The Development of the American Entrepreneurial Economy

Innessa Colaiacovo*
Harvard Business School

Daniel P. Gross†
Duke University and NBER

Jorge A. Guzman ‡
Columbia University

March 2022

Draft v1.0 – comments are welcome

As to whatever may depend on enterprise, we need not fear to be outdone by any people on earth. It may almost be said that enterprise is our element.

—Alexander Hamilton, 1795

Abstract:

We use new data on firm creation to study U.S. entrepreneurship since 1900. New firm creation grew slowly before World War II but flourished after, growing more than twice as fast as GDP over the next 70 years. This break was accompanied by a structural transformation in the industrial composition of new firms, as manufacturing and mineral extraction declined, construction and services grew, and entrepreneurship became increasingly connected to technological change. These latter entrepreneurs were typically users and diffusers of technology, not creators. We embed this insight in an endogenous growth model to articulate a Schumpeterian view of how "Main Street" entrepreneurs transform innovation into productivity growth.

JEL Classification: M13, O31, O41, O51, N52, N62, N72, N82

Keywords: Entrepreneurship, Economic History, Innovation, Technological Change, Structural Change, Endogenous Growth, World War II

*Address: Harvard Business School, Boston, MA 02163, USA; email: icolaiacovo@hbs.edu.
†Address: Duke University Fuqua School of Business, Durham, NC 27708, USA; email: daniel.gross@duke.edu.
‡Address: Columbia Business School, New York, NY, USA; email: jag2367@columbia.edu.

1 Introduction

Entrepreneurship is both critical to the modern American economy and foundational to the American identity (Hamilton 1795). New firms today are responsible for about 20 percent of gross job creation (Haltiwanger et al. 2013, Decker et al. 2014) and play an out-sized role in developing and commercializing new technologies (Kortum and Lerner 2000). Entrepreneurs are vaunted by politicians on both sides of the aisle, and an object of fascination in the press, in business education, and in popular culture. Many of the most celebrated managers in history were founders themselves, from Andrew Carnegie and Henry Ford to Sam Walton and Steve Jobs.

Despite the importance of entrepreneurship, empirical research has mostly studied the post-1980 era, when modern datasets like the Census Bureau's Longitudinal Business Database (LBD) begin. As a result, we lack a comprehensive understanding of the long-run development of American entrepreneurship. Though researchers have produced a mosaic of important facts about historical entrepreneurship in specific regions (Cull et al. 2006) and industries (Lamoreaux and Sokoloff 2000, Lamoreaux et al. 2004, Atack and Margo 2019), and studied technological entrepreneurship via patents (Khan and Sokoloff 1993, Lamoreaux and Sokoloff 2005, Nicholas 2003, 2010, Babina et al. 2020), prior work has left broader questions unanswered, including the most basic question: has the rate of new firm creation increased or decreased across U.S. history?

In this paper, we present new evidence on how U.S. entrepreneurship has changed over the past century and new insights into its Schumpeterian dynamics, which characterize an important subset of the U.S. entrepreneurial economy (Aghion and Howitt 1992). To do so, we introduce novel, comprehensive data on historical U.S. entrepreneurship in the form of business registrations (Andrews et al. 2022), which measure the creation of new legal entities directly from administrative records. For each firm, these data provide the firm's name, physical state, registration state, and registration year. Motivated by research highlighting the information content in firm names (Tadelis 1999, McDevitt 2014, Belenzon et al. 2017), we develop a new procedure for assigning these firms to industry categories based on words in their names. We then use firm names to trace the diffusion of innovation through the entrepreneurial economy. Though inevitably subject to limitations, these new approaches to measurement allow us to bring new insight to long-standing questions in the economics of entrepreneurship, innovation, and economic growth.

We begin by documenting long-run patterns in aggregate entrepreneurship. Our evidence indicates that entrepreneurial dynamism is *not* a fundamental constant of U.S. history. Rather, persistent long-run growth in entrepreneurship is a distinguishing feature of the postwar era. Prior to the

Depression-era 1930s, business registrations per capita grew slowly, matching the pace of per capita real GDP growth. Since the end of World War II, business registrations have grown at a remarkably steady 5.5% per year, more than doubling the average GDP growth rate over the same period. This stylized fact challenges views that entrepreneurship is a permanent feature of the American system. To the contrary, World War II seemingly ushered the U.S. economy from its Chandlerian era of big business (Chandler 1977, 1990) into an era of small business.

We next study how the composition of new firms has changed over time. We first use a complementary sample of firms from Dun & Bradstreet, where World War II again partitions the twentieth century into two periods vis-à-vis the industrial composition of new businesses. We then train and validate a procedure that classifies firms in our administrative registration data to economic sectors on the basis of their names, and use these measures to study how entrepreneurship has reallocated across sectors over time. Consistent with traditional perspectives of structural transformation (e.g., Kuznets 1973, Herrendorf et al. 2014), we document how the *Manufacturing* and *Mining* sectors' shares of new firms declined, and the *Services* sector grew, as the U.S. evolved from an extractive, to industrial, to post-industrial economy. Surges in sector-specific demand, such as a manufacturing boom in World War II or an oil & gas boom in the 1970s, interrupted but did not dislodge these longer-term trends. Nevertheless, a consistent pattern across *all* sectors is that new businesses as a share of all establishments have grown substantially since the 1950s, from under 5% to over 30% in the aggregate, with most of this growth occurring post-1980.

The puzzle we confront in this paper, however, is not the modern period, but rather what changed after World War II. Why did incentives for entrepreneurship change? Did technological innovation change the returns to scale which characterized American business over the prior century (Chandler 1990)? Did population movements and changes in economic geography unlock new entrepreneurial opportunities? Did regulatory changes favoring small businesses change the relative profitability of entrepreneurship? Did small business lending grow? In this paper, we focus on the relationship between entrepreneurship and innovation, motivated by evidence that new firms have become increasingly technological in nature since World War II, and the Schumpeterian view that innovation and entrepreneurship are closely intertwined (Schumpeter 1942).

We do so in two ways. First, we document the emergence of new types of businesses by describing how firm names have changed over time. We then relate specific innovations to entrepreneurship and study how new technologies percolated into the entrepreneurial economy over the last century. Through a series of examples, we show that the rate at which innovation spurs entrepreneurship has accelerated over time, especially in the postwar period. Across all technologies, we find that the

vast majority of entrepreneurship is not by technology producers, but by businesses that identify and exploit the market opportunities that innovation creates.

With this set of results, we provide a new characterization of entrepreneurship and its relationship to innovation. Scholars have long recognized that there is heterogeneity in each new cohort of firms and entrepreneurs, both in terms of innovation intentions and outcomes. Pecuniary returns alone cannot explain entry into self-employment, as the self-employed are disproportionately in both the upper and lower tails of the income distribution (Hamilton 2000, Åstebro et al. 2011), and at their founding, only a minority of all small businesses have a new idea, hold a patent or trademark, serve a new market, or seek to grow "as big as possible" (Hurst and Pugsley 2011). Recognizing as much, recent work has sought to identify, based on founding characteristics, the subset of firms that are likely to achieve growth and scale (Guzman and Stern 2020). Research on entrepreneurship often focuses on this latter category of firms, which could plausibly create a technology and bring it to market. These firms are widely thought to be critical to generating the non-rival innovations at the core of ideas-based models of growth (Romer 1986, 1990).

We highlight that even firms that do not achieve major growth events or technological innovation can be essential to technology diffusion, aggregate growth, and creative destruction. Businesses like auto repair services, video rental stores, or IT consulting are not themselves innovative, but they both exist because of and are critical to realizing the value of underlying technological innovation (motor vehicles, video cassettes, and computers). This is consistent with the full range of economic activities that Schumpeter (1942) recognized as entrepreneurship, who wrote that "the function of entrepreneurs is to reform or revolutionize the pattern of production by exploiting an invention, or, more generally, an untried technological possibility for producing a new commodity, or producing an old one in a new way." For Schumpeter, "[This] function does not essentially consist in either inventing anything or otherwise creating the conditions which the enterprise exploits. It consists in getting things done." Although distinctions between "high growth" and "Main Street" entrepreneurship can be important and clarifying, they do not imply that Main Street entrepreneurship is unrelated to technology or growth: it is critical to both.

We then show how this notion of entrepreneurship can be formalized in a simple model of endogenous growth. Building on Aghion and Howitt (1992) and Howitt (2018)'s simplified representation, we present a model where entrepreneurs drive the diffusion of innovation by creating new businesses around it. In this model, labor can be partitioned into productive labor and entrepreneurial labor (or, in an augmented model, into productive, innovative, and entrepreneurial labor). Entrepreneurs effectively act as multipliers on innovation in driving growth, but they are necessary to realize the

productivity-enhancing effects of innovation. Our goal with this framework is not calibrating parameters or solving a balanced growth path, but rather providing a structured view of the function of Main Street entrepreneurs in the Schumpeterian paradigm.

Put in full view, we believe this article offers insights for several academic literatures, including research on the historical and modern drivers of entrepreneurship, innovation and entrepreneurship, industry evolution, entrepreneurial strategy, and economic growth. Our first contribution is simply an empirical view of how U.S. entrepreneurship has changed over the past 100+ years, which we believe can inform continued research into the development of the American economy. Conceptually, we bring into focus a distinction between technological entrepreneurship and technology-enabled entrepreneurship which often goes unrecognized in theories and discussion of creative destruction. The distinction is crucial, in our view, to making sense of competing views of entrepreneurial strategy as exploration vs. exploitation, and to understanding how entrepreneurs engage with technological change, transform industries, and drive growth.

The paper also points to a multitude of open questions. Some of these we alluded above, including basic descriptive questions—like how did U.S. entrepreneurship rates historically vary with income, education, immigration, financial development, or tax rates—and hypotheses around the postwar take-off in the growth of new businesses. But the evidence also raises deeper questions. Why do some technologies spawn Main Street entrepreneurship, and others do not? What is the function that transforms major innovations into firm creation? How has it changed over time? How sensitive is it to policy? Does it depend on features of the technology, or the economic environment? These are among the many questions we encourage for future research.

The finding that the rate and direction of U.S. entrepreneurship changed sharply after World War II raises questions not only about this period in U.S. history, but also about the persistent effects that crises can have on the entrepreneurial economy. The crucible of a crisis can be fertile ground for economic change (Gross and Sampat 2020a,b, 2021). Short run shocks or supply constraints (such as those generated by World War II, or the COVID-19 pandemic) force substitutions, and these frictions may generate innovation or lock-in that may produce changes in entrepreneurial activity. Though this appears to have been the case after World War II, the story of the COVID-19 pandemic is not yet written. We encourage continued research and attention to these questions, as we continue marching forward into the post-pandemic era.

We proceed as follows. Section 2 reviews the existing canon of research on historical and modern U.S. entrepreneurship. Section 3 describes the various data sources that have been used to system-

atically study entrepreneurship past and present, their limitations, and the new data we bring to the literature. In Section 4, we present an assortment of new facts about historical U.S. business creation. In Section 5, we shift our attention to the Schumpeterian dynamic between innovation and entrepreneurship, and in Section 6 we formalize our intuition in an endogenous growth framework. Section 7 ends the paper with evidence on recent trends and reflections for research on the past, present, and future of the U.S. entrepreneurial economy.

2 What Do We Know about Historical U.S. Entrepreneurship?

2.1 Seeking a comprehensive view of long-run U.S. entrepreneurship

This paper contributes to the literature on U.S. entrepreneurship in three ways: by contextualizing recent research characterizing modern entrepreneurship, providing a more complete picture of pre-1970s entrepreneurship, and exploring the relationship between innovation and entrepreneurship across the U.S. economy and over long spans of history.

Although economists have been writing on entrepreneurship since as early as Adam Smith (Smith 1776), it became a subject of renewed interest in the 1940s and 1950s through the work of Joseph Schumpeter, Arthur H. Cole, and others, who argued that entrepreneurship was a key determinant of national economic performance. Among other insights, early scholars argued that entrepreneurs are supported by innovation, and that business model innovation and its imitation can be just as important to economic growth as technological innovation (Cole 1968).

Evans (1948) is, to our knowledge, the first quantitative effort to systematically study new firm creation, and is in some ways the closest antecedent to this paper. Evans describes the evolution of the legal environment around new firm registrations and as part of this exercise studies business incorporations from 1875 to 1943, using data from sixteen states. In addition to aggregate statistics, Evans measures new businesses by major industrial category, using information in firm names and business charter purpose statements to classify firms into industry categories. Like us, Evans (1948) recognizes that "a reading of the mere names of the newly chartered companies [can] indicate the nature of a wave of entrepreneurial activity ... be it the mania for skating rinks or Tom-Thumb golf courses or the feverish organization of trusts" (Evans 1948).

The past three decades have seen a resurgence of scholarly interest in entrepreneurship, concurrent with growing public interest. Much of this work consists of empirical studies rooted data we describe in Section 3. Using modern Census Bureau and private equity datasets, scholars have,

for example, highlighted that young firms play an important role in job creation and dynamism (Haltiwanger et al. 2013, Decker et al. 2014), that startups are key to translating innovation investments into product development (Kortum and Lerner 2000), and that reallocation across firms can contribute to aggregate productivity growth (Foster et al. 2018). A related stream of work has studied determinants of entry into entrepreneurship, including the importance of non-pecuniary returns to entrepreneurship (Hamilton 2000, Åstebro et al. 2011, Hurst and Pugsley 2011), risk tolerance (Hall and Woodward 2010), and possible changes in profitability at small scales (Colaiacovo et al. 2022a). The past few decades have also borne a revival of research on historical U.S. entrepreneurship, especially technological entrepreneurship, and especially in the context of the developing market for technology and the division of innovative labor (e.g., Khan and Sokoloff 1993, Lamoreaux and Sokoloff 2001, Nicholas 2010, Babina et al. 2020). However, with few exceptions, this historical literature rarely measures firm creation per se. Where it does, the study of entrepreneurship is typically specific to its context.

2.2 Entrepreneurship and growth theory

Reflecting its importance to jobs and productivity, entrepreneurship plays a central role in modern theories of economic growth. The intellectual lineage of these ideas traces in part to Schumpeter, who emphasized the importance of entrepreneurs to market evolution. This evolution "incessantly revolutionizes the economic structure from within," and is the "essential fact about capitalism" (Schumpeter 1942). The principle of 'creative destruction' has since gained traction in ideas-based models of economic growth (e.g., Romer 1986, 1990, Aghion and Howitt 1992, Grossman and Helpman 1994). In this class of models, sustained economic growth is based on the increasing returns derived from non-rival ideas and innovations that enhance productivity or supersede existing product varieties. This foundational work has been further enriched by research that formally describes the joint distributions of R&D, productivity, and firm size; how innovations differ by type of firm; and Schumpeterian aggregate and industry dynamics (e.g., Klette and Kortum 2004, Lentz and Mortensen 2008, Akcigit and Kerr 2018, among others).

We add to this literature in two ways. The first is evidentiary: with long-run data, it is possible to ask whether firm formation co-evolves with growth. The second is conceptual, as we add nuance to academic debate about the types of entrepreneurship essential to growth and technological change. In the growth literature, creative destruction is often implicitly or explicitly confined to a specific innovative sector, or firms that deliberately invest in R&D (Aghion and Howitt 1992). At the same

time, scholars of entrepreneurship have documented that this is only a small subset of each cohort of new firms: most founders do not have a patent, intend to serve a new market, or plan to grow 'as big as possible' (Hurst and Pugsley 2011).[1] However, while differentiating between Silicon Valley and 'Main Street' entrepreneurship is important for many types of studies, we caution that this latter type of entrepreneurship should not be considered unrelated to technology and innovation. As we will show, even businesses not explicitly engaged in developing new technology can be both enabled by and engaged in propagating new innovations.

Notably, this point is consistent with the model Schumpeter articulated. Schumpeter (1942) claims "the function of entrepreneurs is to reform or revolutionize the pattern of production by *exploiting* an invention or, more generally, an untried technological possibility for producing a new commodity or producing an old one in a new way" (emphasis added). In this sense, the entrepreneurial function "does not essentially consist in either inventing anything" but rather is about "getting things done" (Schumpeter 1942). Baumol (1968) explains "it is [the entrepreneur's] job to locate new ideas and to put them into effect."[2] Although research in the growth literature and beyond focuses on the innovative entrepreneur, these are rare, and a Schumpeterian view would maintain that workaday Main Street firms—the exploiters, versus explorers—are important engines of economic change. This is a view which the evidence in this paper will reinforce.

2.3 Entrepreneurship and industry dynamics

Finally, this paper also connects to the large literature in economics and strategic management on industry dynamics—especially when we analyze firm creation around new technology. Alongside theories of industry development from organizational ecology that emphasize population dynamics and the emergence of organizational forms (see Darby and Zucker 2003 for a review), and industrial organization's focus on competitive dynamics, a major branch of this literature focuses in particular on the relationship between technological and industrial change.

Early studies identified empirical regularities in industry evolution using carefully collected data on a small number of industries and products (e.g., Abernathy and Utterback 1978, Gort and Klepper 1982, Tushman and Anderson 1986, Klepper and Graddy 1990). Often, after an initially slow start,

[1] Other research similarly dichotomizes entrepreneurship when studying heterogeneity across entrepreneurs: Levine and Rubinstein (2017) show that among the self-employed, incorporation usually signals a business with larger revenues and with an owner engaged in less routine tasks. In the context of developing countries, Schoar (2010) likewise differentiates between 'subsistence' and 'transformational' entrepreneurs.

[2] Leibenstein (1968) describes entrepreneurs as "gap-fillers" and "input-completers", in the sense that they connect markets and provide inputs necessary to create value from new products. Many of the examples of new businesses we discuss in Section 5, from video rental stores to computer repair shops, have these features.

a period of high net entry is followed by a stabilization, then negative net entry, before the set of industry participants stabilizes (Gort and Klepper 1982).

For industries that develop around new technologies, this pattern of rapid initial entry followed by a "shakeout" may be explained by theories of dominant design. Initially, many firms enter as the technical specifications of a new technology are contested. Eventually, a dominant design emerges that embodies "the requirements of many classes of users of a particular product" (Suarez and Utterback 1995). After the establishment of this standard, the industry contracts as incumbents transition to competing on cost and scale versus innovation.[3]

Our paper adds to this literature by documenting links between technological innovation and entrepreneurial dynamics. Whereas previous studies focus on (often hand-collected) samples of a few products or technology areas, and most study product market entry without distinguishing existing and new firms, our data allow us to provide a comprehensive view across the universe of new firms and technologies, and to study the entry margin specifically.

3 Data: Measuring Historical Entrepreneurship

Our analysis is primarily based on business registration records from the Startup Cartography Project (Andrews et al. 2022) from 1900 to 2009, as well as a cross-sectional sample of U.S. firms in 1990 provided by Dun & Bradstreet. Before delving into the features of these data, we provide a short overview of other data sources and their limitations for our purposes.

3.1 Entrepreneurship Data in Economics

A range of datasets have been used to document the level and characteristics of new startups over time in the United States. We bin these into two categories: modern and historical. Perhaps the most often used data in entrepreneurship research is the Census Bureau's Longitudinal Business Database (LBD) (Jarmin and Miranda 2002). The LBD is a confidential, restricted-use dataset that includes the count and age of all tax-paying employer firms by year, and the derivative Business Dynamics Statistics (BDS) is a public dataset reporting the aggregate counts of new firms by year

[3]These dominant designs are sometimes hypothesized to emerge after a technological discontinuity that changes the price-performance frontier (Anderson and Tushman 1990). However, even incremental innovations in underlying technologies can lead to failure of established firms, if they alter the architecture of a product in ways that incumbents do not perceive or address (Henderson and Clark 1990) or provide a new technical solution for a niche market that allows entrants a foothold to eventually innovate past incumbents (Christensen 1993).

in the LBD (Haltiwanger et al. 2009). Financial economists frequently use datasets that measure investment in the select subset of startups that receive private investment, such as through venture capital (VC) databases (e.g., VentureXpert and Venture Source, or more recently Preqin, CB Insights, and the crowd-sourced platform Crunchbase). In addition to these datasets, there are a series of commercial databases that seek to cover a large portion of the universe of firms in the United States, such as Dun & Bradstreet (D&B) and Infogroup.[4]

These data sources are limited in several ways for research on long-run trends in entrepreneurship, primarily in the set of firms sampled, and in the time periods they cover. The LBD, for example, begins in 1977 and only measures employers. Investment datasets generally have very poor coverage prior to the late 1990s, only measure firms receiving investment, and have meaningful inconsistencies between them (Kaplan and Lerner 2016).[5] Commercial datasets are limited in that they only sample firms relevant to their business purpose, and are also recent.

This does not mean that historical entrepreneurship has not or cannot be empirically studied. Census instruments such as the decennial census of population or the Current Population Survey (CPS) measure the self-employed population and its characteristics—and microdata on the complete U.S. population are available for each census from 1850 to 1940. These datasets too are limited in that they only measure self-employment (not business creation), and they do so in only in decadal cross-sections. Prior work has also used patent records to identify technological entrepreneurs, typically as solo inventors (e.g., Lamoreaux and Sokoloff 1999b,a, Nicholas 2003, 2010, Babina et al. 2020). The patent record does span most of U.S. history, but it covers only a small, distinctive subset of historical entrepreneurship, and even that only imprecisely, since not all solo inventors were business creators—though some studies take the extra step of linking inventors to new businesses (e.g., Khan and Sokoloff 1993, Lamoreaux and Sokoloff 2005).

3.2 Business registration data

To overcome each of these obstacles, this paper takes advantage of business registration data. Business registration is the act of creating a new partnership, corporation, or—since 1993, in most of the U.S.—limited liability companies (LLCs). As such, they represent the legal founding of a

[4]The providers of these datasets developed their products for commercial purposes. For example, D&B data were compiled to assess small business creditworthiness, and Infogroup covers small businesses that listed in the Yellow Pages, and has been used to target small business marketing. Although D&B has been in the credit scoring business for nearly 200 years, electronic data are only available beginning in the late 1960s.

[5]The modern venture capital industry is itself relatively new (Nicholas 2019), and its participation in the U.S. entrepreneurship system only became substantial after the Bayh-Dole Act in 1980 (Lerner 2009).

company, and an anchoring moment of entrepreneurship.

While U.S. corporations have been created since at least the 1600s, it was traditionally a slow process that was tightly controlled by the state (pre-American Revolution, by England; later, by state governments). Since creating new corporations often required state legislatures to pass an act providing it a charter, there was historically a considerable difference between overall entrepreneurship and legal firm formation. The incorporation process was gradually opened in late 1800's, starting with New Jersey in 1896, and quickly expanding to other states. By roughly 1915, two features characterized business incorporation across the U.S. First, all states allowed general (i.e., open) incorporation. Second, firms in every state could choose a legal jurisdiction (a sort of statutory domicile) different from their physical headquarters location. Incorporating firms in practice typically chose between two options: their local state law, which advantages new firms operating in the state, and Delaware corporate law, which is more beneficial for larger firms that engage in business across states, have a large operation, or intend to list on the stock market. In a more recent, post-1988 sample, being registered in Delaware is associated with a 23 times higher probability of reaching a high growth outcome (Guzman and Stern 2020).

We collected data on all business registrations across 50 U.S. states through the Startup Cartography Project (Andrews et al. 2022), of which 47 included historical data.[6] Each record provides the company's name, physical location, state of registration, jurisdiction of choice, corporate form (i.e., corporation versus LLC), and registration date. These data allow us to measure all new legal entities registered in the United States from 1900 onwards, providing a rich characterization of entrepreneurship over time. The major limitations, as we see them, are two. First, the data do not include firms' industry or business characteristics, which will motivate our developing a new way to categorize businesses into sectors. Second, and more importantly, they specifically measure the creation of legal entities, and do not include unincorporated sole proprietors.

Focusing on new legal entities allows us to study an anchoring moment in firm formation that maps naturally to a definition of entrepreneurship: it is the moment a business idea gets translated into an independent organization tasked with executing on it. This is a consistent definition that our data allow us to apply over our entire sample. Other considerations beyond entrepreneurial impetus may also influence registration, such as capital requirements, administrative burden, and the availability of complementary legal services. A sizeable literature documents that changes in benefits and costs can induce (or reduce) incorporation (Klapper et al. 2006, Djankov et al. 2002,

[6]The Startup Cartography Project primarily documents entrepreneurship since 1988, but we are able to retrieve the full history of firm registrations, both active and deceased, from nearly all states.

Landes et al. 2012), and we take the view that any such patterns in our data represent real changes in firm formation, rather than accounting differences.[7] That our key results are similar across U.S. states, despite their disparate registration rules, reinforces this view.

3.3 Dun & Bradstreet data

We supplement these data with D&B data from 1990, which comprises all 6.6 million firms in the D&B universe as of that year. Despite the limitations mentioned above, these data offer a few benefits for our analysis. First, because these firms are classified to an SIC industry, they provide us a sample we can use to train a procedure for classifying businesses to sectors via their name. Second, because they include founding year, they provide a complementary sample to the registration data, especially allowing us to characterize firms that survived for long periods (whereas the registration data measure the universe of firms only at entry). Third, they represent an interesting subset of startups: those which applied for credit, which implies they had an intention to productively use capital. After experimenting with different data years, we chose 1990 to balance recency of the set of firms in the sample against survival of older cohorts.

4 Basic Facts about U.S. Entrepreneurship since 1900

4.1 Aggregate patterns

We begin by describing aggregate patterns in business registrations from 1900 to 2009. Figure 1, Panel (A) plots the total number of new firm registrations over time, in log scale. The blue line charts the raw count, and the red line registrations per capita. Both series show a striking take-off of entrepreneurial activity after World War II. Whereas there is relatively little change in the rate of business creation from 1900 to 1940, especially after adjusting for population, new firms and new firms per capita both begin to grow rapidly in the postwar era. Equally notably, the growth rate in new businesses per capita has been remarkably stable since the 1940s.[8] This marks our first finding, which breaks with conventional views: business registrations suggest a vibrant and growing entrepreneurial economy is not a constant of U.S. history, but rather a recent phenomenon. To our knowledge, this fact was not previously known or documented.

[7] Similar limitations apply to most modern studies of firm formation, including those based on the LBD.

[8] We can also note the modest decline during the Great Depression (1930s) and sharper decline during World War II, when economic resources were shifted into warfighting and war production.

[Figure 1 about here]

In Panel (B), we restrict the sample to firms of Delaware jurisdiction. Although state level corporate law is unique to each state, large companies have coalesced around Delaware as the jurisdiction of choice for U.S. business registration, due to favorable legal treatment, and recent evidence suggests that Delaware registration is a strong predictor of firm growth (Guzman and Stern 2020).[9] We see in Panel (B) that although Delaware incorporation law was not developed until 1899, it quickly gained acceptance in practice. As Evans (1948) explains, the rapid growth of Delaware registrations in the early part of the century reflects the adoption of Delaware incorporation, rather than a boom in high growth entrepreneurship per se. Panel (B) then shows Delaware jurisdiction registrations flattening prior to World War II, declining, and then taking off in the post-war period, with a stable, long-run growth rate that matches the pattern in Panel (A).[10]

We study regional heterogeneity in these trends in Figure 2, where we split our sample across nine U.S. Census divisions, and plot new firm registrations per capita for each. Across all divisions, new firm growth is roughly flat pre-World War II, and subsequently takes off in the postwar era.[11] The war marks not only a trend change in the growth of new firms per capita, but also regional convergence. The consistency of these patterns across the country builds confidence that they are not an artifact of changes in business registration law or practice in any specific state, nor of changes in other local institutions or the industrial composition across states, but rather reflective of a true trend change in the growth of the entrepreneurial economy.

[Figure 2 about here]

These patterns immediately challenge the axiomatic view that the U.S. is and has always been, a nation of entrepreneurs—though this rebuttal is not necessarily at odds with prior evidence. The late nineteenth and early twentieth century was an era of increasing returns to scale and capital deepening, driven by new sources of motive power and organizational innovation, with increasingly large factories replacing artisans, the development of the American system of manufacturing, the

[9] Registering in Delaware allows better corporate law for complex deals and better precedence on legal decisions. However, it is more expensive and requires to maintain two registrations. This creates a separating equilibrium. Firms registered in Delaware are empirically 23 times more likely to achieve a significant growth outcome than those registering under local jurisdiction (Guzman and Stern 2020). Delaware corporate law was not developed until 1899, but it quickly gained acceptance as the cannon of business law in the U.S.

[10] Note that we also observe a few punctuated events, including a temporary increase in registrations in 1967 with the passage of the improved Delaware Corporate Law.

[11] While there are some changes in the overall ranking of these regions over time, these changes are relatively small compared to the consistency with which they co-move.

rise of multi-divisional enterprises, and the formalization of management. On the other hand, a significant literature has also studied entrepreneurs in this period, including the papers we cited above. In comparison to the postwar period, however, the evidence thus far suggests that entrepreneurship in the early twentieth century was relatively feeble or stymied.

Benchmarking growth rate against other series

One possibility is that these trends may be driven by broader economic growth. We evaluate this in Table 1, comparing growth rates in business registrations and real GDP per capita.[12] Column (1) estimates average growth rates in the pre-Depression era. The two series track each other: between 1900 and 1929, business registrations per capita grew at an annual rate of 1.8%, and real GDP per capita at a statistically-indistinguishable 1.5%. Column (2) then estimates these growth rates from 1946 to 2009, showing that business registrations rapidly outpaced real GDP growth, with roughly 2.5x the annual rate of growth (5.5% versus 2.1%). Powered by this post-war growth, the average annual growth rate of per-capita business registrations from 1900-2009 is double that of real GDP. These patterns can be visually seen in Figure 3, Panel (A), which plots the two series against each other, indexing each to its 1900 level (which is normalized to 1).

[Table 1 and Figure 3 about here]

Another benchmark is the count of total firms in the U.S. economy. We use the CBP to normalize business registrations from 1956 onwards. Figure 3, Panel (B) charts new firms' share of establishments from 1956 to 2009. Over the second half of the twentieth century, new businesses have grown ten-fold as a share of all establishments, from roughly 3% in 1956 to 30% in 2009.

4.2 Industrial composition

Figure 4 provides a distinct view of historical entrepreneurship: clusters of 5-year intervals, grouped by their similarity in the industrial composition of new firms. Here we invoke the Dun & Bradstreet 1990 sample, which reports founding years and SIC codes, restricting to headquarters locations and single-establishment firms and measuring industries at the 2-digit SIC level. For each demi-decade from 1910-1914 to 1985-1989, we calculate the distribution of new firms across SIC2 industries. We

[12] Our U.S. GDP series is borrowed from Jordà et al. (2017).

then measure the cosine similarity of every pair of demi-decades and cluster them hierarchically, applying two clustering methods as robustness checks on each other.[13]

[Figure 4 about here]

World War II thus not only produced a trend break in the growth of American entrepreneurship; it also produced a change in its composition, as Figure 4 effectively partitions twentieth century entrepreneurship into two eras: pre- and post-World War II. These results should, however, be caveated by the limitations of the D&B sample, which conditions to firms in the D&B universe (typically, credit applicants) that survived to roughly 1990, and this sample may or may not be representative of all firms founded each period.

4.2.1 Extending industry classification to business registrations

In principle, we would like to know the composition of *all* new businesses. The challenge is that registration data do not contain a measure of a firm's line of business. In administrative datasets such as those produced by the Census Bureau, firms' industry is categorized into a standardized classification such as SICS or NAICS by a professional coder, using information collected through mandatory-response census instruments. State offices that register businesses, however, do not classify new registrations into industries, nor do they consistently collect information that could directly support an ex-post classification by administrators or researchers. Our challenge in seeking to measure new businesses' industry or economic sector is compounded by the fact that in our business registration data, we have only a small number of characteristics we observe for each firm: name, physical state, registration state, and registration year.

Despite these limitations, classification is not necessarily ruled out. Recent papers have shown that firm names can provide information on firms' performance and growth potential (e.g., McDevitt 2014, Belenzon et al. 2017). We contend that firm names also provide useful information about industry. To harness this information, we develop a procedure that associates words in firm names with each of ten economic sectors, which we introduce here and describe in detail in Appendix A. Concretely, we use Dun & Bradstreet data—which include both firm names and SIC codes—to train an algorithm that measures the association of individual words in firm names with the following 10

[13]The two methods are average linkage and Ward's method. Each approach groups observations successively, based on their distance, until all observations belong to one group. Average linkage measures distance between two groups (including singletons) as the average of the distance between the observations in each group. Ward's method creates groupings to minimize within-cluster variance. Other methods produce similar patterns.

sectors, into which the official U.S. SIC classification maps, and apply this to impute firms' sectors based on their names:

1. Agriculture, Forestry, & Fishing
2. Construction
3. Finance, Insurance, & Real Estate
4. Manufacturing
5. Mining
6. Public Administration
7. Retail Trade
8. Services
9. Transportation & Public Utilities
10. Wholesale Trade

As we will show, this simple and intuitive procedure does a reliably good job at identifying firms' economic sectors in multiple validation samples. It comes with the added advantage that it supports a more flexible classification than SICS or NAICS: firms can be fractionally associated with multiple sectors and thereby span sectoral boundaries—as many firms do in practice, and the examples we give will illustrate. We believe that this methodological approach can be generalized to other data sources and even other settings with textual content, such as patent data. Our approach is related in spirit to Hoberg and Phillips (2016), who (i) measure firms' pairwise distance based on the textual similarity of their business descriptions in SEC 10-K annual reports, and (ii) use this to dynamically cluster firms in product space. Our approach is different in that it maps firms to a fixed, time-invariant classification at the sector level—which is the level at which our experience has shown us our method can reliably predict the business that a firm is in on the basis of its name alone—and allows firms to belong to multiple sectors.

Methodological approach

We begin with the Dun & Bradstreet 1990 firm sample, where we observe firms' primary SIC, from which we can obtain economic sectors. We filter this sample to headquarters establishments and single-location firms (dropping branches of multi-establishment companies) and divide the sample into random halves for training and validation. Working with the training sample, we split firms' names into tokens. For each token j in this sample, we compute each sector s's share of all uses of that token, which we will call a token-sector score, τ_{js}. These scores measure the sectoral association of every word in any given firm's name. We aggregate up from tokens to firms by averaging the sector scores of the tokens in a firm's name. Each firm i will then be characterized by a vector of sector associations $\{\varphi_{is}\}_{i=1}^{10}$, where $\varphi_{is} = \frac{1}{J_i}\sum_{j=1}^{J_i}\tau_{js}$, and J_i the number of words in firm i's name. By construction, a firm's sector scores will sum to one: $\sum_{s=1}^{10}\varphi_{is} = 1$.

For a real example from our data, consider the firm "Anderson Home Appliances", whose words

can be fractionally classified across each of the original 10 economic sectors. In our data, these words associate with sectors as shown below (where only the top five sectors are shown), leading to the firm-level scores shown in the last row of the table. Our procedure predicts *Retail Trade* and *Services* as equally likely sectors. The actual primary sector reported for this firm in the D&B data is *Services*, and the secondary sector reported in D&B is *Retail Trade*—though we might infer that Anderson Home Appliances is related to *Construction* too.

Firm-sector scores for "Anderson Home Appliances" (true example)

	Token-sector scores and cross-token averages				
Word	Agriculture	Construction	Fin./Ins./RE	Retail Trade	Services
Anderson	8%	17%	10%	14%	34%
Home	0%	20%	10%	17%	48%
Appliances	0%	2%	0%	71%	21%
Average	3%	13%	7%	34%	34%

Validation

We validate this sector classification procedure in multiple ways. We first do primarily with our D&B test sample, where we can classify firms into sectors and compare our predictions to what is reported in the data. As Figure 5 shows, we are able to match the overall distribution of sectors well. Broadly, it is not hard to understand why: the most common words in firm names are also the most sector-specific, like 'Construction', 'Service', 'Shop', or 'Investments'. In Appendix A we also show that our procedure performs well on measures of precision and recall, as well as on receiver-operating characteristic (ROC) scores. Appendix Tables B.3 and B.4 and Figure B.2 show that a firm's top-predicted sector is by far most likely to be its reported sector. This performance is also stable across founding years (Appendix Figure B.3), and matched in a distinct validation sample comprised of firms in the InfoGroup 2000 sample, which also reports firm names and SIC industries. Note that we do not expect perfect predictive power, given that we believe many firms in practice span sectors, even when the data can only report them as belonging to one sector. In our view, this flexibility is a feature of the method, rather than a bug.

[Figure 5 about here]

Classifying business registrations

Our final step is to classify firms in the business registration data to economic sectors. We clean firm names following the same rules used for the training sample. We then apply the token-sector scores to words in these firms' names, and aggregate them up to the firm level.

This classification nevertheless has limitations. One is that some words in the set of firm names in the registration sample may not be present in the Dun & Bradstreet sample, and thus will be unable to be scored. Due to this, as well as the removal of stop words, some firms in the registration data may not be able to have their sector predicted. We are reassured, however, by the fact that even under our most restrictive sector classification procedure, we classify 90% of firms—and this rate is stable over time. We believe this is high enough to generate useful measures of the sectoral distribution of U.S. entrepreneurship over long periods.

4.2.2 Entrepreneurship by economic sector

In Figure 6, Panel (A), we use our now-classified business registration data to describe the composition of U.S. entrepreneurship over the twentieth century. We first note the relative stability in the composition of startups over long horizons. But we also observe a number of changes, such as the long, slow rise of *Services* (from ∼30% of new businesses in the early 1900s to ∼36% in the early 2000s; the explosion in *Construction* as an entrepreneurial trade after World War II (from ∼6% to ∼10% of new businesses); and the long-run decline of *Mining* (extraction) as a nexus for entrepreneurial activity (from 2-3% to <1% of new businesses). Figure 6 also shows some shocks to business registrations, such as a spike in new manufacturing firms during World War II, or in finance/real estate registrations in 1925, which was driven by a massive real estate bubble in Florida that drew investors from around the country. One consistent pattern across all sectors is that new firms' share of CBP establishments grew substantially over the second half of the century (Panel B).[14] This is especially the case for small sectors such as *Agriculture* and *Mining*, where new businesses comprised 25-30% of establishments by the late 1990s.

[Figure 6 about here]

These findings provide what we believe are new facts about the twentieth century U.S. entrepreneurial economy. In some ways, these patterns mirror broader trends among larger and incumbent firms. But they also offer distinct insights. We see, for example, how quickly U.S. entrepreneurship has reacted to aggregate shocks like World War II government demand or the 1970s energy crises—and an interesting question for dynamism is how the speed of reallocation has changed over time. The biggest takeaway is that an increasing share of production appears to be taking place in new

[14] We plot these sector-level series only until 1997, when the U.S. transitioned from the SIC classification to NAICS, which reallocated some establishments across sectors and led to discrete changes in CBP sector establishment counts that generate artificial discontinuities in our CBP-normalized time series.

companies. This fact represents a departure from the structure of the U.S. economy a century ago, when large, incumbent firms dominated American business.

5 The Relationship of Entrepreneurship to Innovation

These facts leave us with an inescapable question: what changed after World War II that lit a fire in the U.S. entrepreneurial economy? It is well known that the war produced myriad, lasting changes in the U.S. economy (Tassava 2008). Changes in incentives for wage work versus self-employment, population movements and economic dislocations, increasing educational attainment, regulatory changes, changes in small business financing, decreasing returns to scale, and many other factors could conceivably play a role. One striking pattern in our data, however, is the growth of technology-related startups. Figure 7 plots the relative frequency of three words which illustrate how the nature of U.S. entrepreneurship changed over the twentieth century. Across the century, the word "Manufacturing"—which was the 6th most common word in new firm names in the 1900s and 10th most common in the 1940s—declined nearly 90% in its frequency among new businesses, and in the 2000s ranked 254th. The word "Services" crescendoed and reached its peak in the 1960s, and has since been declining. The word "Technology", by comparison, has grown dramatically, and by the 2000s was the 48th most common word in new firm names.

[Figure 7 about here]

Motivated by this evidence, recent research on the changes that World War II brought to the U.S. innovation system (Gross and Sampat 2020a), and the Schumpeterian view of economic change, we focus on one explanation for the post-war take-off in entrepreneurship: changes in the rate of innovation and in the relationship between entrepreneurship and innovation. To understand these changes, we first need to better understand what this relationship is. We thus pose three questions: (i) does new firm creation follow innovation, (ii) what kinds of firms are created, when, and at what rate, and (iii) how have these patterns changed over time?

We take a new approach to answering these questions to which our data on the universe of firm registrations are particularly well-suited: tracing the diffusion of new technology into new business creation via firm names. Building on the intuition in Section 4, we first establish that technology can be measured in firm names, and that some technologies show up widely in firm names. In doing so, we build on recent research using natural language to measure the invention and diffusion of

innovation in the patent and publication record (e.g., Iaria et al. 2018, Arts et al. 2021) and other corpuses (e.g., Gross 2022), research exploiting information content in firm names (e.g., McDevitt 2014, Belenzon et al. 2017, 2020), and a growing literature in economics and the social sciences using text as data (see Gentzkow et al. 2019, for a summary).

We will show that (i) new firms often follow invention, but (ii) the vast majority of these new businesses are neither making nor commercializing innovation. Most firm entry related to innovation takes the form of workaday, "Main Street" entrepreneurs rather than technological innovators. This observation motivates the distinction we draw between technological entrepreneurs, who develop and commercialize new innovation, and *technology-enabled* entrepreneurs, who identify and exploit new market opportunities that innovation creates. Despite the attention technologically-innovative entrepreneurs receive in both research and popular media, we will argue and provide evidence suggesting that Main Street entrepreneurs are the driving force in Schumpeterian growth, as the agents that organize around, deploy, sell, and service the technology that effects economic growth. Finally, we will show that (iii) the number of new firms created around major new innovations has grown more than an order of magnitude over time, and the speed with which they form has accelerated, particularly in the postwar period.

5.1 Measuring technological change in firm names

We provide two additional views of the changing composition of entrepreneurship in Tables 2 and 3. Table 2 shows the most common words, and bigrams, in new firm names by decade.[15] The technological features of U.S. entrepreneurship are not necessarily visible among the most common categories of new firms in this list, though there are exceptions, like "Light [&] Power" or "Air Conditioning". The appearance of *new* n-grams may be especially interesting, as these can indicate the emergence of entire categories of entrepreneurship. Table 3 lists bigrams which first appeared in a firm name in each decade from 1950 to 1980, ranked by their total number of uses in firm names thereafter. We interpret these bigrams as representing new types of businesses which first appeared in a given decade, and then proliferated. The upper panel lists bigrams from business registrations, and the lower panel from the 1990 D&B sample.

[Tables 2 and 3 about here]

[15]The structural transformation of U.S. entrepreneurship from extraction to goods to services is evident here. Financial services (e.g., real estate) is a common category of business throughout the century, and construction after the 1940s. Table 2 also lists words common in community organizations like churches and social clubs—which are included in business registration data, as in other administrative data sources.

The table reveals several patterns. First, it shows business innovation over long horizons, as well as major categories of ordinary (Main Street) entrepreneurs like home and lawn services, or investment and legal services. We can also see how shifts in consumption, like the growth in demand for medical services through public insurance (e.g., Goodman-Bacon 2018) led to new business services like medical billing companies and surgery centers. The D&B bigrams point to another phenomenon: the emergence of new businesses that are technology-forward but not themselves technologically innovative, like video rental stores, one-hour photo labs, quick print shops, or computer services. What these enterprises have in common is that although they were not innovating in technology, they existed as a direct result of innovation—be it in video cassettes and video cassette players, photo minilab devices, printers and copiers, or computers.

These examples bring into focus potential roles that "Main Street" businesses play in technological change. For example, VHS machines were first sold in the U.S. in 1977. Prior to that, consumers had to go to a cinema, or watch a TV broadcast, to see a film. With VHS machines, they could do so at home—if they had a library of VHS cassette tapes to choose from. Dot matrix and laser printers were first produced in the late 1960s and 1970s, but these were expensive alternatives to contemporary substitutes like typesetting or typewriting. Against this wave of technological change, owners of video rental stores and instant print shops essentially performed two functions: identified local demand, and provided capital rental services, leasing access to technology that would be too expensive for households or small businesses to buy on their own.[16]

Finding customers or market opportunities and bundling demand for new technology are only two of the many roles that Main Street entrepreneurs play in the process of technological change. Small businesses also make, sell, install, and service technology. They may advise users on how to use it productively. We will argue in Section 6 that it is as much through these value-adding activities as it is through innovation itself that innovation is transformed into productivity growth, and that much of this activity is performed by atomistic entrepreneurs dispersed throughout the economy. These entrepreneurs may also be advantaged relative to incumbent firms in finding and serving these market opportunities, due to their agility: incumbents may face a higher cost of aligning new activities to their existing business, or (conversely) changing the existing business to serve this new opportunity, which not only involves the same startup costs that entrepreneurs face but may also require depreciating existing assets (e.g., Christensen 1997). Entrepreneurs may also benefit from more intimate knowledge of—and access to—local markets in particular areas (Uzzi 1997).

[16] Put differently, these entrepreneurs were making indivisible capital divisible.

5.2 Systematic evidence: Technologies in patents

In Colaiacovo et al. (2022b), we take a structured approach to studying entrepreneurial dynamics around innovation using population data on new business registrations, and here we use the same measurement approach. We identify major technologies by name using patent data, where we select the top ≈100 keywords from each decade which were new to the patent record in that decade and which had the most subsequent patents thereafter. Our keywords, in this case, are not provided by patent inventors (patents do not themselves list keywords, the way research articles do), but rather are measured and provided by Google as "the top 10 salient terms extracted from the patent's title, abstract, claims, and description." These keywords are nearly always technological, and provide us with more summary information than patent titles would alone.

We take these words (1,092 in total) to the business registration data, identifying firms with each of these words in their names (3.7 million in total). We study a variety of features of technology-driven entrepreneurship in this sample, including the shape of entrepreneurial dynamics and the relationship between a technology's characteristics and associated entrepreneurship. In this paper, we focus on differences between the pre- and post-war periods.

Figure 8 characterizes firm creation over time for two dozen example technologies (words) in this set, from audio, radio, and video to semiconductor, software, and internet. The top panel does so for technologies whose first patent was filed pre-1940, and the bottom panel post-1940. The vertical axis measures the annual fraction of business registrations in which the word was present, and the vertical line in each subfigure identifies the year of a technology's first patent. Each time series is shown as a 5-year rolling average to smooth annual fluctuations.[17]

[Figure 8 about here]

The figure is illustrative of the variation in entrepreneurial dynamics: across technologies, we see variation in the time from the first patent to the growth of related startups. In some cases, firm creation crescendos and quickly reaches its peak; in other cases, it builds slowly and stays elevated or declines gradually over time. In yet other cases it cycles. A broad pattern, however, is that the length of time from the invention of a technology to the take-off of entrepreneurship around that technology has shortened over time, especially for post-1940 inventions.

[17] Each time series is presented as a 5-year rolling average to smooth annual fluctuations. We also drop the first 10 uses of each term from the sample, as these are often errant, either picked up as spelling errors or the occasional firm that was registered in early decades and whose name later changed, where the state registration office provided a record of the firm with its most recent name.

With Figure 9 we extend this evidence to our full sample of technologies. We convert time units to event time around the year of the first patent, and plot the annual average number of startups associated with each technology.[18] Panels (B) and (C) split the sample into pre-1940 and post-1940 technologies. The postwar acceleration of technology-related entrepreneurship is plainly visible in these averages. Moreover, 90% of these firms are local and in non-manufacturing sectors (Appendix Figures C.2, C.3, and C.4). The implication is that the vast majority of new firms created around these technologies—many of which are embodied in physical products—were not making them but rather distributing, selling, installing, or servicing them.

[Figure 9 about here]

Appendix Figure C.5 provides complementary evidence, plotting in each subfigure the fraction of technologies each decade which spawn any related firms within 5, 10, 20, an 30 years. All figures show a sharp uptick from the 1950s onward. Table 4 provides a distinct view, estimating differences across technologies in the fraction of states with a related firm after 5, 10, 20, and 30 years, vis-à-vis the technology's invention decade. Technologies invented in the 1950s onwards were increasingly faster to diffuse into entrepreneurship across the country.

[Table 4 about here]

This evidence is broadly consistent with an increasing, and accelerating, conversion of innovation into entrepreneurship post-World War II, especially startups that create new products and services around this innovation, rather than producing innovation. It thus echoes the Schumpeterian view that the entrepreneur's function is to open up new markets and transform innovation into new methods of production. This has implications for research on entrepreneurship, innovation, and Schumpeterian growth theory. Though non-innovative entrepreneurs do not feature heavily in this literature, the evidence above suggests they may nevertheless be important to technology diffusion and in turn to realizing aggregate productivity growth.

6 Entrepreneurship and Endogenous Growth

To bring structure to these ideas, we propose a model that clarifies the market function that we argue many startups serve vis-à-vis technical change and economic growth. Put briefly, we argue

[18] For this chart we also restrict the sample to technologies first appearing in patents between 1920 and 1979, to ensure we observe at least 20 years pre-patent and 30 years post-patent.

that entrepreneurs—especially "Main Street" entrepreneurs—primarily support the diffusion of innovation by creating new businesses around it. Examples we have in mind are those in Section 5, like instant print shops and video rental stores, television and radio repair, central heating and cooling equipment sales and service, auto mechanics, and so on.

In Colaiacovo et al. (2022b), we use these examples to draw a distinction between two categories of entrepreneurship: technological entrepreneurs, which aim to develop and bring new technology to market, and *technology-enabled* entrepreneurs, which identify and exploit new market opportunities that innovation creates. We argue that it is especially through these businesses that innovation translates to industrial change and growth. The framework we present here is a reductive presentation that builds on the elemental model in Howitt (2018).

6.1 Endogenous growth building blocks

We begin with Howitt's aggregate production function. Suppose aggregate output is produced by a continuum of intermediate products, according to:

$$Y = L^{1-\alpha} \int_0^1 A(i)^{1-\alpha} x(i)^{\alpha} di, \tag{1}$$

where there is a fixed measure of product variety, normalized to unity, and each intermediate product i (which collectively produce a final capital good) has a separate productivity parameter $A(i)$. Each sector is monopolized and produces its intermediate product with a constant marginal cost of unity. The monopolist in sector i faces a demand curve given by the marginal product of that intermediate good in the final good sector:

$$\frac{\partial}{\partial x(i)} Y = \alpha \cdot \left(A(i) \frac{L}{x(i)} \right)^{1-\alpha}$$

Equating marginal revenue (α times this marginal product) to the marginal cost of unity yields the monopolist's profit-maximizing intermediate output: $x(i) = \xi A(i) L$, where $\xi = \alpha^{2/(1-\alpha)}$. Substituting this in for each $x(i)$ in the production function (1), we have:

$$Y = L^{1-\alpha} \int_0^1 A(i)^{1-\alpha} \left(\xi A(i) L\right)^{\alpha} di = \xi^{\alpha} L \int_0^1 A(i) di = \theta A L, \tag{2}$$

where $\theta = \xi^{\alpha}$, and A is average productivity: $A \equiv \int_0^1 A(i) di$.

In the Schumpeterian view of creative destruction and growth, innovation creates improved versions of old products. An innovation in sector i consists of a new version whose productivity parameter $A(i)$ exceeds that of the previous version by the fixed factor $\gamma > 1$.

Suppose that the probability of an innovation arriving in sector i over any short interval of length dt is μdt. Then the growth rate of $A(i)$ is:

$$\frac{dA(i)}{A(i)} \cdot \frac{1}{dt} = \begin{cases} (\gamma - 1) \cdot \frac{1}{dt}, & \text{with probability } \mu \cdot dt \\ 0, & \text{with probability } (1 - \mu) \cdot dt \end{cases}$$

In turn, the expected growth rate is:

$$E(g) = (\gamma - 1)\mu. \tag{3}$$

Following Howitt (2018), we will assume that the flow probability μ of an innovation in any sector is proportional to the current flow of productivity-adjusted R&D expenditures:

$$\mu = \lambda \frac{R}{A}, \tag{4}$$

where R is the amount of final output spent on R&D, and where the division by A reflects decreasing returns to R&D as research advances (e.g., Jones 2009, Bloom et al. 2020). From (2), the growth rate g of aggregate output will be the growth rate of the average productivity parameter A. The law of large numbers guarantees that g equals the expected growth rate ($E(g)$) of the individual productivity parameters (3). From (3) and (4) we have:

$$g = (\gamma - 1)\lambda \frac{R}{A}$$

Recognizing from (2) that $\frac{1}{A} = \frac{\theta L}{Y}$, and letting $n \equiv \frac{R}{Y}$, we have:

$$g = (\gamma - 1)\lambda \theta L n. \tag{5}$$

Thus, innovation-based theory implies that the way to grow rapidly is not to save a large fraction of output (as suggested by traditional growth theory), but rather to devote a large fraction of output to R&D, and is what is called the neo-Schumpeterian view.

6.2 Introducing an entrepreneurial sector

Suppose that in order to generate growth, innovations do not only need to be created: they also need to be widely used, and that this is the role of the entrepreneur. We treat the entrepreneur's role in the economy as coming up with business models that enhance an innovation or facilitate its diffusion (e.g., by identifying customers that lack the scale to profitably adopt a technology themselves, and renting it to them). Suppose there is some allocation of the labor force in production, L_Y, and in entrepreneurship, L_E, where $L_Y + L_E = L$. Let ϵ be the fraction of workers who are entrepreneurs, such that the fraction who are production workers is $1 - \epsilon$.

Let us introduce a new parameter, ν, which is an analog to μ and represents the probability that an entrepreneur conceives of a business model for a given innovation i. The flow probability of entrepreneurship in any sector should be proportional to the stock of entrepreneurs, given that they have some latent Poisson process for business model innovation: $\nu = \phi L_E = \phi\epsilon L$, where in practice ϵ can be $\epsilon(\cdot)$, a function of some policy parameters that we use to affect the fraction of workers who select into the entrepreneurial sector. The role of ν in this case is to multiply μ: $\mu \cdot \nu \cdot dt$ is the flow probability that an innovation in sector i is *both* invented and converted into a business model. It follows that the growth rate from (5) would then be:

$$\begin{aligned} g &= (\gamma - 1)\lambda\theta L_Y n \cdot \phi L_E \\ &= (\gamma - 1)\lambda\theta(1 - \epsilon)Ln \cdot \phi\epsilon L \\ &= (\gamma - 1)\lambda\phi\theta L^2 n \cdot (1 - \epsilon)\epsilon, \end{aligned} \tag{6}$$

When $\epsilon = 0$, $g = 0$ because no innovations make it to Main Street, which plays a pivotal role in transforming innovation into growth. When $\epsilon = 1$, $g = 0$ because nobody produces on Main Street. g is maximized by balancing productive and entrepreneurial labor ($\epsilon = 0.5$).

We might also imagine there to be decreasing returns to entrepreneurship, if we believe the marginal productivity of entrepreneurs declines as the sector grows. From a policy perspective, the rate at which entrepreneurial productivity decreases would determine how big the entrepreneurial sector should be. To see this, let $\nu = \phi \cdot \epsilon^x L$, where $x \in [0, 1]$ is the rate at which the productivity of additional entrepreneurs decays. Then g is:

$$g = (\gamma - 1)\lambda\phi\theta L^2 n \cdot (1 - \epsilon)\epsilon^x$$

which is consistent with (6) and maximized at $\epsilon = \frac{x}{x+1}$.

6.3 Adding innovative labor to the model

The framework above divides labor into the productive and entrepreneurial sectors. Innovation productivity is strictly a function of R&D expenditure, and not of an innovative labor force that is distinct from the production labor force. What if we wanted to add one? Let us now suppose there is some allocation of the labor force in production, L_Y, in innovation, L_I, and in entrepreneurship, L_E, where $L_Y + L_I + L_E = L$. Let ι be the fraction of workers who are innovators and ϵ entrepreneurs, such that the fraction who are production workers is $1 - \iota - \epsilon$.

We can revise the flow probabilities of innovation and entrepreneurship as follows:

$$\mu = \lambda L_I \frac{R}{A} = \lambda \iota L \frac{R}{A} \qquad , \qquad \nu = \phi L_E = \phi \epsilon L$$

It follows that the growth rate would then be:

$$g = (\gamma - 1)\lambda\phi\theta L^3 n \cdot \iota\epsilon(1 - \iota - \epsilon) \tag{7}$$

What values of ι and ϵ maximize (7)? Symmetry implies they will be equal in equilibrium. Without loss of generality, let $\iota = \epsilon$, which leads to the maximand $\epsilon^2(1 - 2\epsilon)$, from which it follows that growth is maximized at $\epsilon = \frac{1}{3}$. As above, this balances innovation (ι), entrepreneurship (ϵ), and scale effects ($1-\iota-\epsilon$) in growth. Note that if we incorporated a demand side, with a positive discount rate, the utility-maximizing allocation of labor to innovation and entrepreneurship would be lower than the even split which maximizes growth, because we will put positive weight on current period consumption, which is a function of current period output. This pressure will optimally reallocate labor from innovation and entrepreneurship into production, putting current and future production in tension because these three sectors split a fixed labor supply.

7 Discussion and Conclusion

In this paper, we used administrative data on business registrations across nearly all U.S. states to show how business creation has evolved since 1900, focusing especially on the increased growth rate, change in composition, and tighter link to technology in the postwar years. Methodologically, we showed that substantial information can be gleaned from firm names regarding a firm's

line of business or its relationship to innovation. Conceptually, we then identified a distinction between new businesses that create versus exploit innovation—what we label technological versus technology-enabled entrepreneurship. Evidence that innovation is increasingly generating "Main Street" business creation may partially explain the post-World War II take-off in firm creation. More broadly, the distinction is crucial, in our view, to understanding how entrepreneurs engage with technological change, transform industries, and drive economic growth.

On their own, these facts shed new light on the development of American entrepreneurship, filling empirical gaps in our understanding of firm creation over long horizons and offering evidence to motivate future research. They also provide a potentially countervailing narrative to recent evidence of declining U.S. business dynamism since 1980 (Decker et al. 2014): our registration data indicate that new firm creation has grown at a relatively steady rate throughout the post-war era (Figure 1), without any discernable break around 1980. This difference could be attributable to differences in samples (registered businesses versus tax-paying employers) or differences in how dynamism is defined and measured (firm creation versus job creation).

We do, however, find some indications in our data that new business creation may be changing in ways masked by the aggregate firm counts. We highlight one of these patterns in Figure 10: the emergence of new types of businesses—as measured by the use of new words in firm names—appears to have slowed since 1980.[19] Elsewhere we find evidence that the elasticity of business registrations to population growth has also declined since 1980, which could potentially reflect increasing returns to scale across the economy over the period (e.g., Autor et al. 2020).

[Figure 10 about here]

A large number of questions inevitably remain. Chief among them is what changed after World War II. Though we have reason to believe technological innovation, and the rate at which innovation begets new firms, is likely important to understanding the postwar take-off of U.S. entrepreneurship, this is by no means an exclusive or definitive explanation. Many of the questions we raised in the introduction—around how entrepreneurship relates to population movements (e.g., postwar suburbanization) or to demographic changes, or to changes in small business regulation or financing—are ripe for further study. So are policy institutions like the Small Business Administration, which was created in 1953 to support entrepreneurs and small businesses. The data we introduce in this paper present myriad opportunities like these for continued research.

[19] Specifically, the figure shows that from the early 1940s until roughly 1980, the vintage of words in new firm names was declining: firm names increasingly included words which were only recently introduced. Since 1980, this pattern has reversed, and the vintage of words in firm names has aged.

A point we wish to reinforce is that the vast majority of entrepreneurs are not themselves innovative, but our data suggest many are nevertheless instrumental to unlocking the value of new innovations. These are the workaday entrepreneurs who "locate new ideas and put them into effect" (Baumol 1968). In this sense, Main Street entrepreneurs effectively act as multipliers on innovation in driving long-run growth. Motivated by this evidence, we believe entrepreneurs should be studied not only as innovators, but also as users and propagators of innovation.

Finally, although modern research on entrepreneurship mostly focuses on the post-1980 (and especially post-2000) era, there is much to learn from history about the nature of business creation, how it changes over time, and how it relates to other economic activities and outcomes. This is perhaps our most important insight: it is only through its historical development that we can discover the foundations of the modern American entrepreneurial economy.

References

Abernathy, William J and James M Utterback. 1978. "Patterns of industrial innovation," *Technology Review*, Vol. 80, No. 7, pp. 40–47.

Aghion, Philippe and Peter Howitt. 1992. "A model of growth through creative destruction," *Econometrica*, Vol. 60, pp. 323–351.

Akcigit, Ufuk and William R Kerr. 2018. "Growth through heterogeneous innovations," *Journal of Political Economy*, Vol. 126, No. 4, pp. 1374–1443.

Anderson, Philip and Michael L Tushman. 1990. "Technological discontinuities and dominant designs: A cyclical model of technological change," *Administrative Science Quarterly*, pp. 604–633.

Andrews, Raymond J, Catherine Fazio, Jorge Guzman, Yupeng Liu, and Scott Stern. 2022. "The Startup Cartography Project: Measuring and mapping entrepreneurial ecosystems," *Research Policy*, Vol. 51, No. 2, p. 104437.

Arts, Sam, Jianan Hou, and Juan Carlos Gomez. 2021. "Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures," *Research Policy*, Vol. 50, No. 2, pp. 104–144.

Åstebro, Thomas, Jing Chen, and Peter Thompson. 2011. "Stars and misfits: Self-employment and labor market frictions," *Management Science*, Vol. 57, No. 11, pp. 1999–2017.

Atack, Jeremy and Robert A Margo. 2019. "Gallman revisited: Blacksmithing and American manufacturing, 1850-1870," *Cliometrica*, Vol. 13, No. 1, pp. 1–23.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. "The fall of the labor share and the rise of superstar firms," *The Quarterly Journal of Economics*, Vol. 135, No. 2, pp. 645–709.

Babina, Tania, Asaf Bernstein, and Filippo Mezzanotti. 2020. *Crisis innovation*. NBER Working Paper No. 27851.

Baumol, William J. 1968. "Entrepreneurship in economic theory," *The American Economic Review*, Vol. 58, No. 2, pp. 64–71.

Belenzon, Sharon, Aaron K Chatterji, and Brendan Daley. 2017. "Eponymous entrepreneurs," *American Economic Review*, Vol. 107, No. 6, pp. 1638–55.

——— . 2020. "Choosing between growth and glory," *Management Science*, Vol. 66, No. 5, pp. 2050–2074.

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb. 2020. "Are ideas getting harder to find?" *American Economic Review*, Vol. 110, No. 4, pp. 1104–44.

Chandler, Alfred D. 1977. *The visible hand*. Cambridge: Harvard University Press.

——— . 1990. *Scale and scope: The dynamics of industrial capitalism*. Cambridge: Harvard University Press.

Christensen, Clayton M. 1993. "The rigid disk drive industry: A history of commercial and technological turbulence," *Business History Review*, Vol. 67, No. 4, pp. 531–588.

——— . 1997. "The innovator's dilemma: When new technologies cause great firms to fail."

Colaiacovo, Innessa, Margaret G Dalton, Sari Pekkala Kerr Kerr, and William R Kerr. 2022a. *The transformation of self employment*. NBER Working Paper No. 29725.

Colaiacovo, Innessa, Daniel P Gross, and Jorge Guzman. 2022b. *Innovation and the dynamics of new enterprise*. Working paper.

Cole, Arthur H. 1968. "Introductory remarks," *The American Economic Review*, Vol. 58, No. 2, pp. 60–63.

Cull, Robert, Lance E Davis, Naomi R Lamoreaux, and Jean-Laurent Rosenthal. 2006. "Historical financing of small-and medium-size enterprises," *Journal of Banking & Finance*, Vol. 30, No. 11, pp. 3017–3042.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2014. "The role of entrepreneurship in US job creation and economic dynamism," *Journal of Economic Perspectives*, Vol. 28, No. 3, pp. 3–24.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2002. "The regulation of entry," *The Quarterly Journal of Economics*, Vol. 117, No. 1, pp. 1–37.

Evans, George H. 1948. *Business incorporations in the United States, 1800-1943*. New York: National Bureau of Economic Research.

Foster, Lucia, Cheryl Grim, John C Haltiwanger, and Zoltan Wolf. 2018. *Innovation, productivity dispersion, and productivity growth*. NBER Working Paper No. 24420.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as data," *Journal of Economic Literature*, Vol. 57, No. 3, pp. 535–74.

Goodman-Bacon, Andrew. 2018. "Public insurance and mortality: Evidence from Medicaid implementation," *Journal of Political Economy*, Vol. 126, No. 1, pp. 216–262.

Gort, Michael and Steven Klepper. 1982. "Time paths in the diffusion of product innovations," *The Economic Journal*, Vol. 92, No. 367, pp. 630–653.

Gross, Daniel P. 2022. *The hidden costs of securing innovation: the manifold impacts of compulsory invention secrecy*. Forthcoming at Management Science.

Gross, Daniel P and Bhaven N Sampat. 2020a. *Inventing the endless frontier: The effects of the World War II research effort on post-war innovation*. NBER Working Paper No. 27375.

——— . 2020b. *Organizing crisis innovation: Lessons from World War II*. NBER Working Paper No. 27909.

——— . 2021. *Crisis innovation policy from World War II to COVID-19*. Forthcoming in Entrepreneurship and Innovation Policy and the Economy.

Grossman, Gene M and Elhanan Helpman. 1994. "Endogenous innovation in the theory of growth," *Journal of Economic Perspectives*, Vol. 8, No. 1, pp. 23–44.

Guzman, Jorge and Scott Stern. 2020. "The state of American entrepreneurship: New estimates of the quantity and quality of entrepreneurship for 32 US states, 1988-2014," *American Economic Journal: Economic Policy*, Vol. 12, No. 4, pp. 212–43.

Hall, Robert E and Susan E Woodward. 2010. "The burden of the nondiversifiable risk of entrepreneurship," *American Economic Review*, Vol. 100, No. 3, pp. 1163–94.

Haltiwanger, John, Ron S Jarmin, and Javier Miranda. 2009. *Business dynamics statistics: An overview*. Kauffman Foundation Other Research.

——— . 2013. "Who creates jobs? Small versus large versus young," *Review of Economics and Statistics*, Vol. 95, No. 2, pp. 347–361.

Hamilton, Alexander. 1795. "The Defence No. XI," *Papers of Alexander Hamilton*, Vol. 19, p. 90.

Hamilton, Barton H. 2000. "Does entrepreneurship pay? An empirical analysis of the returns to self-employment," *Journal of Political Economy*, Vol. 108, No. 3, pp. 604–631.

Henderson, Rebecca M and Kim B Clark. 1990. "Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms," *Administrative Science Quarterly*, pp. 9–30.

Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi. 2014. "Growth and structural transformation," in Aghion, Philippe and Steven N Durlauf eds. *Handbook of Economic Growth*, Vol. 2, pp. 855–941.

Hoberg, Gerard and Gordon Phillips. 2016. "Text-based network industries and endogenous product differ-

entiation," *Journal of Political Economy*, Vol. 124, No. 5, pp. 1423–1465.

Howitt, Peter. 2018. *Endogenous growth theory*. pp. 3632–3636, London: Palgrave Macmillan UK.

Hurst, Erik and Benjamin Wild Pugsley. 2011. "What do small businesses do?" *Brookings Papers on Economic Activity*, Vol. Fall 2011, pp. 73–142.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger. 2018. "Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science," *Quarterly Journal of Economics*, Vol. 133, No. 2, pp. 927–991.

Jarmin, Ron S and Javier Miranda. 2002. *The longitudinal business database*. Available at SSRN 2128793.

Jones, Benjamin F. 2009. "The burden of knowledge and the "Death of the Renaissance Man": Is innovation getting harder?" *The Review of Economic Studies*, Vol. 76, No. 1, pp. 283–317.

Jordà, Òscar, Moritz Schularick, and Alan M Taylor. 2017. "Macrofinancial history and the new business cycle facts," *NBER Macroeconomics Annual*, Vol. 31, No. 1, pp. 213–263.

Kaplan, Steven N and Josh Lerner. 2016. "Venture capital data: Opportunities and challenges," in Haltiwanger, John, Erik Hurst, Javier Miranda, and Antoinette Schoar eds. *Measuring entrepreneurial businesses: Current knowledge and challenges*, pp. 413–431.

Khan, B Zorina and Kenneth L Sokoloff. 1993. ""Schemes of practical utility": entrepreneurship and innovation among "great inventors" in the United States, 1790-1865," *The Journal of Economic History*, Vol. 53, No. 2, pp. 289–307.

Klapper, Leora, Luc Laeven, and Raghuram Rajan. 2006. "Entry regulation as a barrier to entrepreneurship," *Journal of Financial Economics*, Vol. 82, No. 3, pp. 591–629.

Klepper, Steven and Elizabeth Graddy. 1990. "The evolution of new industries and the determinants of market structure," *The RAND Journal of Economics*, pp. 27–44.

Klette, Tor Jakob and Samuel Kortum. 2004. "Innovating firms and aggregate innovation," *Journal of Political Economy*, Vol. 112, No. 5, pp. 986–1018.

Kortum, Samuel and Josh Lerner. 2000. "Assessing the contribution of venture capital to innovation," *The RAND Journal of Economics*, Vol. 31, No. 4, pp. 674–692.

Kuznets, Simon. 1973. "Modern economic growth: Findings and reflections," *The American Economic Review*, Vol. 63, No. 3, pp. 247–258.

Lamoreaux, Naomi R, Margaret Levenstein, and Kenneth L Sokoloff. 2004. *Financing invention during the Second Industrial Revolution: Cleveland, Ohio, 1870-1920*. NBER Working Paper No. 10923.

Lamoreaux, Naomi R and Kenneth L Sokoloff. 1999a. *Inventive activity and the market for technology in the United States, 1840-1920*. NBER Working Paper No. 7107.

——— . 1999b. "Inventors, firms, and the market for technology in the late nineteenth and early twentieth centuries," in Lamoreaux, Naomi R, Daniel M Raff, and Peter Temin eds. *Learning by doing in markets, firms, and countries*, pp. 19–60.

——— . 2000. "The geography of invention in the American glass industry, 1870-1925," *The Journal of Economic History*, Vol. 60, No. 3, pp. 700–729.

——— . 2001. "Market trade in patents and the rise of a class of specialized inventors in the 19th-century United States," *American Economic Review*, Vol. 91, No. 2, pp. 39–44.

——— . 2005. *The decline of the independent inventor: A Schumpeterian story?*. NBER Working Paper No. 11654.

Landes, David S, Joel Mokyr, and William J Baumol. 2012. *The invention of enterprise: Entrepreneurship from ancient Mesopotamia to modern times.*: Princeton University Press.

Leibenstein, Harvey. 1968. "Entrepreneurship and development," *The American Economic Review*, Vol. 58, No. 2, pp. 72–83.

Lentz, Rasmus and Dale T Mortensen. 2008. "An empirical model of growth through product innovation," *Econometrica*, Vol. 76, No. 6, pp. 1317–1373.

Lerner, Josh. 2009. *Boulevard of broken dreams: Why public efforts to boost entrepreneurship and venture capital have failed–and what to do about it*. Princeton: Princeton University Press.

Levine, Ross and Yona Rubinstein. 2017. "Smart and illicit: who becomes an entrepreneur and do they earn more?" *The Quarterly Journal of Economics*, Vol. 132, No. 2, pp. 963–1018.

McDevitt, Ryan C. 2014. ""A" business by any other name: firm name choice as a signal of firm quality," *Journal of Political Economy*, Vol. 122, No. 4, pp. 909–944.

Nicholas, Tom. 2003. "Why Schumpeter was right: innovation, market power, and creative destruction in 1920s America," *The Journal of Economic History*, Vol. 63, No. 4, pp. 1023–1058.

——— . 2010. "The role of independent invention in US technological development, 1880-1930," *The Journal of Economic History*, Vol. 70, No. 1, pp. 57–82.

——— . 2019. *VC: An American History*. Cambridge: Harvard University Press.

Romer, Paul M. 1986. "Increasing returns and long-run growth," *Journal of Political Economy*, Vol. 94, No. 5, pp. 1002–1037.

——— . 1990. "Endogenous technological change," *Journal of Political Economy*, Vol. 98, No. 5, Part 2, pp. S71–S102.

Schoar, Antoinette. 2010. "The divide between subsistence and transformational entrepreneurship," *Innovation Policy and the Economy*, Vol. 10, No. 1, pp. 57–81.

Schumpeter, Joseph. 1942. *Capitalism, socialism and democracy*. New York: Harper & Brothers.

Smith, Adam. 1776. *An inquiry into the nature and causes of the wealth of nations*. London: W. Strahan and T. Cadell.

Suarez, Fernando F and James M Utterback. 1995. "Dominant designs and the survival of firms," *Strategic Management Journal*, Vol. 16, No. 6, pp. 415–430.

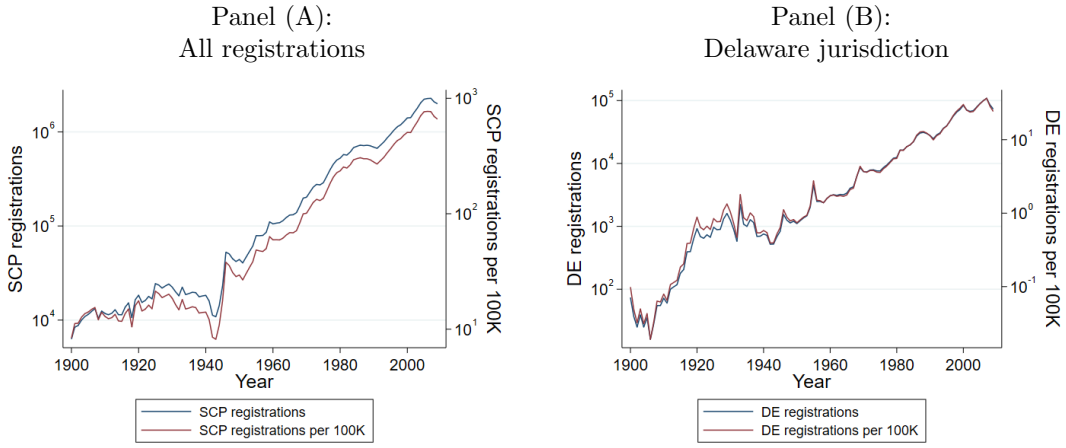
Tadelis, Steven. 1999. "What's in a name? Reputation as a tradeable asset," *American Economic Review*, Vol. 89, No. 3, pp. 548–563.

Tassava, Christopher J. 2008. "The American economy during World War II," in Whaples, Robert ed. *EH.Net Encyclopedia*.

Tushman, Michael L and Philip Anderson. 1986. "Technological discontinuities and organizational environments," *Administrative Science Quarterly*, pp. 439–465.

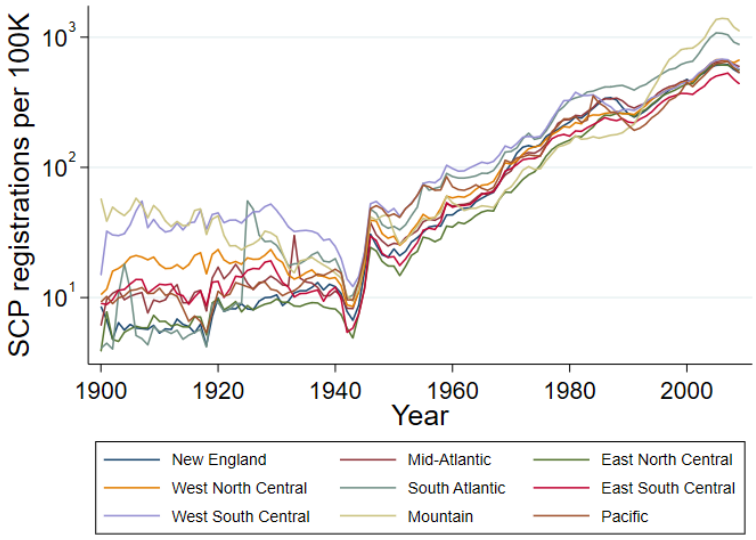
Uzzi, Brian. 1997. "Social structure and competition in interfirm networks: The paradox of embeddedness," *Administrative Science Quarterly*, pp. 35–67.

Figure 1: Business registrations, levels and per capita, 1900-2009



Notes: Panel (A) shows aggregate U.S. business registrations and per-capita registration from 1900 to 2009, in logarithmic scale. Panel (B) shows Delaware jurisdiction registrations. Sample excludes businesses domiciled in three states (AK, HI, NE) for which data are not available for the entire period, as well as those domiciled in Delaware.

Figure 2: Business registrations per 100,000 population, 1900-2009, by census division

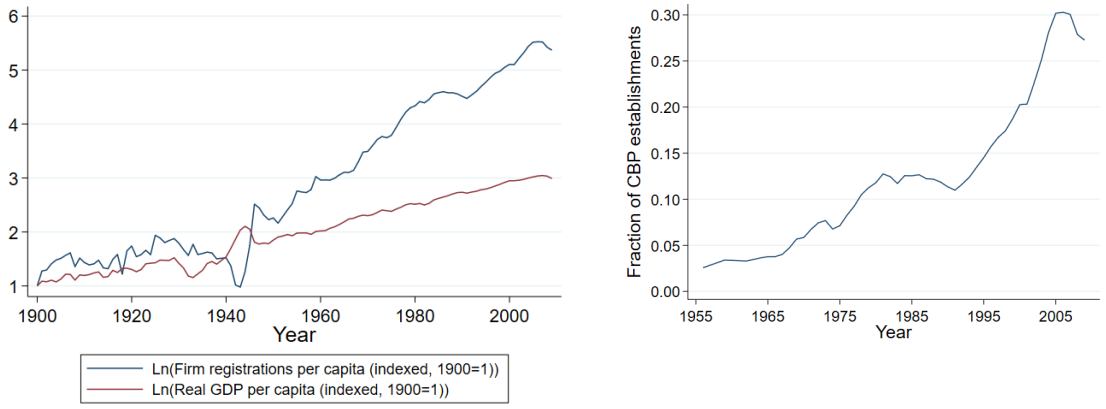


Notes: Figure shows per-capita business registrations from 1900 to 2009 by census division, in logarithmic scale. Sample excludes businesses domiciled in three states (AK, HI, NE) for which data are not available for the entire period, as well as those domiciled in Delaware.

Figure 3: Growth of firm registrations vs. other time series

Panel (A):
Business registrations and GDP per capita, 1900-2009 (log values; indexed; 1900 = 1)

Panel (B):
Business registrations as a share of CBP establishments, 1956-1998

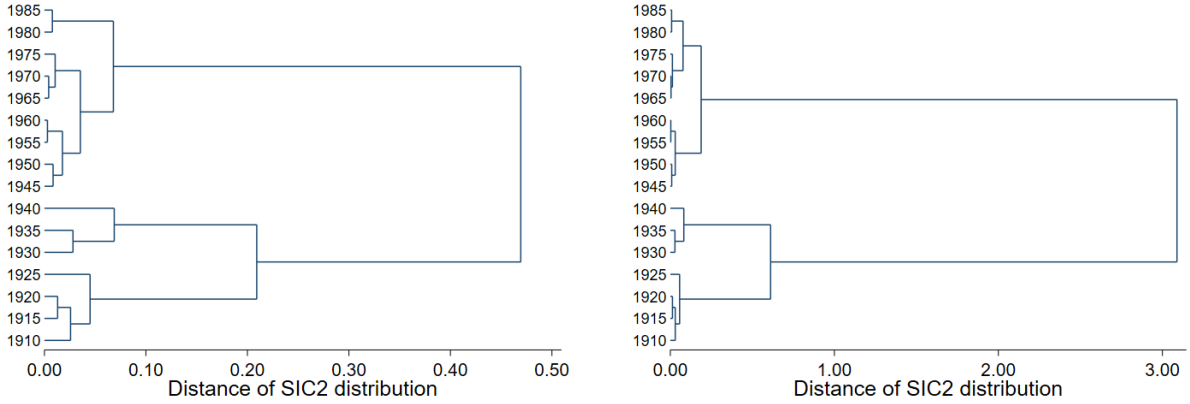


Notes: Panel (A) shows the growth of U.S. business registrations per capita and real GDP since 1900, with both series indexed to 1900 values. Panel (B) shows business registrations as a share of CBP establishments from 1956 onwards. Sample excludes businesses domiciled in three states (AK, HI, NE) for which data are not available for the entire period, as well as those domiciled in Delaware.

Figure 4: Clustering firm founding years on SIC2 similarity, 1910-1989 (5-year intervals; based on firms in Dun & Bradstreet 1990 sample)

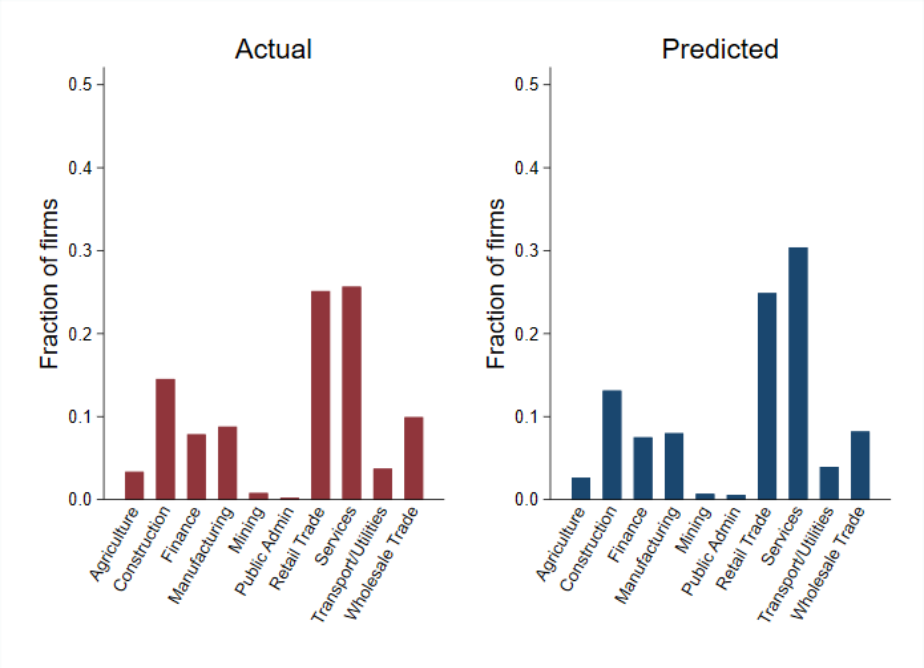
Panel (A):
Average Linkage

Panel (B):
Ward's method



Notes: Figure clusters decades by their similarity in the SIC2 industry distribution of new firms, using the D&B 1990 sample. Results from two different hierarchical clustering methods are shown. Sample excludes businesses domiciled in three states (AK, HI, NE) for which data are not available for the entire period, as well as those domiciled in Delaware.

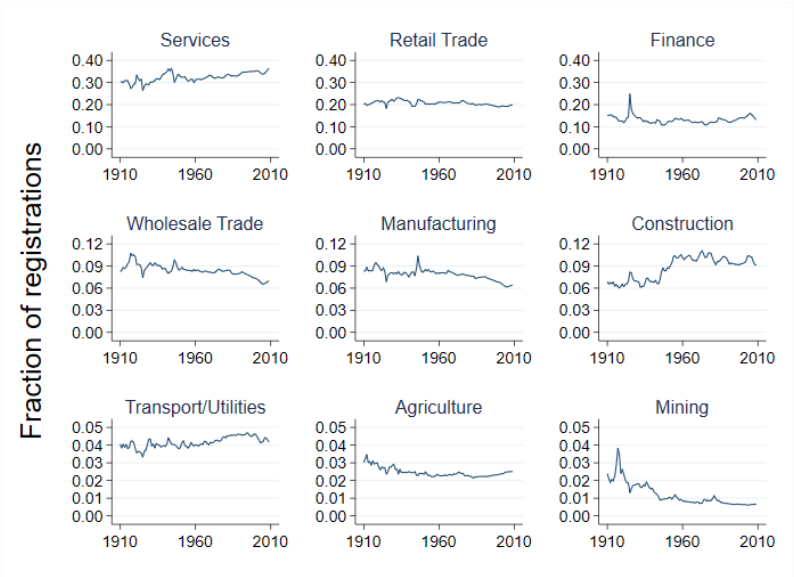
Figure 5: Predicted vs. actual distribution of firms across sectors in D&B testing sample



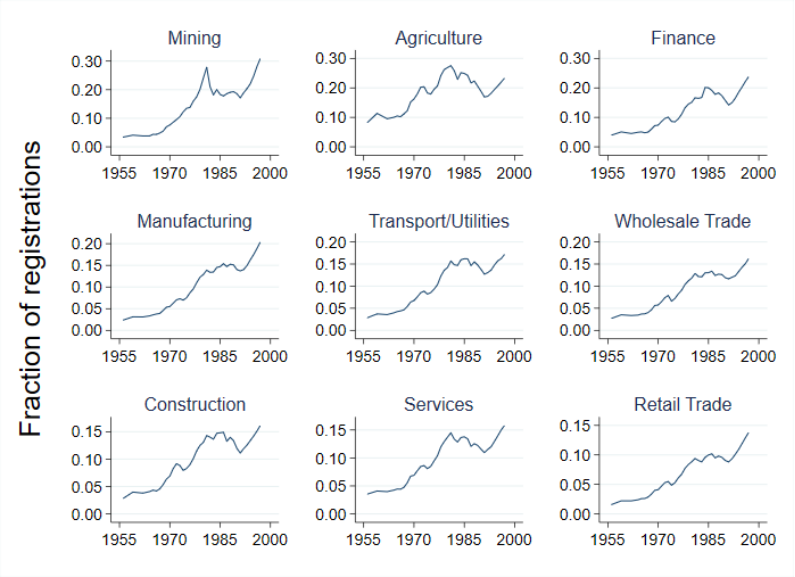
Notes: Figure shows the predicted and actual distribution of sectors in our D&B test sample, using the sector classification algorithm described in the paper.

Figure 6: Composition of business registrations, 1910-2009

Panel (A):
Sector shares of business registrations, 1910-2009



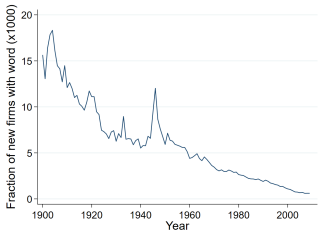
Panel (B):
Business registrations as a fraction of all establishments, 1956-1998



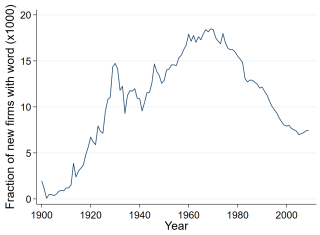
Notes: Panel (A) shows estimated sector shares of U.S. business registrations, over time. Panel (B) shows business registrations as a share of CBP establishments, by sector.

Figure 7: Changing technological orientation of U.S. entrepreneurship post-World War II

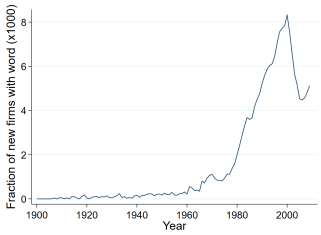
Panel (A):
Frequency of "Manufacturing"



Panel (B):
Frequency of "Service"



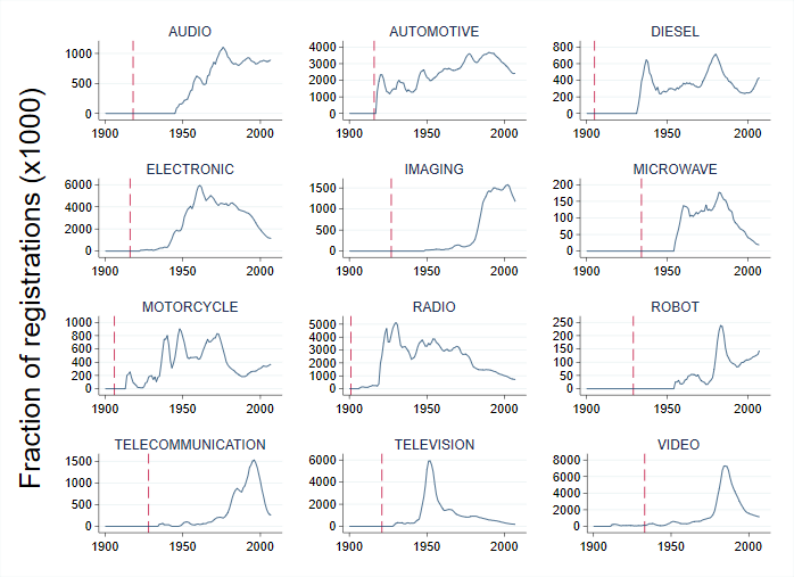
Panel (C):
Frequency of "Technology"



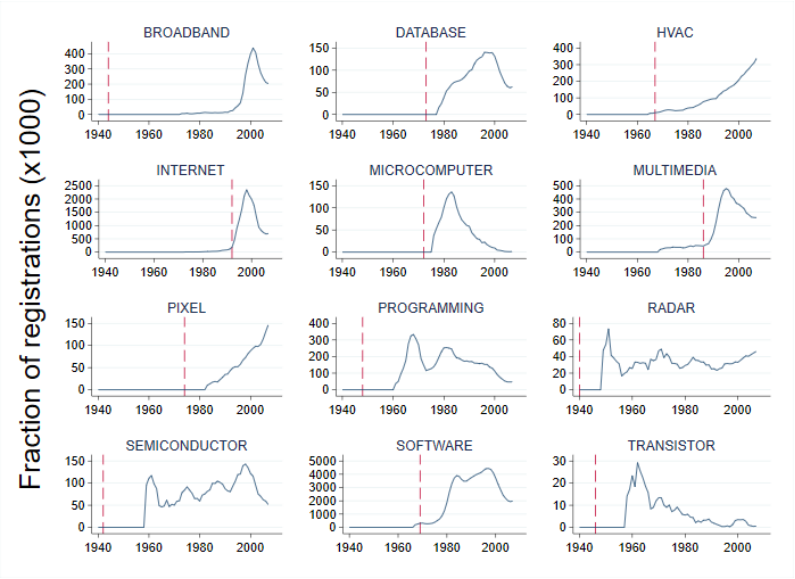
Notes: Figure shows the time series of the frequency of specific words, measured as a fraction of all words in firm names. *Manufacturing* subsumes all words that begin with "Manufactur-" (e.g., Manufacturer, Manufacturers, Manufacturing); *Service*, all words that begin with "Servic-" (e.g., Service, Services, Servicing); *Technology*, all words that begin with "Technolog-" (e.g., Technology, Technologies, Technological).

Figure 8: Diffusion of new technologies into new firm names: select technologies

Panel (A): Pre-1940

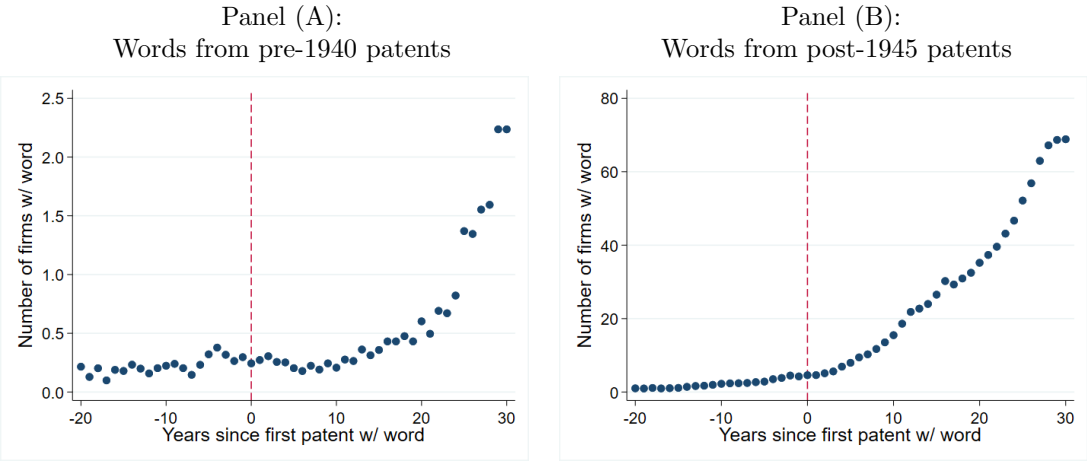


Panel (B): Post-1940



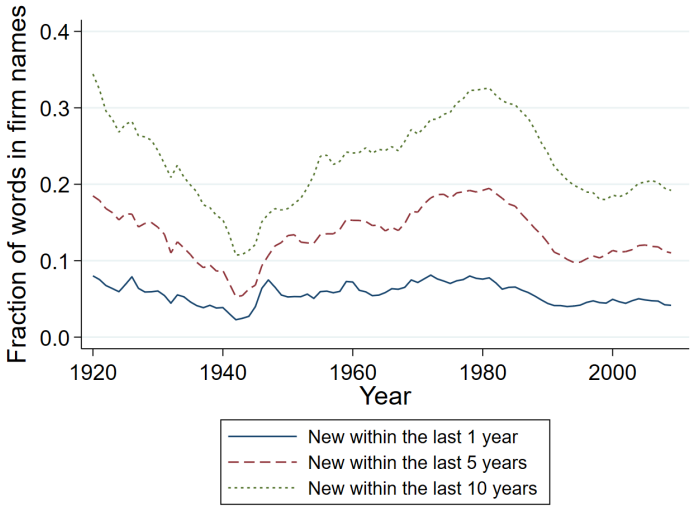
Notes: Figure shows diffusion of specific technologies into the entrepreneurial economy as measured in firm names, for technologies developed before World War II (Panel A) and after (Panel B). Vertical axis measures the share of business registrations in a given year containing the given word. The dotted line in each subfigure marks the filing year of the first patent with that term as a keyword. Figures present 5-year rolling averages to smooth annual fluctuations, and exclude the first ten appearances of each term in firm names, which are sometimes measurement error. See text for further details.

Figure 9: Diffusion of new technologies into new firm names: population averages



Notes: Figure plots the mean diffusion of major technologies into the entrepreneurial economy, as measured in firm names, before and after the year of the technology's first patent, which approximates a date of invention. The sample for these figures comprises keywords from patents granted in the 20th century which were new to the patent record when their first patent granted, and which were in the top 100 new keywords of that decade in their number of subsequent patents. (Put simply, the sample covers 100 major technologies from each decade between 1900 and 2000.) Panels (A) and (B) divide the sample into keywords from before and after World War II. Figures exclude the first ten appearances of each term in firm names, which are sometimes measurement error. See text for further details.

Figure 10: Vintage of words in firm names



Notes: Figure shows time series measuring what fraction of words used in firm names each year were new to (i.e., never-before used in) firm names.

Table 1: Growth rate of firm registrations and GDP per capita, 1900-2009

	Average growth rate		
	1900-1929	1946-2009	1900-2009
Registrations p.c.	0.018	0.055	0.043
	(0.002)	(0.001)	(0.001)
Real GDP p.c.	0.015	0.021	0.020
	(0.002)	(0.001)	(0.001)
R^2	0.78	0.99	0.93
Difference (p-val)	0.271	0.000	0.000

Notes: Table estimates long-run average business registration and real GDP per capita growth rates. We do so over three periods: 1900-1929, 1946-2009, and the full 1900-2009 sample. p-values reported for a test of equal growth rates. Sample excludes businesses domiciled in three states (AK, HI, NE) for which data are not available for the entire period, as well as those domiciled in Delaware.

Table 2: Most common words and bigrams in firm names, by decade

Most common words in new firm names, by decade

Rank	1900s	1910s	1920s	1930s	1940s
1	OIL	OIL	OIL	CLUB	CLUB
2	CHURCH	GAS	CLUB	OIL	CHURCH
3	MINING	CHURCH	REALTY	REALTY	REALTY
4	LUMBER	CLUB	BUILDING	SERVICE	SERVICE
5	BANK	BANK	CHURCH	CHURCH	COUNTY
6	MANUFACTURING	MINING	LOAN	COUNTY	SUPPLY
7	CLUB	LUMBER	INVESTMENT	SUPPLY	CONSTRUCTION
8	LAND	REALTY	MOTOR	MOTOR	PRODUCTS
9	TELEPHONE	MANUFACTURING	SERVICE	PRODUCTS	OIL
10	GAS	COAL	LUMBER	INVESTMENT	MANUFACTURING

Rank	1950s	1960s	1970s	1980s	1990s
1	CLUB	SERVICE	SERVICE	CONSTRUCTION	MANAGEMENT
2	CHURCH	CLUB	CONSTRUCTION	SERVICE	CONSTRUCTION
3	REALTY	CONSTRUCTION	DEVELOPMENT	MANAGEMENT	SERVICE
4	SERVICE	CHURCH	REALTY	DEVELOPMENT	PROPERTIES
5	CONSTRUCTION	REALTY	CLUB	CENTER	DEVELOPMENT
6	DEVELOPMENT	DEVELOPMENT	CENTER	REALTY	CENTER
7	SUPPLY	SUPPLY	MANAGEMENT	PROPERTIES	FAMILY
8	COUNTY	CENTER	INDUSTRIES	AUTO	INVESTMENTS
9	INVESTMENT	COUNTY	CHURCH	CLUB	REALTY
10	OIL	PRODUCTS	AUTO	INVESTMENTS	AUTO

Most common bigrams in new firm names, by decade

Rank	1900s	1910s	1920s	1930s	1940s
1	OIL_GAS	OIL_GAS	BUILDING_LOAN	OIL_GAS	BAPTIST_CHURCH
2	STATE_BANK	STATE_BANK	OIL_GAS	SOCIAL_CLUB	SOCIAL_CLUB
3	STREET_RAILWAY	BUILDING_LOAN	SOCIAL_CLUB	LEGION_DEPARTMENT	VETERANS_FOREIGN
4	MINING_MILLING	PRESBYTERIAN_CHURCH	BAPTIST_CHURCH	CHAMBER_COMMERCE	CHAMBER_COMMERCE
5	BUILDING_LOAN	BAPTIST_CHURCH	STATE_BANK	CREDIT_UNION	LUTHERAN_CHURCH
6	BAPTIST_CHURCH	LIGHT_POWER	REAL_ESTATE	BAPTIST_CHURCH	FOREIGN_UNITED
7	GOLD_MINING	SOCIAL_CLUB	PLEASURE_CLUB	INSURANCE_AGENCY	VOLUNTEER_FIRE
8	CHURCH_LATTER	LUTHERAN_CHURCH	LUTHERAN_CHURCH	FUNERAL_HOME	METHODIST_CHURCH
9	PRESBYTERIAN_CHURCH	METHODIST_CHURCH	CHAMBER_COMMERCE	POST_LEGION	FIRE_DEPARTMENT
10	REAL_ESTATE	TOWNSHIP_ELECTRIC	ATHLETIC_CLUB	ATHLETIC_CLUB	KIWANIS_CLUB

Rank	1950s	1960s	1970s	1980s	1990s
1	BAPTIST_CHURCH	BAPTIST_CHURCH	INSURANCE_AGENCY	REAL_ESTATE	REAL_ESTATE
2	INSURANCE_AGENCY	INSURANCE_AGENCY	REAL_ESTATE	INSURANCE_AGENCY	INSURANCE_AGENCY
3	METHODIST_CHURCH	REAL_ESTATE	BAPTIST_CHURCH	BAPTIST_CHURCH	HEALTH_CARE
4	CHAMBER_COMMERCE	METHODIST_CHURCH	AIR_CONDITIONING	OIL_GAS	BAPTIST_CHURCH
5	LUTHERAN_CHURCH	PLUMBING_HEATING	PLUMBING_HEATING	HEALTH_CARE	PROPERTY_MANAGEMENT
6	KIWANIS_CLUB	SOCIAL_CLUB	SOCIAL_CLUB	AIR_CONDITIONING	AIR_CONDITIONING
7	SOCIAL_CLUB	REALTY_TRUST	OIL_GAS	PLUMBING_HEATING	IMPORT_EXPORT
8	VOLUNTEER_FIRE	LUTHERAN_CHURCH	SERVICE_CENTER	CARE_CENTER	AUTO_REPAIR
9	UNITED_METHODIST	AIR_CONDITIONING	FUNERAL_HOME	AUTO_BODY	ASSET_MANAGEMENT
10	FIRE_DEPARTMENT	UNITED_METHODIST	CARE_CENTER	PROPERTY_MANAGEMENT	CAPITAL_MANAGEMENT

Notes: Table lists the most common words and bigrams in firm names in each decade shown in the column headings.

Table 3: Most common new bigrams in firm names, by decade

Most common new bigrams in new firm names (business registrations)

Rank	1950s	1960s	1970s	1980s
1	CAPITAL_MANAGEMENT	LAW_FIRM	LAWN_CARE	HOME_SOLUTIONS
2	INVESTMENT_PROPERTIES	BUSINESS_SOLUTIONS	LAW_OFFICE	PROPERTY_SOLUTIONS
3	LEARNING_CENTER	CONSTRUCTION_MANAGEMENT	HOME_INSPECTIONS	AFFORDABLE_HOUSING
4	HOME_IMPROVEMENTS	PROPERTY_INVESTMENTS	HOME_INSPECTION	MARKETING_SOLUTIONS
5	ESTATE_INVESTMENTS	PROFESSIONAL_LAW	SELF_STORAGE	TOWING_RECOVERY
6	LAWN_SERVICE	FAMILY_INVESTMENTS	ASSISTED_LIVING	ENERGY_SOLUTIONS
7	LIMOUSINE_SERVICE	MANAGEMENT_CONSULTING	MEDICAL_BILLING	ESTATE_SOLUTIONS
8	MEDICAL_EQUIPMENT	MEDICAL_MANAGEMENT	HAIR_SALON	CUTTING_EDGE
9	EQUIPMENT_LEASING	PROFESSIONAL_MEDICAL	MANAGEMENT_SOLUTIONS	NETWORK_SOLUTIONS
10	HOUSING_DEVELOPMENT	TECHNOLOGY_SOLUTIONS	SURGERY_CENTER	SPORTS_BAR

Most common new bigrams in new firm names (D&B 1990 sample)

Rank	1950s	1960s	1970s	1980s
1	HAIR_DESIGN	LAWN_CARE	HALLMARK_SHOP	HOUR_PHOTO
2	PRO_SHOP	GOLF_SHOP	HAIR_STUDIO	VIDEO_WORLD
3	JANITORIAL_SERVICE	HAIR_SALON	HOME_VIDEO	MAIL_BOXES
4	ONE_HOUR	LIMOUSINE_SERVICE	FITNESS_CENTER	FROZEN_YOGURT
5	JOINT_VENTURE	QUICK_STOP	CONVENIENCE_STORE	HOME_MEDICAL
6	TRUCK_REPAIR	SCREEN_PRINTING	SIR_SPEEDY	FAMILY_VIDEO
7	OFFICE_PRODUCTS	FRIED_CHICKEN	SPEEDY_PRINTING	VIDEO_PLUS
8	FAMILY_RESTAURANT	LEARNING_CENTER	VIDEO_PRODUCTIONS	PHOTO_EXPRESS
9	LAWN_SERVICE	COMPUTER_SERVICE	MEDICINE_SHOPPE	COMPUTER_SOLUTIONS
10	APPLIANCE_REPAIR	HAIR_DESIGNS	VIDEO_CENTER	CARE_MEDICAL

Notes: Table lists the most common new bigrams in firm names in each decade shown in the column headings, first among business registrations, and then for the 1990 D&B sample (comprised of firms in the D&B data—broadly, firms which filed an application for business credit and survived to the data year).

Table 4: Fraction of states with a related startup after 5, 10, 20, and 30 years

	All firms			
	(1)	(2)	(3)	(4)
	5 years	10 years	20 years	30 years
1(Decade==1910s)	-0.003	0.001	-0.000	0.000
	(0.007)	(0.009)	(0.014)	(0.019)
1(Decade==1920s)	0.002	0.003	-0.003	0.001
	(0.009)	(0.010)	(0.014)	(0.019)
1(Decade==1930s)	-0.003	-0.004	0.001	0.022
	(0.007)	(0.009)	(0.015)	(0.022)
1(Decade==1940s)	0.006	0.015	0.037*	0.072***
	(0.009)	(0.013)	(0.019)	(0.026)
1(Decade==1950s)	0.035**	0.051**	0.097***	0.159***
	(0.017)	(0.020)	(0.026)	(0.033)
1(Decade==1960s)	0.036**	0.062***	0.119***	0.172***
	(0.017)	(0.021)	(0.028)	(0.034)
1(Decade==1970s)	0.108***	0.154***	0.232***	0.300***
	(0.027)	(0.032)	(0.039)	(0.044)
1(Decade==1980s)	0.116***	0.159***	0.241***	0.000
	(0.025)	(0.028)	(0.034)	(.)
1(Decade==1990s)	0.172***	0.228***	0.000	0.000
	(0.024)	(0.027)	(.)	(.)
N	1092	1092	929	820
R^2	0.12	0.15	0.15	0.14
Y mean				

Notes: Table estimates the fraction of states with a firm related to each technology after 5, 10, 20, and 30 years (Columns 1 to 4, respectively), as a function of invention decade. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Robust SEs in parentheses.

Web Appendix

A Data Appendix

A.1 Business registration data

PENDING

A.2 Dun & Bradstreet data

PENDING

A.3 Infogroup USA data

PENDING

A.4 Other supporting datasets

A.4.1 Patent data

The patent-level dataset used in this paper begins with the USPTO historical master file (Marco et al. 2015), which provides a master list of granted patents with grant dates, patent class/subclass (USPC), and two-digit NBER category (Hall et al. 2001).

We obtain from Google BigQuery three other patent-level variables for all patents in this sample: filing dates, titles, and "top terms", which, according to Google, are "the top 10 salient terms extracted from the patent's title, abstract, claims, and description" (no further documentation is available on how—e.g., via what algorithm— these terms are derived, but visual browsing suggest they are sensible, and they are what Google itself uses to identify related patents). These terms effectively serve as keywords, for our purposes in this paper, and are what we use to identify and label technologies which could potentially appear in firm names.

We use several sources to compile the historical network of front-page patent citations. These sources are FreePatentsOnline.com, Derwent Innovation, the Berkes CUSP dataset (Berkes 2018), and PatentsView. The data in these sources mostly coincide in the years they overlap, but we use their union to ensure complete measurement. Finally, we merge in patent-level measures developed in prior research, including Kogan et al. (2017) patent values, which are available for patents granted to public companies from 1926 onwards; Kelly et al. (2021) breakthrough patent indicators; and Ganglmair et al. (2021) measures of product and process claims, which can be used to try to distinguish between product and process inventions.

A.4.2 Historical U.S. County Business Patterns data

The U.S. County Business Patterns (CBP) data used in this paper cover the 1956 to 2009 period. The CBP measures establishments, employment, and wages at the level of states, counties and industries, and was published at irregular intervals from 1946 to 1964 and annually thereafter. PDFs of complete printed volumes are available for most years from HathiTrust, and electronic CBP data are available from NARA for 1969 onwards. Although CBP was first administered in 1946, we begin our CBP data series in 1956 because this was the first year in which all U.S. counties were surveyed, such that the data represent the universe of U.S. business establishments. The CBP data used in this paper were obtained from three sources. Because the analysis in this paper is at the state and sector level, we retrieved or compiled data at state x sector level.

- 1956-1969, 1971, 1973: Manual transcription.
- 1970, 1972, 1974: U.S. National Archives (see https://catalog.archives.gov/id/613576). NARA makes available for download historical CBP data files spanning 1967-2007. We use the state-level data files from 1970, 1972, and 1974. Because the pre-1970 data are incomplete, and the 1971 and 1973 records are missing state files, we had these data transcribed from these years' print CBP publications. Later years are covered next.
- 1975-2009: Eckert et al. (2021) (see http://fpeckert.me/cbp/). We use the CBP data collected and published by Eckert et al. (2021) for remaining years.

We identify sectors in the CBP data as the categories that subsume two-digit SIC industries, and are denoted in the CBP data as numbers with two trailing hyphens.

Our sector-code crosswalk is as follows:

- 07--: Agriculture, Forestry, & Fishing
- 10--: Mining
- 15--: Construction
- 19--: Manufacturing (through 1985)
- 20--: Manufacturing (1986 onwards)
- 40--: Transportation & Utilities
- 50--: Wholesale trade
- 52--: Retail Trade
- 60--: Finance, Insurance, & Real Estate
- 70--: Services

In 1998, the CBP transitioned from the SIC to NAICS classification. Sector definitions persist, but some industries were reallocated across sectors. Where we invoke the sectoral CBP data in Section 4, we thus cut our analysis at 1997.

A.4.3 Macroeconomic and population variables

We use GDP data from the Jordà-Schularick-Taylor MacroHistory Database (Jordà et al. 2017) where we compare business registration growth to GDP growth in Section 4. We obtain national and state-level population series from FRED (https://fred.stlouisfed.org/).

B Sector Classification Method

In this appendix, we describe our sector classification procedure in detail, and provide supplementary validation results to those in the body of the paper. Portions of this section may reproduce passages in the paper, adding detail and context. We begin as follows.

Let i index firms, j tokens, s sectors. Each firm i has J_i tokens in its name, and there are J unique tokens in the corpus. Let τ_{js} represent a token-sector score, which will be defined below as the fraction of token j's total occurrences that are in each sector s. φ_{is} will represent a firm-sector score, which will be an average of τ_{js} across all J_i tokens in firm i's name. τ_{js} and φ_{is} will thus measure each token or firm's association with each of $S=10$ sectors. To aid the exposition below, let n_{js} represent the number of times token j is in the name of firms in sector s, and $N_J=\sum_{s=1}^{10} n_{js}$ represent the number of firms in which token j appears overall.

B.1 Measuring token-sector scores

We first undertake several steps to prepare the D&B data for training and validating our classification procedure. We begin by cleaning firm names of stop words, special characters, tokens of length $\leq$2, and all-numeric tokens. We then split out the tokens in firm names, and reshape the firm-level data to a token-level dataset, with one observation per firm-token. Because the D&B company name field is fixed in width, words in firm names are sometimes abbreviated. We thus extend our data cleaning effort by creating an extensive, part-manual and part-automated crosswalk from tokens which appear to be abbreviations or minor misspellings to full words. A total of roughly 65,000 abbreviated tokens get crosswalked to full words through this approach. For example, "IMPRVMNT" is updated to "IMPROVEMENT", and "PIZZARIA" to "PIZZERIA". We believe these changes ultimately increase the fidelity of our data to the underlying economic reality and will improve the quality of our sector classification. A total of 65,000 unique tokens in the D&B 1990 sample are revised through these methods, out of roughly 10 times as many tokens in the data. The three most common words in this sample are "SERVICE", "CONSTRUCTION", and "SHOP"—words which are (seemingly) immediately revealing of the sector that a firm with those words is in, and illustrate the face validity of our proposed approach.

We take two approaches to further reducing (or not) the set of tokens which we will use to predict firms' economic sector. In the first variant, we deploy this set of tokens as is. In our second variant, we remove additional stop words, excluding names which are *not* an English-language word but are any of (i) a word in a U.S. state name, (ii) a word in a U.S. city name, (iii) a common given name, or (iv) a common surname.[1] We exclude these tokens because we think they are both frequently

[1] Our sample of given names consists of the union of the 1,000 most common baby names for birth years 1880 to 2009, according to the Social Security Administration, obtained from https://github.com/hadley/data-baby-names. Our sample of surnames consists of all surnames which appeared $\geq$100 times in the 2010 census, obtained from https://www.census.gov/topics/population/genealogy/data/2010_surnames.html. In addition to excluding these names,

occurring in firm names, due to geographic specialization or eponymy, and because they are sector-agnostic and thus may obscure our sectoral predictions. The resulting dataset has firm-tokens and their associated sector, obtained from each firm's D&B-reported SIC code.

We then measure τ_{js} as sector s's share of all uses of token j: $\tau_{js} = \frac{n_{js}}{N_J}$. Although we experimented with other approaches to scoring tokens' associations with each sector, we found that this simple approach outperformed others across many validation tests.[2] By construction, each token's sector-level scores τ_{js} will add to 1 when summed across sectors.

B.2 Measuring firm-sector scores

To get from token-sector scores to firm-sector scores (our target output), we need to aggregate across tokens in firm names—reducing the dimensionality of each firm i from $J_i \times 10$ to 10. We experimented with two approaches: straight averages and weighted averages, weighting by (the square root of) a token's total usage. This latter approach will overweight more-common words, whose sector associations we might measure more precisely. Formally, these two approaches to measuring the firm-sector score φ_{is} can be characterized as follows:

$$\textit{Approach 1}: \qquad \varphi_{is}^{uwtd} = \frac{1}{J_i} \sum_{j=1}^{J_i} \tau_{js}$$

$$\textit{Approach 2}: \qquad \varphi_{is}^{wtd} = \frac{1}{\sum_{j=1}^{J_i} \sqrt{N_j}} \sum_{j=1}^{J_i} \sqrt{N_j} \tau_{js}$$

To illustrate, suppose there are two tokens (A and B) and two sectors (1 and 2). To make this example concrete, let us specifically imagine a firm with the name "Master Plumbing", and that there are two sectors, *Construction* and *Retail Trade*. We don't ex-ante know what sector the firm is in. It might be a construction contractor. It might be a plumbing fixture store. But the words in its name give us clues—especially when we can gauge how often these words associate with each of these sectors in our training data. For this example, let us denote "Master" and "Plumbing" as tokens A and B, and *Construction* and *Retail Trade* as sectors 1 and 2.

Suppose that in our training sample, token A ("Master") is used in 225 firm names, and appears 50% of the time in sector 1 (*Construction*) and 50% in sector 2 (*Retail Trade*). Token B ("Plumbing"), on the other hand, is in 25 firm names, and appears 90% of the time in sector 1 and 10% in sector 2. Table B.1 provides the computed sector scores under each of these two approaches. In practice, we find that the simpler aggregation method (unweighted averages) better predicts the true distribution, which makes this our preferred approach.

we also exclude variants of these names plus the letter 'S', to account for the sizable number of businesses named in the possessive (e.g., "MIKE'S FOOD & SPIRITS" or "DAVE'S FRESH PASTA").

[2] An alternative approach we explored was to square these shares, to overweight high token-sector associations. Doing so, however, reduced the rate at which we correctly predicted firms' actual sector.

Table B.1: Firm-sector scores for "Master Plumbing" (contrived example)

	Sector 1	Sector 2
Approach 1: Unweighted	$\varphi_{i1} = \frac{0.5+0.9}{2} = 0.7$	$\varphi_{i2} = \frac{0.5+0.1}{2} = 0.3$
Approach 2: Weighted	$\varphi_{i1} = \frac{\sqrt{225}\cdot 0.5+\sqrt{25}\cdot 0.9}{\sqrt{225}+\sqrt{25}} = 0.6$	$\varphi_{i2} = \frac{\sqrt{225}\cdot 0.5+\sqrt{25}\cdot 0.1}{\sqrt{225}+\sqrt{25}} = 0.4$

For a real example from our data, consider the firm "Anderson Home Appliances", whose words can be fractionally classified across each of the original 10 economic sectors. In our data, these words associate with sectors as shown in Table B.2 (where only the top five sectors are shown), leading to the firm-level scores shown in the last row of the table. Our procedure predicts *Retail Trade* and *Services* as equally likely sectors. The actual primary sector reported for this firm in the D&B data is *Services*, and the secondary sector reported in D&B is *Retail Trade*—though we might infer that Anderson Home Appliances is related to *Construction* too.

Table B.2: Firm-sector scores for "Anderson Home Appliances" (true example)

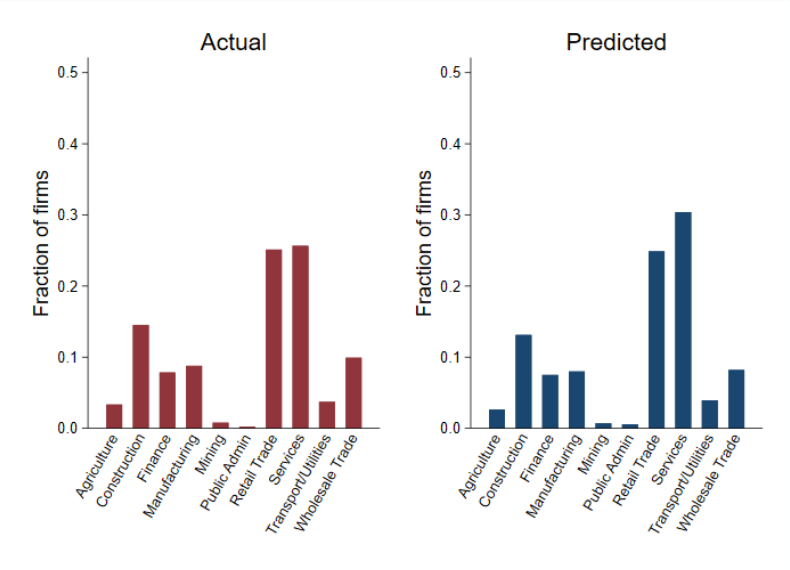
	Token-sector scores and cross-token averages				
Word	Agriculture	Construction	Fin./Ins./RE	Retail Trade	Services
Anderson	8%	17%	10%	14%	34%
Home	0%	20%	10%	17%	48%
Appliances	0%	2%	0%	71%	21%
Average	3%	13%	7%	34%	34%

B.3 Validating our procedure

We validate this sector classification procedure in several ways. We begin by validating within our D&B test sample, where we can classify firms into sectors and compare our predictions to what is reported in the data. We focus our validation evidence on our more restrictive variant, where we remove place and person names from firm names before classifying sectors.

Figure B.1 begins with a high level view, aggregating sector shares across the sample and showing the overall frequency of firms in each of our 10 sectors as observed (left panel) and as predicted (right panel). The distributions line up closely, though we slightly overpredict *Services* and slightly underpredict *Wholesale Trade*. Table B.3 provides a more nuanced view of our predictive validity, showing, for each sector (first column), how often firms predicted to be in that sector are in fact reported by D&B to be in that sector (second column), as well as how often in each such case our second-most likely prediction is the D&B-reported sector (third column). The final column shows the sum: the probability that one of the top two predicted sectors is the reported sector. Across all predicted sectors, we match D&B 70 to 90% of the time with our top one or two predictions. We consider these rates to be high, particularly as we do not expect to match D&B 100% of the time, given that many firms span boundaries between sectors.

Figure B.1: Predicted vs. actual distribution of firms across sectors in D&B testing sample



Notes: Figure shows the predicted and actual distribution of sectors in our D&B test sample, using the sector classification algorithm described in the paper.

Table B.3: Sector classification algorithm: Predictive performance, by sector

	Fraction of firms where:		
Predicted sector	Prediction correct	Runner-up prediction correct	Total
Agriculture, Forestry, & Fishing	78%	9%	87%
Construction	84%	7%	91%
Finance, Insurance, & Real Estate	82%	8%	90%
Manufacturing	67%	15%	82%
Mining	58%	13%	71%
Public Administration	81%	9%	91%
Retail Trade	73%	13%	86%
Services	66%	16%	83%
Transportation & Public Utilities	84%	6%	91%
Wholesale Trade	57%	21%	78%
Overall	71%	14%	85%

Notes: Table shows, for each sector (first column), how often firms predicted to be in that sector (i.e., whose top predicted sector is that shown in the left column) are in fact reported by D&B to be in that sector (second column), as well as how often in each such case our second-most likely prediction is the D&B-reported sector (third column).

In Table B.4, we break these patterns down even further. Here we document, for each top-predicted sector (left column), the fraction of firms with that prediction that are reported by D&B in each of our 10 sectors. The first features of this table to observe is that the mass is overwhelmingly concentrated along the diagonal (which reproduces Column 2 of Table B.3). In addition, the table also illustrates the sectors that are likely to be jointly present—or incorrectly classified, depending on interpretation. Firms which we predict to be in *Wholesale Trade* (row 10), for example, are 18% of the time actually reported as being in *Retail Trade*—illustrating that firms may sometimes be a bit of each sector, and the occasional challenging of distinguishing the two.

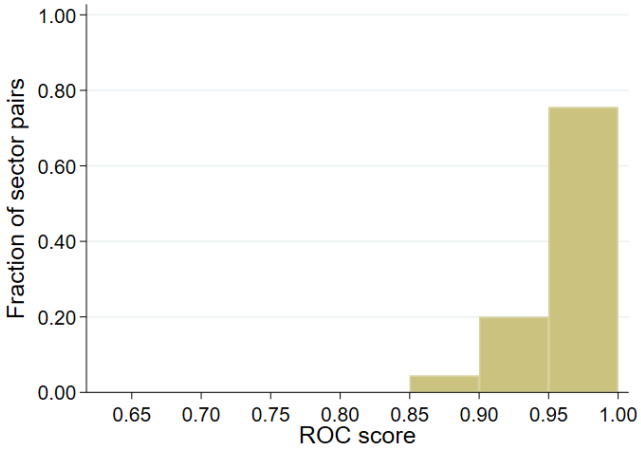
Table B.4: Sector classification algorithm performance: Predicted vs. actual sector

When top predicted sector is...		Pr(Actual sector is...)									
Sector	Description	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	Agriculture, Forestry, & Fishing	**78%**	2%	2%	1%	0%	0%	7%	5%	1%	4%
2	Construction	1%	**84%**	1%	3%	1%	0%	3%	4%	1%	3%
3	Finance, Insurance, & Real Estate	1%	4%	**82%**	1%	0%	0%	3%	7%	1%	2%
4	Manufacturing	1%	4%	1%	**67%**	0%	0%	7%	9%	1%	11%
5	Mining	1%	4%	4%	4%	**58%**	0%	9%	6%	3%	12%
6	Public Administration	1%	1%	4%	1%	0%	**81%**	2%	6%	3%	2%
7	Retail Trade	2%	2%	2%	5%	0%	0%	**73%**	9%	1%	7%
8	Services	2%	4%	4%	4%	0%	1%	11%	**66%**	2%	5%
9	Transportation & Public Utilities	1%	2%	1%	1%	0%	0%	3%	5%	**84%**	2%
10	Wholesale Trade	2%	3%	1%	8%	2%	0%	18%	7%	2%	**57%**

Notes: Table shows, for each sector, how often firms predicted to be in that sector (i.e., whose top predicted sector is that shown in the left column) are fact reported by D&B to be in each of the ten sectors. The diagonal reproduces the rates from the second column of Table B.3.

As another check, we make all unique pairs of sectors, and for each pair, we filter to firms whose reported sector is one of the two and use receiver-operating characteristic (ROC) analysis to evaluate for how many firms the more probable sector is the reported one. Figure B.2 plots the distribution of ROC scores among these pairs. The scores are near one, indicating that in each of these two-sector horseraces, we nearly always predict the D&B-reported sector.

Figure B.2: Distribution of ROC scores for prediction in sector pairs in D&B testing sample

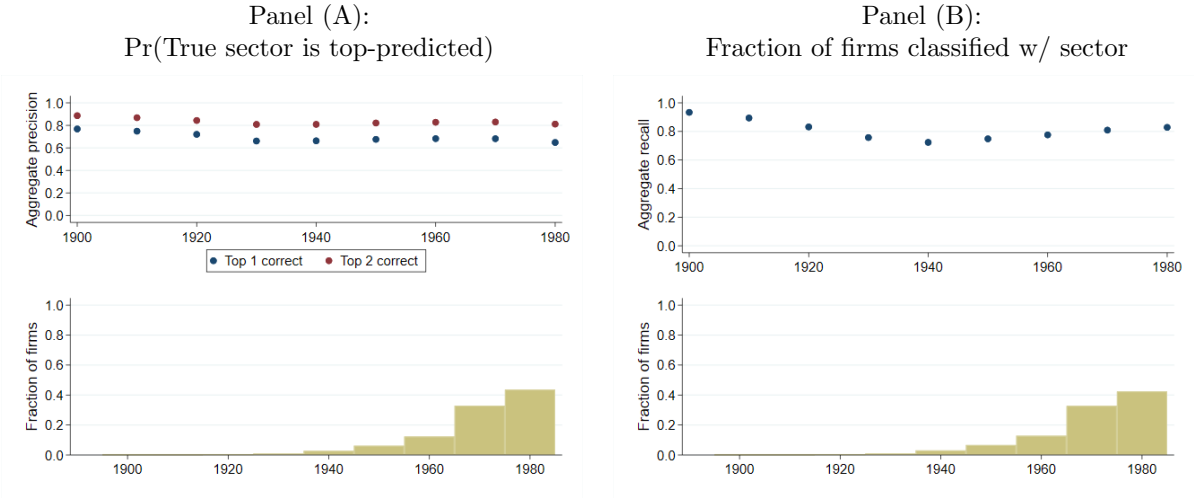


Notes: Figure shows distribution of ROC scores over sector pairs, evaluating across our test sample how often the more-predicted sector in each pair is the firm's reported sector, conditional on one of the two sectors in the pair being the true sector. With 10 sectors, the distribution in this figure is over $\sum_{s=1}^{9} s$ pairs. In the vast majority of these pairings, our algorithm predicts a firm's true sector over the alternative >95% of the time.

In additional validation tests, we examine the stability of this performance across time, for firms with different D&B-reported founding years, and we find that our procedure's performance is stable in both the rate at which its predictions match D&B-reported sectors, and in the fraction of firms that get classified to a sector (Figure B.3). We also perform validation tests against an independent third sample of firms in the Infogroup USA 2000 data extract. The performance of our procedure

in this sample is similar to that in our D&B test sample, suggesting that the results above are not driven by distinctive shared features of the D&B training and test samples.

Figure B.3: Precision of sector predictions vs. rate of prediction in D&B testing sample



Notes: The top chart in each panel shows how precision and recall rates vary over time (across firm registration decade) in the D&B testing sample. Bottom panel shows the distribution of firms in this sample across decades. Aggregate precision and recall is a weighted average of those in the upper scatterplots, weighted by the density in the lower histograms.

B.4 Classifying the business registration sample

Our final step is to classify firms in our business registration data to economic sectors. We clean firm names by the same approach applied to D&B firms in Section B.1, up to expanding abbreviations and correcting minor misspellings, and again remove place and person names that are not English-language words. We then apply the token-sector scores calculated in Section B.1 to these firms' names, and aggregate to the firm level following Section B.2. The resulting dataset has sector scores for each registered firm, across each of ten economic sectors (*Agriculture* to *Wholesale Trade*). Recall that these scores will add to one, and can alternatively be interpreted as the probability that a firm is in a given sector, or the firm's fractional association with that sector—which allows firms to span sectoral boundaries, as many firms do in practice.

This classification nevertheless bears limitations. One is that some words in the set of firm names in the registration sample may not be present in the Dun & Bradstreet sample, and thus will be unable to be scored. Due to this, as well as the removal of stop words, some firms in the registration data may not be able to have their sector predicted. We are reassured, however, by the fact that even under our more restrictive sector classification procedure, we classify 90% of firms—and this rate is stable over time. We believe this is high enough to generate useful measures of the sectoral distribution of U.S. entrepreneurship over long periods.

C Supplementary Results

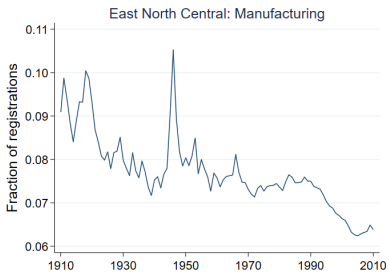
C.1 Regional patterns in entrepreneurship

In Figure C.1, we highlight two specific stories we find embedded in our data. The first is long-run changes in the strength of the *Manufacturing* sector in the East North Central census division (Ohio, Indiana, Illinois, Michigan, and Wisconsin)—what might be traditionally viewed as the Rust Belt. Panel (A) shows a long-run decline in manufacturing's share of new businesses, dating back at least a century, including a material post-1990 decline, consistent with the "China shock" that a large literature has recently documented (e.g., Autor et al. 2013). What the figure illustrates, however, is that reference dates matter. The decline in Midwestern manufacturing startups as a share of the regional economy is not only a postwar, or post-1990, phenomenon: it has been going on for a century. Moreover, our long series—the first of its kind—shows us that the late 1940s is an artificial reference point that will overstate the decline of Midwest manufacturing, since the World War II years produced an explosion of manufacturing businesses in the region, presumably to serve government demand. Recognizing that manufacturing entrepreneurship declined rapidly in the first half of the twentieth century raises new questions for research on how this reallocation might have affected a number of different regional outcomes.

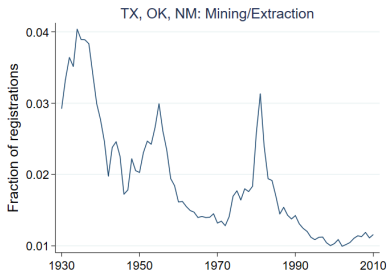
Figure C.1, Panel (B) highlights a different story: boom-and-bust cycles in the oil and gas industry in the major oil-producing states of the postwar era: Texas, Oklahoma, and New Mexico. What is especially striking is the more than doubling of new mining/extraction businesses during and after the oil crises of the 1970s. This evidence, coupled with the previously-discussed spike in real estate registrations seen in 1925 in Figure 6, illustrates how closely new business activity follows sectoral economic cycles, including transient shocks and manias.

Figure C.1: Regional entrepreneurship in select sectors

Panel (A):
Manufacturing in the Rust Belt



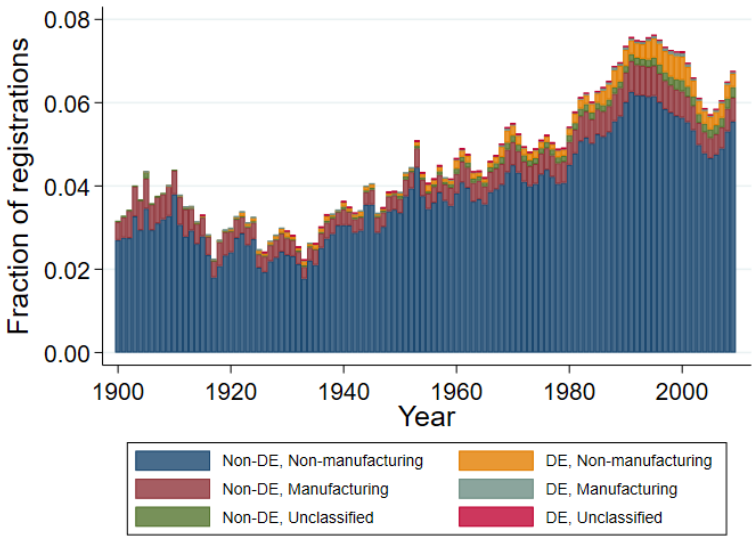
Panel (B):
Mining/Extraction in Oil-producing States



Notes: Panel (A) shows the *Manufacturing* sector's share of new businesses over time in the East North Central census division, which includes several states traditionally viewed as the U.S. Rust Belt. Panel (B) shows the *Mining and Extraction* sector's share of registrations in the country's major oil-producing states (TX, OK, NM).

C.2 Level and composition of technology-based entrepreneurship

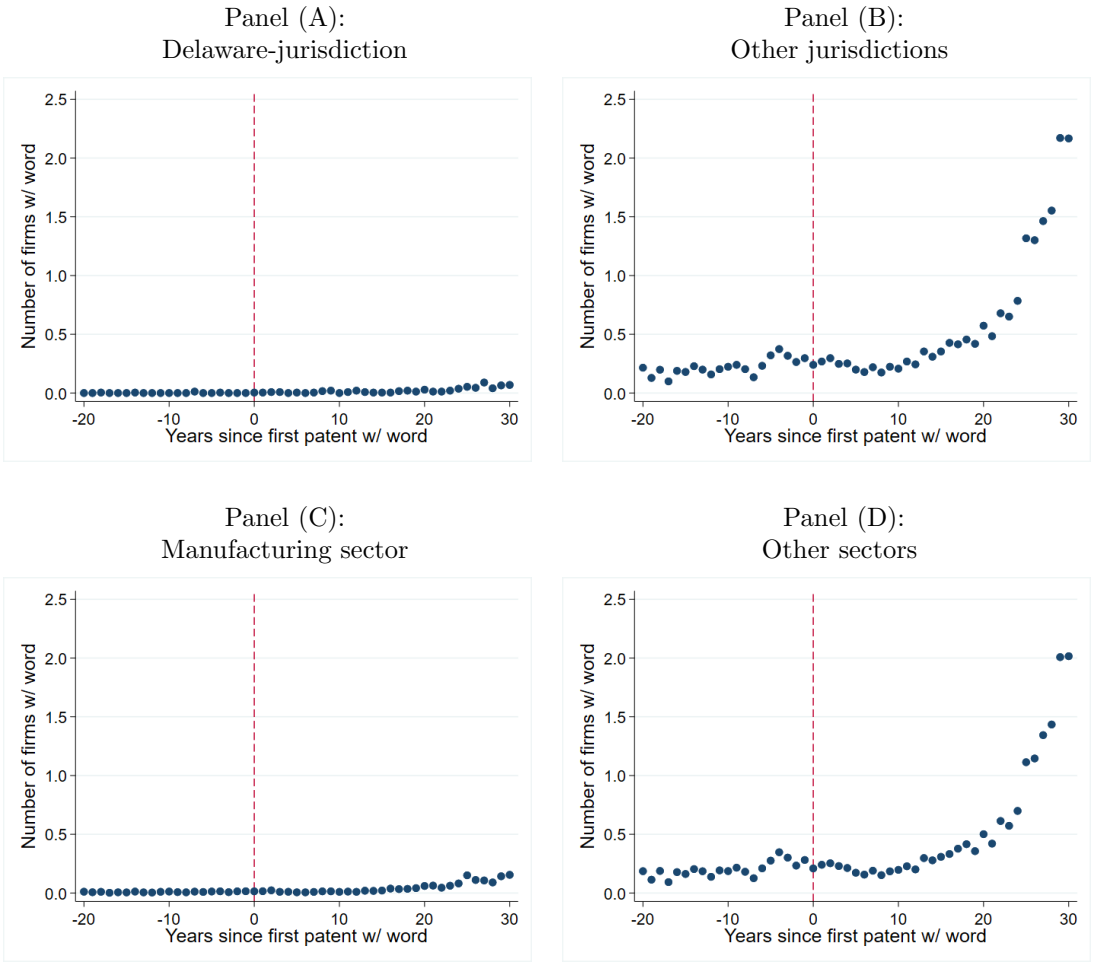
Figure C.2: Technology-related firms' share of registrations, by firm type



Notes: Figure plots the annual share of firm registrations that are semantically related to (i.e., contain in their name) one of the 1,092 technologies we sample across the 20th century. We distinguish these firms by sector (manufacturing vs. others) and jurisdiction (Delaware vs. others). The figure shows that a large majority of technology-related entrepreneurship appears to be among local, non-manufacturing enterprises.

C.3 Diffusion of technology-based entrepreneurship: patterns by firm type

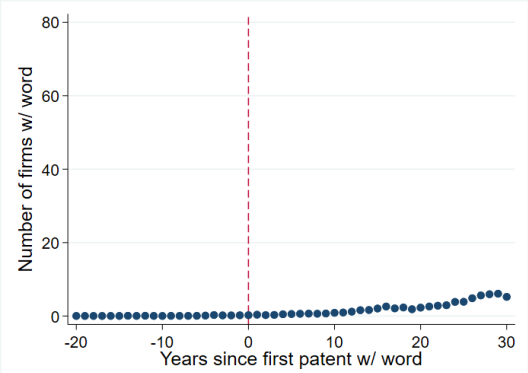
Figure C.3: Diffusion of new technologies into new firm names
words from pre-1940 patents, by firm type



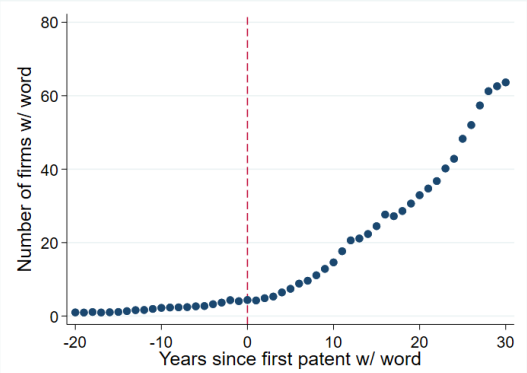
Notes: Figure plots the mean diffusion of major technologies into the entrepreneurial economy, as measured in firm names, before the year of the technology's first patent, which approximates a date of invention. The sample for these figures comprises keywords from patents granted in the 20th century which were new to the patent record when their first patent granted, and which were in the top 100 new keywords of that decade in their number of subsequent patents. Sample restricted to technologies invented before World War II. Panels (A) and (B) divide the sample into Delaware-jurisdiction and other firms, and Panels (C) and (D) into manufacturing firms and other firms.

Figure C.4: Diffusion of new technologies into new firm names
words from post-1940 patents, by firm type

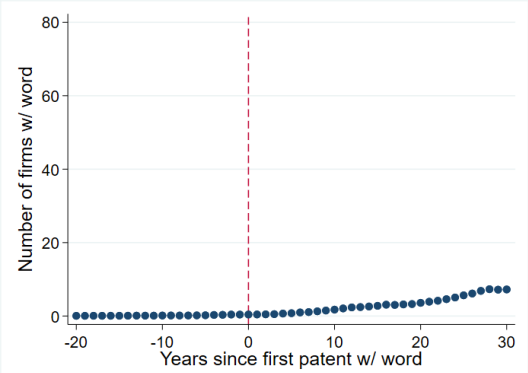
Panel (A):
Delaware-jurisdiction



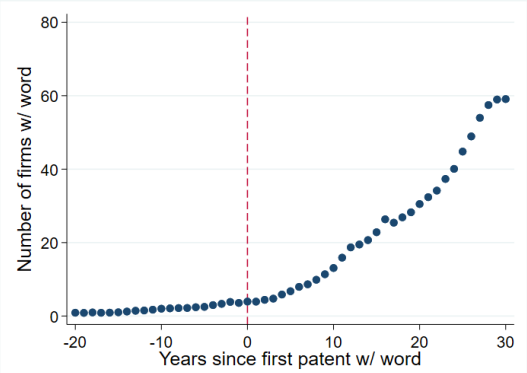
Panel (B):
Other jurisdictions



Panel (C):
Manufacturing sector



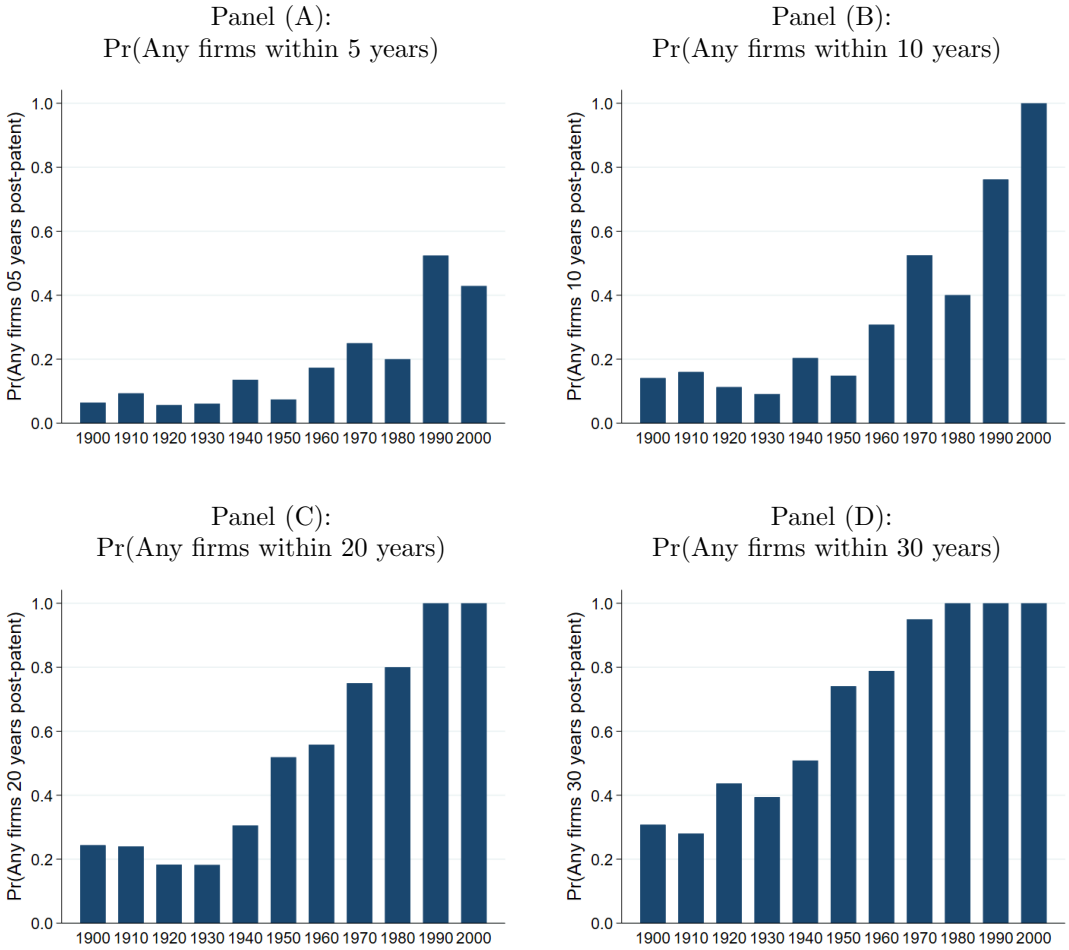
Panel (D):
Other sectors



Notes: Figure plots the mean diffusion of major technologies into the entrepreneurial economy, as measured in firm names, before the year of the technology's first patent, which approximates a date of invention. The sample for these figures comprises keywords from patents granted in the 20th century which were new to the patent record when their first patent granted, and which were in the top 100 new keywords of that decade in their number of subsequent patents. Sample restricted to technologies invented after World War II. Panels (A) and (B) divide the sample into Delaware-jurisdiction and other firms, and Panels (C) and (D) into manufacturing firms and other firms.

C.4 Changes in technology-based entrepreneurship over time

Figure C.5: Changes over time in rate of business creation around new technologies



Notes: Figure plots the probability that a technology first patented in the decade shown along the horizontal axis has any firm with that word in its name after 5 years (Panel A), 10 years (Panel B), 20 years (Panel C), and 30 years (Panel D). The sample for these figures comprises keywords from patents granted in the 20th century which were new to the patent record when their first patent granted, and which were in the top 100 new keywords of that decade in their number of subsequent patents.

D Reflections from the Growth Model

Two reflections on the model in Section 6 are below:

- This model raises the question of how to optimally trade off present output vs. future growth in how we allocate labor across entrepreneurship and production. Because L is fixed, these are intrinsically in tension with each other. In equilibrium, what fraction of labor is in each of these activities? To answer this question, we would need to add consumption to the model (the demand side). If we did so, we could obtain a balanced growth path, equilibrium interest rate, and equilibrium division of labor between sectors. Conceptually, the discount rate should govern the tradeoff between present and future consumption, and thus how under an optimal economic policy, labor would be allocated to present output vs. future growth. A high discount rate suggests allocating more labor to production.
- Observe that when $n = \frac{R}{Y}$ increases (i.e. when R&D intensity increases) by one percent, so will the growth rate, irrespective of the size of the entrepreneurial sector. This feature of the model could be revisited: for example, perhaps the optimal level of entrepreneurship (i.e. the optimal share of the labor force that is in the entrepreneurial sector) should vary with the R&D rate. Incorporating this feature into the model will require increasing the degree of complementarity between innovation and entrepreneurship.

Appendix references

Autor, David H, David Dorn, and Gordon H Hanson. 2013. "The China syndrome: Local labor market effects of import competition in the United States," *American Economic Review*, Vol. 103, No. 6, pp. 2121–68.

Berkes, Enrico. 2018. *Comprehensive Universe of US Patents (CUSP): data and facts*. Working paper.

Eckert, Fabian, Teresa C Fort, Peter K Schott, and Natalie J Yang. 2021. *Imputing missing values in the US Census Bureau's county business patterns*. NBER Working Paper No. 26632.

Ganglmair, Bernhard, W. Keith Robinson, and Michael Seeligson. 2021. *The rise of process claims: Evidence from a century of US patents*. Working paper.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg. 2001. *NBER patent citations data file: Lessons, insights and methodological tools*. NBER Working Paper No. 8498.

Jordà, Òscar, Moritz Schularick, and Alan M Taylor. 2017. "Macrofinancial history and the new business cycle facts," *NBER Macroeconomics Annual*, Vol. 31, No. 1, pp. 213–263.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. 2021. "Measuring technological innovation over the long run," *American Economic Review: Insights*, Vol. 3, No. 3, pp. 303–20.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. "Technological innovation, resource allocation, and growth," *The Quarterly Journal of Economics*, Vol. 132, No. 2, pp. 665–712.

Marco, Alan C, Michael Carley, Steven Jackson, and Amanda Myers. 2015. *The USPTO historical patent data files: Two centuries of innovation*. SSRN working paper, available at https://ssrn.com/abstract=2616724.