012]	[0.006]	[0.010]	[0.003]	[0.003]	[0.005]	[0.003]	[0.004]
Standardized Coefficient	*0.314*	*0.313*	*0.076*	*0.346*	*0.064*	*0.285*	*0.288*	*0.065*	*0.300*	*0.058*
Observations	10,894	10,894	10,894	13,632	13,632	53,215	53,215	53,215	74,491	74,491
Clusters (firms)	1,823	1,823	1,823	2,369	2,369	5,549	5,549	5,549	10,224	10,224
R^2	0.10	0.11	0.43	0.14	0.48	0.08	0.09	0.36	0.10	0.43
Industry FE	N	Y	N	Y	N	N	Y	N	Y	N
Firm FE	N	N	Y	N	Y	N	N	Y	N	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: This table replicates the results in Table 2 using scientific workers reported in the IRLUS surveys instead of all research workers. The number of observations is smaller because breakdowns of the research worker totals into scientific research workers and other workers were not reported in all cases. The dependent variables are patents (panel A), scaled citations (panel B) and scaled generality (panel C) in logarithms with one added to rescale zero values. Patenting firms are those that patented at least once during the time period. All firms include those firms that did not patent. Standardized coefficients measure the effect of a one standard deviation change. Standard errors in parentheses are clustered by firm: ***p < 0.01, **p < 0.05, *p < 0.1.

Appendix. Table A2. Estimates for Public, Private and External Finance Dependence Firms: Firm Fixed Effects

	Panel A: Public and Private Firms											
	Patents				Scaled Citations				Scaled Generality			
	Interwar & WWII		Post-WWII		Interwar & WWII		Post-WWII		Interwar & WWII		Post-WWII	
	Private	Public	Private	Public	Private	Public	Private	Public	Private	Public	Private	Public
Log (Research Workers)	0.111***	0.227***	0.075***	0.159***	0.113***	0.233***	0.071***	0.160***	0.024**	0.033**	0.020***	0.016**
	[0.030]	[0.047]	[0.016]	[0.037]	[0.029]	[0.048]	[0.016]	[0.037]	[0.010]	[0.013]	[0.006]	[0.007]
Standardized Coefficient	*0.120*	*0.211*	*0.100*	*0.149*	*0.119*	*0.210*	*0.092*	*0.146*	*0.063*	*0.114*	*0.059*	*0.060*
Observations	21,014	4,556	42,062	11,799	21,014	4,556	42,062	11,799	21,014	4,556	42,062	11,799
Clusters (firms)	2,108	378	4,831	1,063	2,108	378	4,831	1,063	2,108	378	4,831	1,063
R^2	0.74	0.82	0.70	0.81	0.69	0.80	0.66	0.79	0.34	0.32	0.32	0.38
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

	Panel B: External Finance Dependence											
	Patents				Scaled Citations				Scaled Generality			
	Interwar & WWII		Post-WWII		Interwar & WWII		Post-WWII		Interwar & WWII		Post-WWII	
	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
Log (Research Workers)	0.158***	0.138***	0.122***	0.104***	0.160***	0.149***	0.120***	0.104***	0.023**	0.038***	0.020***	0.020***
	[0.031]	[0.045]	[0.022]	[0.026]	[0.031]	[0.046]	[0.022]	[0.027]	[0.009]	[0.014]	[0.007]	[0.007]
Standardized Coefficient	*0.171*	*0.136*	*0.135*	*0.124*	*0.169*	*0.143*	*0.129*	*0.121*	*0.065*	*0.116*	*0.061*	*0.072*
Observations	16,528	9,042	30,176	23,685	16,528	9,042	30,176	23,685	16,528	9,042	30,176	23,685
Clusters (firms)	1,467	873	2,700	2,861	1,467	873	2,700	2,861	1,467	873	2,700	2,861
R^2	0.75	0.79	0.79	0.77	0.71	0.76	0.76	0.74	0.35	0.35	0.37	0.36
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: This table reports coefficients for the top and middle panels of Figure 9. The dependent variables are patents, scaled citations and scaled generality in logarithms with one added to rescale zero values. Estimates are for patenting firms that patented at least once during the time period. Standardized coefficients measure the effect of a one standard deviation change. Standard errors in parentheses are clustered by firm: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.