Human-capital shocks and innovation: Evidence from Britain's Lost Generation

Luca Repetto ®[1]

Uppsala University

Davide Cipullo ®[2]

Università Cattolica del Sacro Cuore

Edward Pinchbeck ®[3]

University of Birmingham

Jan Bietenbeck ®[4]

Lund University

Abstract

This paper studies how World War I mortality shocks to British communities affected long-run innovation. Linking parish-level military deaths to universal patent data (1895–1979) and inventor records, we compare high- and low-mortality areas. A 10 percent increase in deaths reduces the probability that a parish produces any patent by 0.09–0.12 percentage points and the probability to produce a breakthrough patent by three times as much. Mortality depresses both the entry of new inventors and the productivity of established ones, particularly in frontier and technologically complex fields. Mobility, collaboration, and stronger local innovation ecosystems mitigate these effects, albeit only partially.

Keywords: World War; Innovation; Human Capital; Patents; Lost Generation

JEL classification: D74, O15, O31

First version: Sept. 2024. This version: February 2026

---

[1]Department of Economics, Uppsala University. Email: luca.repetto@nek.uu.se.

[2]Department of Economics, Università Cattolica del Sacro Cuore. Email: davide.cipullo@unicatt.it

[3]Department of Economics, University of Birmingham. Email: e.w.pinchbeck@bham.ac.uk.

[4]Department of Economics, Lund University. Email: jan.bietenbeck@nek.lu.se

Authors' order was randomised. We would like to thank Jingyuan Zeng, Gustaf Holmberg and Rebecca Wiechers for excellent research assistance, as well as Felipe Carozzi, Adrian Gortzak, Elias Papaioannou, Vincenzo Scrutinio, Marco Tabellini and participants in seminars held at ASREC Europe Conference 2023 and 2025, University of Birmingham, Uppsala University, Università Cattolica in Milan, UIL/HEUU/Urban lab workshop in Sigtuna, ASWEDE 2024, SIEP 2025, Università di Bologna, CEPR Applied Micro History Workshop (St. Gallen) for comments and suggestions. This work was supported by BA/Leverhulme Small Research Grant SRG2021-210936, Jan Wallanders och Tom Hedelius stiftelse samt Tore Browaldhs stiftelse grant P22-0060, and the Centre for Economic Performance. This work is based on data provided through www.VisionofBritain.org.uk and uses statistical material which is copyright of the Great Britain Historical GIS Project, Humphrey Southall and the University of Portsmouth.

1. Introduction

Technological innovation is central to long-run economic growth (e.g. Romer, 1990; Aghion and Howitt, 1992; Glaeser et al., 1992), and human capital is one of its key determinants. Theories of directed technical change predict that labour scarcity increases incentives for firms to develop labour-saving technologies (Hicks, 1932; Acemoglu, 2002), and evidence from historical population shocks is broadly consistent with this demand-side channel (Habakkuk, 1962; Hanlon, 2015; Bergeaud, Chaniot and Malgouyres, 2025). At the same time, innovation depends on knowledge transmission and the supply of scarce skills – inventors, engineers, and other complementary workers (Bell et al., 2019*b*; Akcigit et al., 2018), which may be adversely affected by such shocks. While existing research has convincingly documented evidence of demand-side responses to labour scarcity, we know much less about whether human-capital losses inhibit innovation by disrupting the local supply of ideas.

In this paper, we study how a large shock to local human capital resulting from mortality in World War I (WWI) affected the supply of innovation in Britain in the long run. Mortality varied widely across space and fell disproportionately on young and relatively educated men – the so-called "lost generation" (Winter, 1977) – creating a large and plausibly exogenous shock to communities' human-capital resources. Because fighting took place abroad, Britain saw no wartime destruction of domestic physical capital, allowing us to isolate the consequences of these losses. We use this setting to test whether parish-level mortality hindered innovation by inhibiting the productivity of incumbent inventors and deterring the entry of successful new innovators.

Our main results show that WWI mortality had large negative effects on the innovative output of both pre-existing and first-time inventors that lasted for several decades. These effects are stronger – roughly three times larger – for higher-quality patents. Additional results consistently point to a human-capital supply-side channel. Effects on established inventors are concentrated in knowledge-intensive sectors and are more pronounced for inventors whose work is more innovative and complex, but are attenuated by relocation and co-authorship. The adverse effects of war deaths on the emergence of new first-time inventors are not driven by scale, and are amplified when losses include highly skilled individuals but mitigated by proximity to universities and role models.

To document these effects, we build a new dataset that links WWI mortality records to patent data for England and Wales, 1895–1979. Mortality and military participation come from linked Commonwealth War Graves Commission death records and military service data. Innovation is measured by combining PatentCity (geocoded patent filings aligned to 1911 parish boundaries) with PATSTAT inventor identifiers and patent abstracts. The linked data cover the universe of geo-referenced WWI deaths and patents filed in England and Wales over the period, allowing us to analyse responses at both the parish and inventor levels.

The first part of our analysis examines how WWI fatalities affected parish-level patenting. Patenting is rare at this level – only one in ten parishes register a patent in a given year – so we aggregate outcomes to 5-year periods and use as baseline outcome an indicator equal to one if a parish registers at least one patent. Building on Kelly et al. (2021), we also construct text-based measures of "breakthrough" patents.[1] We estimate difference-in-differences and event-study specifications relating parish outcomes to local WWI deaths, with parish, and period-by-county fixed effects that control for local labour market conditions and plausibly absorb induced aggregate demand effects. The estimates show an immediate and persistent decline in innovation: a 10 percent increase in WWI deaths lowers the probability that a parish produces any patent in a 5-year period by about 0.09–0.12 percentage points and the probability that a parish produces any breakthrough patents by roughly three times as much. Taken together, these results indicate that WWI altered both the geography and the quality distribution of British invention for decades.[2]

Having documented the negative effect of WWI mortality on innovation, we turn to the underlying mechanisms. At the most general level, the decline in innovation can arise through two channels: lower output of pre-existing inventors, who may perish during the war, relocate, or become less productive; and the emergence of fewer new inventors ("Lost Einsteins," Bell et al. 2019*a*). Our results show that both channels are at play. WWI mortality reduces patenting of both pre-existing inventors, i.e., patentees who were already active before the War, and the long-run entry of successful new inventors.

To study the first channel, we construct an inventor-level panel of roughly 54,000 patentees who registered at least one patent in the pre-war period and follow them through 1979. Inventors in higher-mortality parishes patent less after the war, and the effect is larger for more innovative patents, consistent with the parish-level results.[3]

The inventor-level analysis uncovers several dimensions of heterogeneity. The effect is not uniform across technological fields, but is larger amongst inventors patenting in advanced sectors such as electricity, machinery, transportation, physics, and chemistry, and modest and not statistically significant in human necessities and construction. In addition, inventors with engineering occupations and those whose first patents are more complex or more important (based on the importance metric of Kelly et al. 2021) are more heavily affected.[4] Overall, the declines are concentrated among inventors operating in more skill-intensive and frontier domains (Smil, 2005), consistent with innovation in these sectors relying on specialised teams, complementary technicians and engineers, and supporting infrastructure (Neffke, 2019; Jaravel, Petkova and Bell, 2018; Hanlon, 2022).

[1] Focusing on breakthrough patents helps address the well-known issue that patents are an imperfect measure of innovation, as not all innovations are patented and not all patents represent true innovation (e.g. Griliches, 1990; Nagaoka, Motohashi and Goto, 2010).

[2] Our continuous-treatment DiD relies on strong parallel trends (Callaway, Goodman-Bacon and Sant'Anna, 2024), an assumption that would be automatically satisfied under fully exogenous treatment assignment. Because WWI mortality is unlikely to be exogenous – for example, declining communities may have sent less prepared soldiers to the front – in a complementary analysis we instrument parish deaths using a shift-share IV that follows Carozzi, Pinchbeck and Repetto (2023) and exploits plausibly exogenous battalion assignments and battalion-level death rates. The IV estimates closely match our main findings.

[3] The effect is similar when we exclude WWI fatalities and when we restrict to inventors that produced at least one patent over the post-WWI period, indicating that our estimates are not driven by the mechanical loss of inventors.

[4] To ensure comparability across specifications, we assign to each inventor a field, an occupation, and measures of patent complexity based on the first patent that they publish in the pre-war period.

Subsequently, we examine whether inventors can mitigate the adverse productivity effects of mortality. We consider three margins: switching field, changing co-authorship strategy, and relocating. Relocation to another parish attenuates the shock and may reverse the inventor's fortune, whereas the negative effect is larger than the baseline estimate for inventors who remain in their initial location. This pattern aligns with prior evidence on exposure effects (Bell et al., 2019*b*), suggesting that moving to highly innovative areas raises inventorship through improved human-capital transmission, better infrastructure, stronger networks, or a combination of these factors. By contrast, switching sectors does not mitigate the impact, possibly because retraining is costly. Co-authorship also provides a buffer: while the likelihood of producing solo-authored patents drops substantially for inventors in high-mortality areas, the probability of patenting with at least one co-author is unaffected, consistent with access to co-inventors and broader collaboration networks mitigating productivity losses after shocks (Jaravel, Petkova and Bell, 2018).

To investigate the second channel, we test the "lost generation" hypothesis: the loss of highly educated and technically skilled individuals reduced communities' long-run innovative capacity. Conditional on WWI mortality, losing an elite school alumnus is associated with a significant additional decline in subsequent innovation, though precision varies across specifications. Using detailed soldier records, we also show that the death of an officer with an engineering background has an additional negative effect, whereas the loss of other officer types does not.[5] These results suggest that the loss of young, talented, and technically skilled servicemen had persistent local effects: communities lost not only potential future inventors and role models, but also part of the human-capital base that supports the production of new ideas.

We next examine heterogeneity by local innovation ecosystems, which prior work has shown to matter (see, e.g., Akcigit, Grigsby and Nicholas 2017*b*; Bell et al. 2019*b*; Andrews 2023). The effect is attenuated in areas with members of engineering institutes and proximity to universities, suggesting that inventors that are supported by an ecosystem favourable to innovation can partly offset the negative consequences of a human-capital shock.

Our paper contributes to the literature on the relationship between labour scarcity and technological change. Previous theoretical and empirical work has examined when economies respond to worker shortages by adopting or developing labour-saving technologies (Acemoglu, 2007; Voth, Caprettini and Trew, 2023). Our findings instead show that the loss of human capital – especially at the upper tail – can produce a persistent decline in the local supply of innovative capacity. While this result may seem at odds with earlier studies, one reconciliation is that much of the prior evidence concerns low-skilled labour (e.g. Hanlon 2015; Andersson, Karadja and Prawitz 2022; Bergeaud, Chaniot and Malgouyres 2025), whose tasks were more readily substituted by machines. Moreover, Andersson, Karadja and Prawitz 2022's study of mass emigration in Sweden 1867–1914 and Bergeaud, Chaniot and Mal-

[5]This latter finding aligns with recent evidence that engineers are pivotal to the production of innovative output (Hanlon, 2022) and underscores the role of technically oriented individuals in driving technology adoption and economic growth (Maloney and Valencia Caicedo, 2022).

gouyres 2025's work on WWI fatalities in France both examine labour scarcity during the transition from agriculture, whereas our evidence comes from an industrialised economy.

A second difference concerns the level of analysis. We observe both wartime mortality and invention at a very fine geographic scale, which allows us to condition on market-level shocks and study local responses in innovative activity. This level of detail is important if demand-side adjustments toward labour-saving technologies are more likely to respond to broader market-level scarcity, whereas supply-side responses by inventors and their collaborators react to very localised losses of human capital. If inventors and complementary workers react to local economic conditions, human-capital shocks can reduce innovative capacity in ways that analyses at a broader geographic aggregation may attenuate or miss. In this sense, our evidence complements prior work on demand-side effects by highlighting a localised supply-side mechanism.

Our paper also contributes to the large body of research on the determinants of innovation, and its distribution in space and over time. Recent research has highlighted the importance of individual background as well as external factors, such as exposure to an innovative environment (Bell et al., 2019*b*; Akcigit, Grigsby and Nicholas, 2017*a*; Andrews, 2023). The spatial distribution of innovation is remarkably persistent over time (Andrews and Whalley, 2022) and often concentrated in highly productive clusters (e.g., Moretti 2021; Audretsch and Feldman 1996). However, history offers striking examples of the rapid decline of innovation hubs that once played pivotal roles such as Florence during the Renaissance, Germany's Ruhr Valley, and Detroit's automobile industry. Our results provide new evidence that exposure to major external shocks such as wars may persistently harm the long-run innovative potential of communities, both in terms of quantity and quality.[6] We also bridge the gap between studies analysing community-level outcomes and those focusing on individual inventors by providing evidence on both in the same setting and over the long run.

We also contribute to the literature on the long-run economic consequences of war.[7] Prior work finds heterogeneous effects across settings and conflict types: bombings can have persistent negative impacts (Lin, 2022; Riaño and Valencia Caicedo, 2024), yet population and activity often recover (Davis and Weinstein, 2002; Brakman, Garretsen and Schramm, 2004). Chupilkin and Koczan (2022) show that long-run GDP effects vary with conflict characteristics, with on-territory wars yielding larger persistent losses than off-territory wars. Our setting isolates military human-capital losses without physical capital destruction and at a granular spatial scale, allowing us to show that these losses depress innovation at the community level for decades. Besides this finding, we also advance this literature by tracing the

[6]A notable exception to the general lack of long-term focus is the work on the impact of the Nazi regime on science. Waldinger (2016) use WW2 destruction as a shock to physical and human capital and show the dismissal of scientists in Nazi Germany had long-lasting negative effects on academic productivity and the capacity of departments to attract new talent.

[7]Many papers also consider how war-related losses affect outcomes in the short run, including on marriage patterns, female labour force participation and gender norms (Abramitzky, Delavande and Vasconcelos, 2011; Boehnke and Gay, 2020; Brainerd, 2017), while Carozzi, Pinchbeck and Repetto (2023) find that WWI mortality led to the creation of civic capital in Britain. A broader body of related papers also examine other mortality shocks, for example due to infectious diseases, see e.g. Franck (2024) for an example that also looks at innovation.

evolution of outcomes before, during, and after the war. This allows us to validate our results by inspecting pre-trends and the dynamics of the treatment effect, mitigating concerns about time-varying shocks and the compression of history between treatment and a later outcome in persistence studies (Voth, 2021).

2. Background

2.1. The evolving patent landscape in Britain

We begin with a brief overview of key institutional details. The British Patent Office opened in 1852 and by the start of the period we study (1895–1979), the British patent system was well established. In 1883 an Act significantly reduced patent fees, and around the same time, Britain began to provide full patent rights to signatory countries under an international convention. Patents became subject to an investigation into the novelty of an invention prior to grant from 1902. The legislative framework was modernised by a 1977 Act, which underpins the current system.

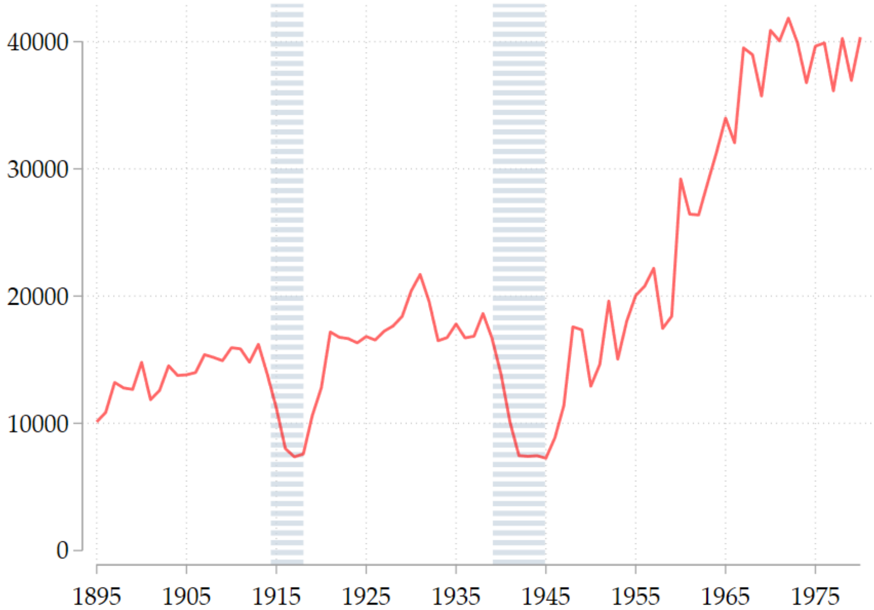
By the turn of the 20^{th} century, the global distribution of international patents was heavily concentrated in a small group of industrialised nations. Although no longer the leading or fastest-growing industrial power, Britain retained substantial inventive capacity in areas such as mechanical engineering, shipbuilding, textiles, and emerging electrical industries. British inventors were active domestically but also in filing patents in major markets such as the United States and parts of continental Europe. Although Britain's share of global patenting declined, the number of patents registered in Britain increased substantially over the time period we consider in our analysis, albeit activity was reduced during both World Wars (Figure 1). There was also a marked shift towards corporate invention: until the 1920s, the vast majority of patents were registered to individual inventors, but by 1950 companies were registering as many patents as individuals (Nicholas, 2011).

There is no single dataset that provides comprehensive information about the age, occupation, education and income of inventors in England and Wales in the first half of the 20^{th} century – few individuals who generated patents described their occupation as inventors in the Census, and patent databases do not contain comprehensive inventor characteristics. Data provided by Nicholas (2011), however, suggests that between 35 and 45% of independent inventors registering patents between 1910 and 1930 were engineers. Unfortunately no British data sources contain educational information on inventors. However, in a different context Akcigit, Grigsby and Nicholas (2017*b*) show that around 40% of US inventors in 1940 had a university degree (relative to 10% of non-inventors).

2.2. The British Armed Forces during WWI

Over 4.5 million men from England and Wales served in the British Army during WWI, and around 800,000 more served in the Royal Navy and what became the Royal Air Force (Winter, 1977). The Army expanded rapidly after heavy early casualties, first through volunteering and then through conscription from 1916. Of those who served, around 330,000

FIGURE 1
PATENTS REGISTERED IN ENGLAND AND WALES



Notes: Total number of patents registered at the UK Patent Office, all patentees, yearly totals for all periods from 1895 to 1979. Source: authors' elaboration from PatentCity.

were commissioned officers (Winter, 1977). At the beginning of the war, the 20,000 or so officers of the serving and reserve armed services were primarily drawn from Britain's upper middle and upper classes, recruited through prestigious public schools and universities. As the demand for officers grew, however, men from outside these traditional backgrounds were promoted from the ranks. Although exemptions existed for some "reserved occupations," scientific and technical workers were not systematically shielded from frontline service, especially in the early years of the war. As a result, the war generated large losses among young cohorts and plausibly affected the local stock of skills relevant for innovation.

2.3. Mortality in WWI and the Lost Generation

Because the war was fought overseas, war-related deaths in Britain were overwhelmingly military: over 750,000 armed-forces deaths versus fewer than 17,000 civilian deaths from enemy action. The mortality shock was also strongly age-specific, with very high losses among young men and especially among cohorts under 30 (Winter, 1977). This cohort concentration underpins the notion of a British "Lost Generation".

Mortality was not evenly distributed across social groups. Officers – who faced high frontline exposure – had substantially higher death rates than enlisted soldiers, especially in 1914–15 (about 14% versus 6%) (Winter, 1977). Since officer recruitment in the early war years remained closely tied to elite educational institutions, university-educated men were disproportionately affected; among Oxford and Cambridge graduates, wartime death rates were around 19% and 18%, with rates above 25% for students who matriculated between 1910 and 1914.

Evidence for non-elite groups is less clear. Bailey, Hatton and Inwood (2023) show that among non-officer servicemen, wartime mortality was strongly related to branch and timing of enlistment, with weaker evidence of disadvantage for men from white-collar backgrounds. At the same time, demobilisation statistics indicate that engineering workers represented a substantial share of non-officer servicemen (BWO, 1922). This is relevant because invention in our period is heavily concentrated among engineers. Taken together, these patterns motivate our focus on how localised wartime losses of high-skill and technically trained men affected subsequent innovation.

3. Data and descriptives

We construct a parish-level panel dataset for England and Wales covering 5-year intervals from 1895 to 1979, and an auxiliary panel at the inventor level for the same period.[8]

Parishes were the primary administrative unit of local government in England and Wales during the early 20th century and closely approximate local communities. For consistency, we use parish boundaries as defined in 1911, the last Census year before WWI, throughout our analysis to ensure uniformity in geographic units across time. Below, we describe the main data sources used and the procedures for constructing our sample.

3.1. Data on WWI deaths

Our main data on WWI soldier deaths come from the casualty database of the Commonwealth War Graves Commission (CWGC) (Commonwealth War Graves Commission, 2023), complemented by Forces War Records (FWR) (Forces War Records, 2023). These sources provide soldier-level information on name, age, date of death, rank, regiment, service unit, and place of birth or residence. We geolocate soldiers to parishes of origin using reported locations; details on construction and validation are in Appendix B. See Carozzi, Pinchbeck and Repetto (2023) for additional details on the geolocation process.

To construct measures of loss of elite individuals, we match CWGC-FWR records to the Oxford University Roll of Service (Craig, 1920) and the Cambridge War List (Carey, 1921), and we use rank information to identify officers. In addition, we exploit descriptive text fields in the CWGC records to identify further alumni of Oxford, Cambridge, and elite schools in the UK.[9] For mechanism analysis, we further split officers into two groups: engineers (identified through Royal Engineers service and engineering-related rank or commemoration strings) and other officers.

We then aggregate soldier-level records to the parish level to construct our main treatment variable (total WWI deaths) and indicators for whether a parish lost at least one elite school graduate, at least one officer, and at least one officer with engineering background.

[8]We exclude Scotland from our analysis as Census data and digital maps for Scotland are unavailable for this period.

[9]Elite schools are defined as the nine leading schools identified by the Clarendon Commission in 1861: Eton, Charterhouse, Harrow, Rugby, Shrewsbury, Westminster, Winchester, St Paul's, and Merchant Taylors (Shrosbree, 1988).

3.2. Data on patents

Our primary source of patent data is PatentCity (Bergeaud and Verluise, 2024), which we restrict to patents registered at the Great Britain patent office between 1895 and 1979.[10] We treat each patent-inventor pair as one observation, yielding about 1.7 million patents and 2.8 million inventor-patent pairs before sample restrictions.

A key feature of PatentCity is geocoded inventor locations. We map inventor-patent records to 1911 parish boundaries and exclude foreign residents and records without sufficiently precise geolocation.[11] The resulting sample contains about 855,000 inventor-patent pairs (around 600,000 patents).

We merge PatentCity with PATSTAT (European Patent Office, 2023) using the common patent identifier. This provides inventor identifiers, abstracts, and co-authorship information, allowing us to track inventors over time.[12]

Using PATSTAT abstracts, we construct a text-based patent-importance measure following Kelly et al. (2021) (Appendix C).[13] We also use Cooperative Patent Classification (CPC) codes and, when needed, predict likely fields from patent titles using NLP methods (Appendix D).

3.3. Further data sources

We use three additional sources. First, we use I-CeM Census micro-data for England and Wales in 1911 from IPUMS International (Minnesota Population Center, 2019) to measure parish population and pre-war covariates, including the share of households with no servants, the share of male white-collar workers, and a proxy for unemployment.

Second, we use British Army Service Records (1914–1918) from FamilySearch (FamilySearch, 2023) to construct parish-level measures of mobilisation.[14] We combine these records with CWGC deaths to construct the shift-share IV sample, applying the restrictions described in Appendix B.

Third, we construct 5-year parish populations using Census data from "A Vision of Britain through Time" (VoB), spatially re-weighting to 1911 parish polygons and linearly interpolating between Census years.

We use additional sources to construct proxies for regional innovation ecosystems before the start of the war, such as membership lists of four engineering trade bodies in pre WWI years; a list of pre-WWI universities; a list of rail stations in 1910; and, finally, a georeferenced list of all post offices in 1900 from the GB1900 Gazetteer. We then use these sources

[10]PatentCity includes some records before 1895 and after 1980. We exclude these years because of incomplete early coverage and a structural break in later years. We note that the documentation states that the dataset creators' goal was to mimic as closely as possible the concept of first publication of a granted patent, but then later states that the UK data is based on Patent Application documents. We will typically refer to the patents in the data as being registered patents.

[11]See Appendix B.5 for geolocation and assignment details.

[12]The match is near-complete. See Appendix B.5 and Figure A.1 for details on unmatched records.

[13]Missing abstracts are concentrated in the first years of the sample and in 1979.

[14]These records are incomplete because part of the collection was destroyed in a 1940 fire. FamilySearch combines surviving service records ("Burnt documents") and pension-related records ("Unburnt documents"), for about four million records in total. Appendix B describes parish assignment.

to construct a series of separate county-level ecosystem accessibility measures, computing the average distance to the nearest ecosystem factor in each county by taking the weighted average of the parish level distances, using as weights 1911 parish populations. We then take the natural log of this weighted average distance.

3.4. Sample selection and descriptive statistics

Our final parish-level dataset comprises 10,807 grouped parishes.[15] We retain patents geolocated at least to the city level and aggregate outcomes into 5-year periods from 1895–1899 to 1975–1979, yielding a balanced parish-by-period panel.

Our second dataset is an inventor-level panel spanning 1895–1979. We restrict to 54,642 patentees (excluding firms and other non-natural entities) who registered at least one patent in the pre-war period (1895–1914) and track their patenting over time using PATSTAT inventor identifiers.[16] For each inventor, we construct time-invariant characteristics (residence, WWI mortality exposure, field, and patent importance) based on the first patent observed in the sample.

Table A.1 in the Appendix presents summary statistics for key parish characteristics. In 1911, the average parish had a population of about 3,300 inhabitants and covered 14 square kilometres. On average, each parish contributed around 285 servicemen to the WWI mobilisation effort, implying a parish-level mobilisation rate of about 5.5 percent of the population.

Figure 2 provides a visual overview of the variation in WWI losses and patenting activity across parishes in our final dataset. Panel A shows the spatial distribution of soldier deaths per parish, revealing substantial geographic variation. Panel B presents a histogram of soldier deaths, highlighting that roughly one in nine parishes did not experience any soldier fatalities, and that the distribution is highly skewed. On average, parishes reported 54 soldier deaths, equivalent to about 1 percent of the 1911 population, as shown in Table A.1. Additionally, 6 percent of parishes had at least one soldier who attended an elite school and died in the war, while 11 percent experienced the death of a soldier with an engineering background.

In terms of patenting activity, Panel C of Figure 2 maps the average number of patents registered per 5-year period across parishes over the full sample. Consistent with prior findings, patenting is highly spatially concentrated in England and Wales. Panel D presents a histogram of the number of patents per 5-year period, showing a highly skewed distribution: even when aggregated over 5 years, around three-quarters of parish-period observations have zero patents. Among parishes with patenting activity, the mode is a single patent, although the distribution features a long right tail. Due to this skewness, our main analysis focuses on a binary outcome variable indicating whether at least one patent was registered in a given 5-year period. Table A.1 in the Online Appendix shows that the mean of this indicator in our sample is 0.24.

[15] For details on parish grouping, see Appendix B.

[16] Appendix B.5 provides details on the sample construction.

Patenting activity is also spatially clustered within counties.[17] We compute a Herfindahl-Hirschman index (HHI) of spatial concentration of patenting at the county level and use it to infer how many parishes effectively contribute to each county's innovation. On average, innovation clusters span only about 5 percent of a county's parishes, supporting the use of a geography finer than the county. In Figure A.3 in the Appendix, we compute the spatial HHI by county for different levels of patent importance using the measure of Kelly et al. (2021), then take the national average. Concentration is present for all patents but rises with importance: more consequential innovations are more geographically concentrated than less consequential ones.

4. Empirical strategy

4.1. Empirical specifications

Our baseline model is a continuous-treatment difference-in-differences (DiD) specification that compares changes in innovation across parishes exposed to different WWI mortality intensity:

$$y_{it} = \alpha + \beta log(d_i^{WWI}) \times Post\ WWI + \gamma' W_{it} + FE + \epsilon_{it}, \tag{1}$$

where y_{it} is an innovation outcome (e.g. an indicator for at least one patent) for parish i in 5-year period t.[18] $log(d_i^{WWI})$ is the log number of servicemen from parish i killed in WWI, and $Post\ WWI$ indicates periods from 1915 onward.[19] We also include controls W_{it} and different sets of fixed effects that vary by specification.[20]

To study dynamic treatment effects and to evaluate pre-trends, we also estimate the following dynamic DiD model:

$$y_{it} = \alpha + \sum_{k=1895, k \neq 1910}^{1975} \beta_k log(d_i^{WWI}) \times D_k + \gamma' W_{it} + FE + \epsilon_{it}, \tag{2}$$

where D_k is an indicator for 5-year period k. As it is customary, in estimation we exclude the pre-treatment period 1910-1914 which, therefore, becomes the baseline.

As well as the parish level model, we also use an inventor-level panel for 1895-1979. This allows us to study effects on pre-war inventors, heterogeneity by sector and patent complexity, and the mitigating role of mobility and co-authorship. Using this panel, we

[17]In Figure A.2 in the Appendix, we randomly sample ten counties and reproduce Panel C of Figure 2 to illustrate within-county variation.

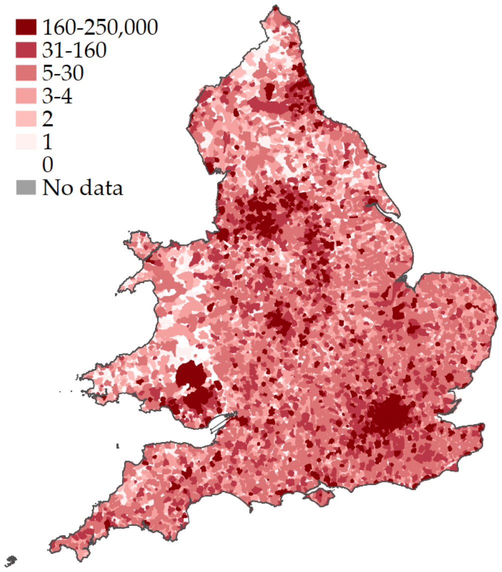
[18]The first period is 1895-1899 and the last is 1975-1979.

[19]Because we take logs, the baseline model excludes parishes with zero deaths. Results are robust to using $log(1 + d_i^{WWI})$ (Table A.6).

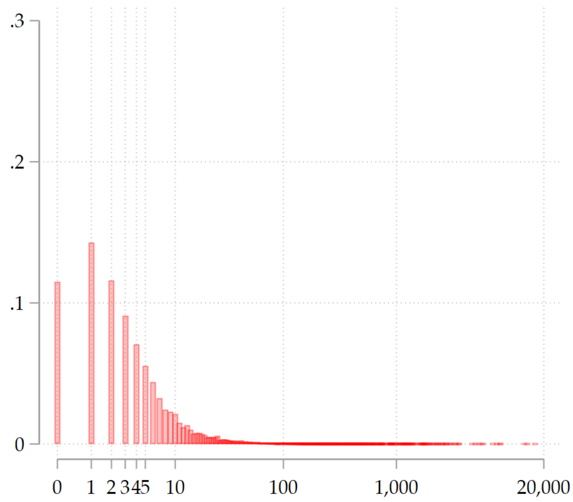
[20]These controls are the log of WWI mobilisation at the parish level, as well as variables from the 1911 Census, always interacted with linear time trends. Specifically, we include total population, the share of households in the parish with no servants (as a proxy for income), the share of white collar workers and a proxy for unemployment (to control for labour market conditions prevailing before the War). The set of controls is purposefully parsimonious because i) the difference-in-difference assumptions rely on parallel trends – for which there is evidence already in the model without controls, and ii) to avoid the inclusion of an excessive number of controls and fixed effects (that have to be interacted with linear trends or time effects).

FIGURE 2

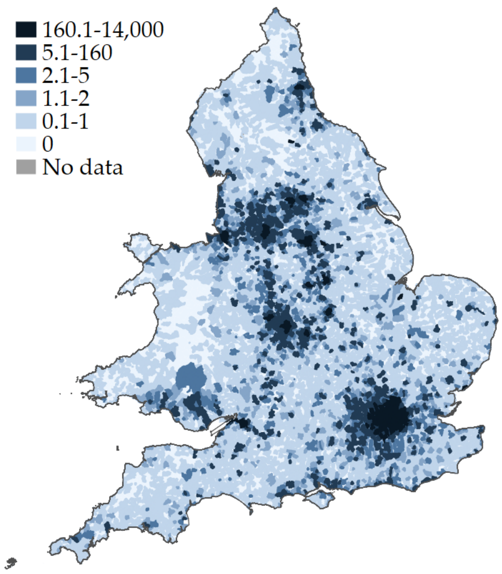
PATENTS REGISTERED AND WWI DEATHS IN ENGLAND AND WALES



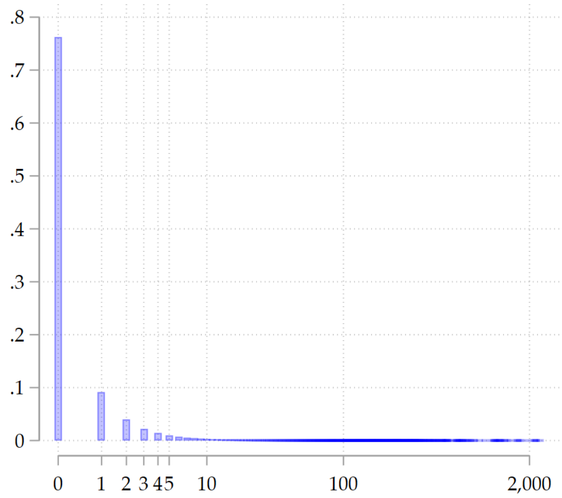
(A) N. DEAD WWI SOLDIERS



(B) N. DEAD WWI SOLDIERS – DENSITY



(C) N. OF PATENTS OVER 5-YEAR PERIOD – AVERAGE



(D) N. OF PATENTS OVER 5-YEAR PERIOD – DENSITY

Notes: Historical (grouped) parishes in England and Wales. Panel A displays the number of soldiers killed in WWI originating in each parish. Panel B reports the density of the number of soldiers killed in WWI. Panel C shows the number of patents registered over a 5-year period, averaged over all periods in 1895-1979. Panel D shows the density of number of patents registered over a 5-year period (where each period counts as an observation).

estimate the following inventor-level continuous DiD model:

$$y_{ijt} = \alpha + \beta log(d_i^{WWI}) \times Post\ WWI + \gamma' W_{it} + FE + \epsilon_{ijt}, \tag{3}$$

where y_{ijt} is an indicator for inventor j from parish i having registered a patent in period t, and the rest is as above.[21]

4.2. Identification assumptions

Because (log) WWI mortality varies continuously across parishes, identification comes from differences in treatment intensity rather than from a treated-versus-untreated comparison. The key DiD assumption is that, conditional on controls and fixed effects, parishes with different mortality intensity would have followed parallel outcome trends absent differential mortality exposure. Under this assumption, the interaction between post-war indicators and $log(d_i^{WWI})$ identifies the marginal causal effect of WWI mortality on innovation outcomes (Callaway, Goodman-Bacon and Sant'Anna, 2024).

Since all parishes experienced positive (log) mortality, our estimates do not identify the effect of moving from zero to positive exposure. Instead, they identify how outcomes vary with marginal changes in mortality intensity within the observed support.

Although the parallel trends assumption is not directly testable, we provide several validation checks in Section 8. First, we show no systematic pre-trend differences across deciles of $log(d_i^{WWI})$ and find similar conclusions when using a binary above-median treatment (Table A.2). Finally, we show that results from a complementary design where we instrument parish-level WWI deaths using a shift-share IV following Carozzi, Pinchbeck and Repetto (2023) yield similar results. Specifically, we use as instrument for actual mortality predicted mortality that combines battalion-level mortality rates with each parish's distribution of servicemen across battalions. The exclusion restriction is that, conditional on controls, battalion-level mortality reflects differences in combat exposure rather than unobserved parish-level determinants of innovation. We show that the first stage is strong and that IV estimates closely track the baseline DiD in Section 8.

5. Main results

5.1. The local effect of human-capital shocks on innovation in the long-run

Table 1 reports DiD estimates from equation 1 of the effect of parish-level WWI mortality on parish-level patenting.[22] column 1 includes parish and time fixed effects only. The estimates imply that higher parish-level fatalities reduce the probability that residents produce any patent in subsequent decades. Specifically, a 10 percent increase in deaths lowers

[21] For the inventor-level analysis, the vector FE includes inventor fixed effects, time fixed effects, and year of first patent fixed effects interacted with time fixed effects.

[22] In this table and in those that follow, we restrict the sample to that used in column 4 of Table 2, the most demanding specification, which also controls for log parish population. Doing so keeps the number of observations constant across columns and avoids variation due to missing covariate values. The restriction is imposed purely for comparability; none of the results are sensitive to relaxing it.

the probability of patenting by 0.09–0.12 percentage points, a 0.3–0.5% decline relative to the baseline mean. In column 2, we add (log) county-level mortality as a proxy for labour-market-wide mortality exposure. This allows us to compare highly localised war losses with broader regional shocks. The county-level measure has no statistically significant long-run effect on patenting in our setting.

In column 3 (our preferred specification) we introduce county-time fixed effects, thus limiting comparisons to parishes within the same county and eliminating county-wide factors. The coefficient on parish mortality becomes larger in absolute terms by around a quarter, suggesting that county-level factors that correlate with parish level mortality attenuate our estimates. Column 4 shows that this estimate is robust to including interactions of baseline controls (log 1911 population, share of households with servants, share white collar workers, share unemployed) with a linear time trend. Finally, in column 5 we control for initial parish size less parametrically by including the three way interaction between time, counties, and decile of 1911 parish population within each county.

Figure 3 examines the dynamics of these effects by presenting estimates and 95 percent confidence intervals of the event study specification in equation 2 that includes parish and county-time fixed effects (i.e. the specification corresponding to column 3 of Table 1). Inspection of the figure reveals that all estimated pre-trend coefficients are small and not statistically different from zero. In the 5-year period that includes the Great War, we observe a pronounced drop in patenting activity, with a coefficient about twice as large as the one estimated using the full post-WWI period in Table 1, suggesting that the impact of war mortality was sizeable at the onset of the war and then decreased in the subsequent decades. This is indeed the case as evidenced by inspecting the figure further. After a recovery – almost to pre-war levels – in the inter-war period, patenting activity declines again before and during WW2 in parishes with high WWI mortality, and remains low for the remainder of the sample period.

For comparison, Appendix Figure A.4 plots an event-study specification that includes both parish- and county-level mortality (corresponding to column 2 of Table 1). The pattern for parish-level deaths is effectively identical to that in Figure 3. By contrast, county-level mortality has a different effect on parish innovation: larger county shocks are linked to no change, or even a modest increase, in patenting in the short to medium run, but to lower patenting in the long run. These results underscore that the spatial scale at which population shocks are measured is crucial for understanding their effects on places.

The results using our instrumental variables approach for WWI deaths closely mirror those from the difference-in-differences specification discussed here. Therefore, to maintain clarity and focus, we present the IV results separately in Section 8, and concentrate on the difference-in-difference specification throughout the main analysis.

5.2. Effects by patent quality

The results presented thus far demonstrate that WWI mortality led to a decline in the production of local innovation. However, this analysis centers on the *quantity* of innovation and does not consider its *quality* – that is, the technological significance and impact of each

Table 1
WWI deaths and patenting – OLS estimates

	(1) Any patent	(2) Any patent	(3) Any patent	(4) Any patent	(5) Any patent
$Log(d^{WWI})\times$ Post WWI	-0.009***	-0.010***	-0.012***	-0.009***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
$Log(d^{County})\times$ Post WWI		0.001			
		(0.002)			
Mean dep.var.	0.26	0.26	0.26	0.26	0.26
R2	0.48	0.48	0.49	0.49	0.51
Observations	162004	162004	162004	162004	162004
Parish FE	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y
County $\times$ Time	N	N	Y	Y	Y
Controls $\times$ Linear trend	N	N	N	Y	N
County $\times$ Time $\times Q_n$ Pop	N	N	N	N	Y

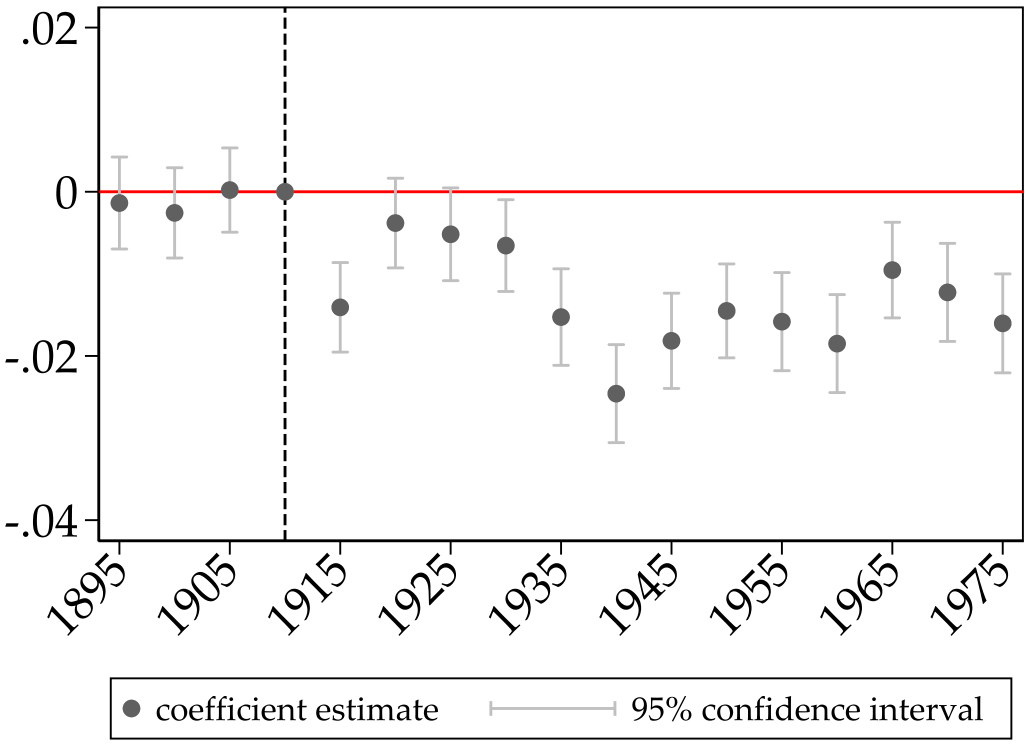
Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by a resident of parish i during 5-year period t. Column 1 includes parish and time fixed effects. Column 2 adds log county-level mortality. Column 3 adds county-time fixed effects. Column 4 adds baseline controls interacted with a linear time trend. Column 5 controls for a three way interaction between time, counties, and deciles of 1911 parish population within each county. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

patent. It is plausible that WWI mortality altered the balance between the number and the importance of innovations, potentially reducing the overall volume of patenting without necessarily affecting the production of high-impact inventions. To investigate this question, we study whether WWI human-capital losses affected not just the output but also the quality of innovations. Patent quality is measured using a text-based indicator of innovativeness, constructed according to the approach of Kelly et al. (2021) briefly outlined earlier. This allows us to assess whether WWI mortality lowered the likelihood that communities generated more influential patents, in addition to reducing patenting activity overall.

Figure 4 presents a series of estimates from repeated applications of equation 1, where the dependent variable is progressively restricted to patents above increasing importance thresholds. For each 5-year window, we rank patents by their importance and construct a binary variable indicating whether parish i had at least one patent meeting or exceeding a given threshold during period t. The initial estimate includes all patents, treating any patent as "important". Subsequent iterations successively tighten the criterion: first to only patents in the top 90% of the importance distribution, then the top 80%, and so forth, culminating with an indicator that equals one only if a parish resident published at least one patent in the top 10% of importance within that window.

The results suggest that the negative impact of the WWI death shock on innovation potential grows with patent importance. In absolute terms, a higher exposure to the WWI mortality shock reduces the probability of publishing at least one patent in the top half of the importance distribution by approximately one-third more than the effect on the probabil-

Figure 3
WWI deaths and patenting – Event-study results



Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by a resident of parish i during 5-year period t. Estimates from equation 2, including time, parish, and county-time fixed effects. 95% confidence intervals constructed using standard errors clustered at the parish level.

ity of publishing a patent of any importance. This relationship is even more pronounced in the right tail of the importance distribution: the effect of the WWI shock on the probability of producing a patent in the top 10 percent of importance is roughly three times as large as the effect on the probability of registering a patent of any importance, suggesting that WWI mortality affected both the quantity and the quality of the innovation produced by British communities.[23]

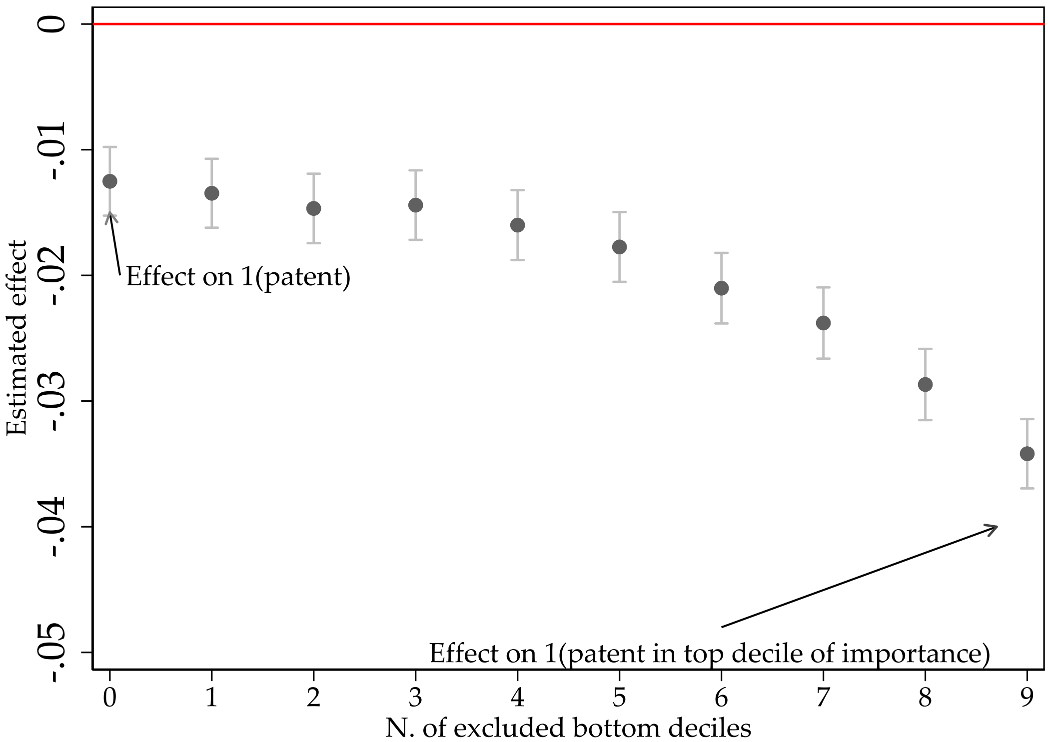
5.3. *Effects on patents by established and new inventors*

Highly localised war mortality could affect communities' innovative output in the long run in several ways. A natural starting point, grounded in recent theories of innovation, is to consider implications for the productivity of existing inventors and for the creation of new inventors (Bell et al., 2019*a*; Akcigit et al., 2018). In both cases, mechanical channels are

[23] By construction, the mean of the dependent variable – critical for assessing the relative magnitude of the estimated coefficients – declines with each iteration, as the threshold for a patent being classified as important increases. The fact that we estimate a gradient between effect size and patent importance even without explicitly accounting for this by rescaling the coefficients and the confidence intervals by the mean of the dependent variable in the corresponding estimation sample indicates that, in relative terms, the effect is substantially concentrated among breakthrough patents.

Figure 4

WWI deaths and patenting – Important patents



Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent whose innovativeness (measured as in Kelly et al. (2021)) is in the top deciles of the distribution of patents published at the same time is registered by a resident of parish i during 5-year period t. Each coefficient refers to a different regression, so that the left-most coefficient reflects the estimates reported in Table 1 and the right-most coefficient is an indicator equal to 1 if at least one patent registered belongs to the 10 percent of the most important patents published during the 5-year period t. Estimates include time, parish, and county-time fixed effects. 95% confidence intervals constructed using standard errors clustered at the parish level.

possible. War deaths may directly reduce innovation if inventors are killed. Even when they are not, a large mortality shock could persistently reduce local population, for example if villages are abandoned. To the extent that new inventors scale with population, that would mechanically reduce future innovation.

We explore these channels in Table 2. We define a "new" inventor as a patentee who did not appear in the sample earlier, and a pre-existing inventor as one who registered at least one patent in the pre-WWI period.[24] column 1 estimates the effect of WWI deaths on patenting by pre-WWI inventors; the effect is large and negative, suggesting that war mortality negatively affected previously active inventors. In column 2 we construct our outcome

[24]In our data we observe the details of patents but not the decisions of individuals to become inventors. This means our variable capturing a patent by a new patentee has two limitations. First, it may incorrectly include patents by inventors who last patented before our sample period starts. Second, it does not count those who become inventors but do not generate a patent. For this reason we can think of the variable as capturing successful new inventors.

restricting to patents registered by inventors who we can confirm were not killed in WWI,[25] and find that results are broadly unchanged. Columns 3 and 4 focus on new inventors. Column 3 shows that patenting by new inventors also declines in high-mortality parishes. The coefficient on population is positive as expected. While population may be an outcome and therefore a "bad control", its inclusion does not change our main estimate. Taken together, the results suggest that the war shock reduced both the productivity of existing inventors and the rate at which parishes produce successful new inventors, for reasons beyond mechanical death or population effects. We explore the mechanisms that underpin these results further in the next sections.

TABLE 2
WWI DEATHS AND PATENTING – OLS ESTIMATES – NEW AND EXISTING INVENTORS

	Any patent existing inv.	Any patent existing inv.	Any patent new inv.	Any patent new inv.
$Log(d^{WWI})\times$ Post WWI	-0.053***	-0.040***	-0.018***	-0.018***
	(0.002)	(0.002)	(0.001)	(0.001)
Log Population				0.055***
				(0.003)
Sample	Full	Not killed in WWI	Full	Full
Mean dep. var.	0.15	0.13	0.19	0.19
R2	0.46	0.46	0.45	0.45
Observations	162004	162004	162004	162004

Notes: OLS estimates of the effect of WWI deaths on parish-level patenting. Column 1: outcome is an indicator equal to one if at least one patent is registered by a resident of parish i during 5-year period t who was already patenting before 1915. Column 2: same outcome, excluding inventors matched to CWGC fatalities (to assess mechanical death effects). Column 3: outcome is an indicator equal to one if at least one patent is registered by a resident of parish i during 5-year period t who is a "new" inventor (i.e., one who did not appear in the sample earlier). Column 4: same as column 3, adding a control for log parish population. Parish, time, and county-time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

6. Human-capital shocks and the productivity of established inventors

Motivated by the evidence previously shown that active inventors were also affected, in this section we move to an inventor-level analysis using the individual-level panel of inventors covering the period 1895–1979. Leveraging the detailed information available in this dataset, we are able to examine heterogeneity in the effects of WWI mortality on innovation across several dimensions. In particular, we are able to explore the impact of WWI mortality on the production of patents by inventors with different backgrounds, fields of specialisation, co-authorship patterns and their location.

[25]We match inventors to the CWGC WWI fatalities list using full first and surname. This approach is conservative, in that we may misclassify some surviving inventors as killed when they share a name with a deceased soldier.

6.1. Baseline inventor-level results

We start by estimating the baseline DiD model in equation 3 by OLS. Column 1 of Table 3 reports the results and shows a negative effect, in line with our parish-level estimates. In column 2, we restrict the sample to patentees who were not killed in WWI.[26] The effect is very similar in the restricted sample, which reassures us that our parish-level results are not driven by a mechanical fall in patenting due to the death of inventors. Instead, they reflect the negative impact of WWI mortality on both surviving inventors who were active before the war and on individuals who became inventors afterwards. In column 3, the outcome is an indicator for registering a "breakthrough" patent, defined as one whose importance (calculated following Kelly et al. 2021) exceeds the 80th percentile of the distribution of all patents registered in the past 5 years. In column 4, we use a citation-based measure of importance, defining a highly cited patent as one that receives more citations than 80 percent of all patents registered in the past 5 years.[27] Under both the breakthrough and citation-based measures, we find that inventors in higher-mortality areas are less likely to produce highly innovative patents.[28] Compared with the sample mean of each outcome, the point estimates in columns 3 and 4 imply that the decline in high-quality patenting is proportionally larger than the decline in patenting of any type (column 1).

TABLE 3
WWI DEATHS AND PATENTING – INVENTOR-LEVEL ESTIMATES

	Any patent	Any patent	Any breakthr. patent	Any highly cited patent
$Log(d^{WWI})\times$ Post WWI	-0.001***	-0.002***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Sample	Full	Not killed in WWI	Full	Full
Mean dep. var.	0.12	0.13	0.03	0.02
R2	0.60	0.58	0.19	0.16
Observations	924477	489549	924477	924477

Notes: OLS estimates of the effect of WWI deaths on an indicator equal to one if inventor j from parish i registers a patent during 5-year period t. Column 1 presents results from equation 3 for the full sample, whereas in column 2 the sample is restricted to individuals who do not appear in the CWGC dataset of WWI fatalities, matched using the initial of the first name and surname. In columns 3 and 4, the outcomes are, respectively, an indicator for the inventor registering a "breakthrough" patent or a highly-cited patent (see text for details). Individual, time, and year of first patent by time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

The corresponding event-study results are reported in Figure A.6 in the Appendix, and show no evidence of pre-WWI differences in the outcomes and a negative, persistent effect of

[26]To identify these, we match patentees to the CWGC WWI fatalities list using the initial of the first name and the full surname. This approach is conservative, in that it may incorrectly classify some surviving inventors as having been killed when they share a first-name initial with a deceased soldier. Alternative matching methods (e.g., using more initial letters or the full first name) yield very similar results; see Appendix Table A.3.

[27]This measure is imperfect because citation counts in PATSTAT have been systematically recorded only since the 1980s, so highly influential patents that lost relevance before then may go undetected.

[28]Using one- or ten-year windows instead of 5 years yields similar results; we omit them for brevity.

WWI mortality on innovation in all specifications.

6.2. The negative effect of mortality on highly innovative and complex patents

By concentrating on inventors who were already active before WWI, we can identify which fields of knowledge experienced a greater impact from the mortality shock. To do this, we estimate an extended version of equation (3) in which the $log(d_i^{WWI}) \times Post$ term is interacted with a series of indicators, one for each inventor's pre-war field of specialisation, measured by the one-digit CPC code of the first patent they produced in the pre-War period. This approach enables us to break down the average effect of WWI mortality by field, revealing how the shock differently affected areas that require different skills and expertise to generate innovation.

Results are presented in Figure 5, which reports the estimated coefficients for each interaction term. Results show substantial heterogeneity across fields, with the largest effects in electricity, mechanics, transportation, physics and chemistry. These science-intensive sectors, at the forefront of innovation at the time, required highly trained scientists and the execution of increasingly complex tasks. The need to commercialise scientific breakthroughs made laboratory research, specialised skills, and coordinated teams and infrastructure essential (Edgerton and Horrocks, 1994).[29]

Figure 6 examines heterogeneity in the effects of mortality on patenting by different measures of complexity and importance of the inventor's first patent. The figure illustrates that the negative effect of WWI deaths on patenting is not uniform: it is significantly more pronounced among inventors whose early work or training marked them as central to highly innovative or complex fields.

We start by contrasting inventors who recorded engineering as their occupation at the time of registering their first patent with those who were not. Results indicate that engineers experienced a notably larger negative impact from mortality, underscoring that the loss of human capital due to the War was particularly damaging to highly skilled inventors.

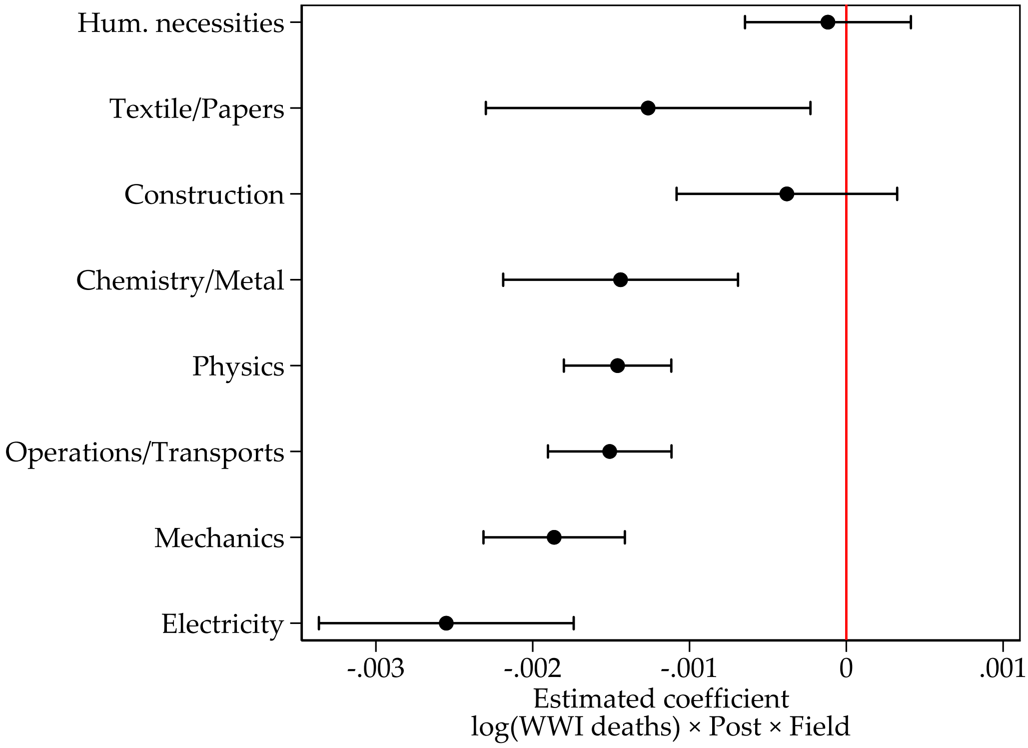
In addition, we examine heterogeneity by dividing inventors according to the complexity and importance of their first patent. Specifically, inventors whose initial patent falls above the median in either complexity (proxied by abstract length) or importance (measured using the metric from Kelly et al. 2021) experience greater declines in subsequent patenting activity, although in the latter case the difference is marginal and not statistically significant.[30] The last coefficients reported in the figure report results using the first principal component of patent complexity and importance, providing a summary index. These estimates reiterate that the negative impact of the mortality shock is most pronounced among inventors whose

[29]These sectors were also those where engineers were mostly active, as shown in Appendix's Figure A.8. Figure A.9 in the Appendix reports scatter plots of the share of engineers among inventors against two proxies for technological complexity: the length of the patent's abstract (panel A), and patent importance, computed following Kelly et al. (2021) (panel B). In both cases, we observe a strong positive relationship: patents that are more complex and more influential were disproportionately produced by engineers.

[30]Throughout our study period, the length of patent abstracts was not regulated, resulting in considerable variation. Prior to 1978, providing an abstract was optional; nevertheless, fewer than 9 percent of patents in our PATSTAT-based sample lack an abstract.

FIGURE 5

HETEROGENEOUS EFFECTS BY FIELD OF SPECIALISATION



Notes: Inventor-level OLS estimates of the effect of WWI deaths on an indicator equal to one if at least one patent is registered during 5-year period t, interacted by the field of specialisation of each inventor, as proxied by the field of specialisation of the first patent produced by the inventor in the pre-War period (panel a). Estimates from equation 3 where the term $log(d_i^{WWI} \times Post)$ has been interacted with each field, as reported on the vertical axis. Individual, time, and year of first patent by time fixed effects are included. 95% confidence intervals constructed using standard errors clustered at the parish level. Fields of specialisation are assigned according to the NLP technique explained in section 3.

early work was more innovative or complex.

Taken together, these results indicate that the decline in innovation after WWI was disproportionately concentrated among inventors operating at the technological frontier or engaged in producing more complex inventions. This pattern underscores how the human-capital shock induced by WWI had profound and lasting consequences for British technological progress.

6.3. *The mitigating effects of mobility and co-authorship*

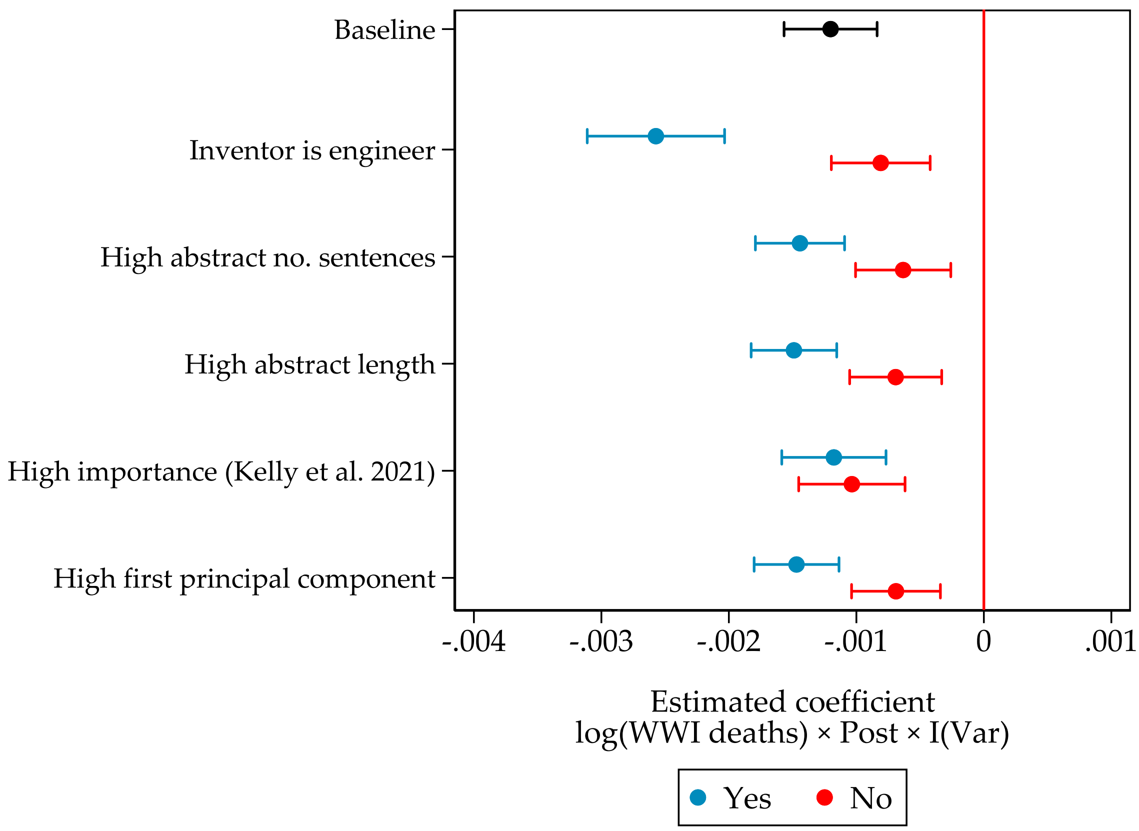
The long-run decline in innovation following WWI mortality suggests that the human-capital losses undermined both incumbent and future inventors, possibly by disrupting co-authorship networks and degrading the broader innovation ecosystem through the loss of complementary workers.

We explore this mechanism using our sample of inventors in Figure 7. In each column, we redefine the dependent variable to capture how changes in inventor behaviour induced by the WWI mortality shock affect inventors' productivity.[31] In column 2 we redefine our

[31]To facilitate interpretation across columns, all coefficients are rescaled by the corresponding sample mean

Figure 6

Heterogeneous effects by patent complexity



Notes: This figure presents inventor-level OLS estimates of the impact of WWI deaths on patenting outcomes, examining how effects vary according to characteristics of the inventor's first patent (see main text for details). The first row displays the overall (uninteracted) effect of WWI mortality. The following rows show estimated effects separately for inventors who reported being engineers versus non-engineers at the time of their first patent. Additional rows present effects for inventors whose first patent was of high or low complexity, defined as above or below the median based on measures such as abstract length or sentence count. The figure further distinguishes between high and low patent importance, and includes a summary using the first principal component of all measures. The outcomes are binary indicators for having patented in a given period, and coefficients are reported for each interaction term. Individual, time, and year of first patent by time fixed effects are included in all specifications. 95% confidence intervals constructed using standard errors clustered at the parish level.

patenting indicator to take value 1 if the inventor publishes at least one patent within the inventor's original field – as defined by the field of the inventor's first patent published prior to WWI, while in column 3 the dependent variable is equal to 1 if the inventor publishes at least one patent within a different field. Both estimates are negative and relatively similar to the baseline effect (reported for comparison in column 1), although the one for patenting in a new sector is larger in magnitude, suggesting that mid-career retraining into a new technical field is not effective as a response to the human-capital shock.

In columns 4 and 5 we perform a similar decomposition but now for geographic mobility. In column 4, the dependent variable takes value 1 if the inventor publishes at least one patent

of the outcome. See Table A.4 in the Appendix for full, unrescaled, underlying estimates.

while residing in the same parish they used to live in upon publishing their first patent. In column 5, instead, the dependent variable is equal to 1 if the inventor publishes at least one patent upon relocating to a different parish. Here we observe a clear margin of adjustment: higher WWI mortality reduces the probability of patenting in the original parish but raises the probability of patenting in a new one. This pattern accords with recent evidence showing positive effects of being exposed to high-innovation areas on becoming inventors and patenting. Similarly to Bell et al. (2019*b*) and Moretti (2021), the positive effect we document for movers could be due to gaining access to better networks, role models or mentors, as well as accessing specific infrastructures and complementary workers, who appear to be crucial for technologically advanced research. Importantly, these results suggest that relocation of successful inventors is another important channel through which patenting activity at the parish level is negatively hit by the human capital shock.

Finally, in columns 6-7 the dependent variables take the value 1 if the inventor publishes at least one single-authored patent and at least one co-authored patent, respectively. We find that the negative mortality effect on single-authored patents again mirrors the baseline, but its impact on co-authored patents is statistically indistinguishable from zero. As in Jaravel, Petkova and Bell (2018), this suggests that working in teams provides a buffer against human-capital losses – co-authors help sustain productivity – while solo projects are more vulnerable to negative productivity shocks and might be more likely to be de-prioritised.

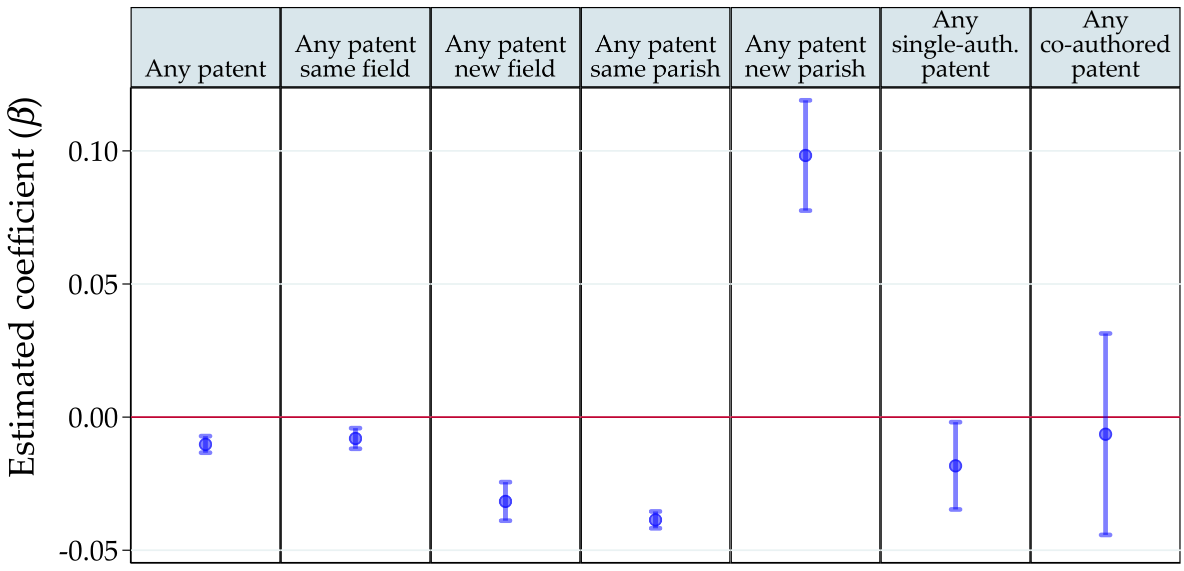
In sum, our results provide evidence consistent with recent research on exposure effects. Relocating to communities less affected by human-capital losses enables inventors to offset the negative impact on their productivity, likely by accessing a larger and higher-quality pool of potential collaborators and co-workers. At the same time, we confirm the importance of team-specific innovative capital, as previously documented for the United States (Jaravel, Petkova and Bell, 2018), by showing that inventors who relied more heavily on teamwork were better able to mitigate the adverse effects of the shock.

7. Human-capital shocks and the supply of new inventors

In this section, we return to the parish-level analysis to explore how human-capital shocks affect the local supply of new inventors. As shown earlier in Table 2, patent outputs by new inventors declined at the parish level in response to exposure to WWI mortality and this effect is not driven by changes in the scale of local population. The evolution of this outcome measure may capture both the entry/exit of individuals into research activities or changes in the productivity of researchers who have yet to publish a patent. As such we can use it to shed light on mechanisms that affect at least one of these phenomena, although typically we will be unable to distinguish between them. We start by examining how the wartime loss of social elites and skilled engineers affects the cultivation of successful new inventors. We then explore the extent to which pre-existing institutions and innovation ecosystems can mitigate the effects of mortality on the same outcome.

FIGURE 7

INVENTOR-LEVEL ESTIMATES – SECTORS, MOVERS, AND CO-AUTHORSHIP



Notes: $N = 924,528$. Inventor-level OLS estimates of the effect of WWI deaths on patenting from eq. 3. All coefficients are rescaled by the corresponding sample mean of the outcome to facilitate comparability across specifications. See Table A.4 in the Appendix for the underlying unrescaled estimates. Column 1 reports the baseline estimate (analogous to column 1 of Table 3) for comparison. In column 2 (resp., 3), the dependent variable is an indicator taking the value 1 if inventor j registers at least one patent during 5-year period t within the same sector (resp., within a different sector) as the one the inventor first patented in. In column 4 (resp., 5), the dependent variable is an indicator taking the value 1 if inventor j registers at least one patent during 5-year period t while residing in the same parish (resp., while residing in a different parish) as the one in which they were residing upon registering their first patent. Finally, in column 6 (resp., 7), the dependent variable is an indicator taking the value 1 if inventor j registers at least one single-authored (resp., co-authored) patent during 5-year period t. Notice that inventor j in the same 5-year period t can produce patents within different fields, residing in multiple locations, and with different co-authorship teams. Individual fixed effects, time fixed effects, and year of first patent by time fixed effects are included in all specifications. 95% confidence intervals constructed using standard errors clustered at the parish level.

7.1. The importance of social elites and engineers

As previously discussed, a very high proportion of inventors in the first half of twentieth century Britain classified themselves as engineers. Although we do not have any data to confirm it, it also seems possible that many of those who would go on to become inventors had attended university. We therefore expect that losing social elites and men with engineering skills could reduce the supply of future inventors. Besides this, we may also expect that engineers could serve as role models for potential future innovators (Bell et al., 2019*b*). To test the extent to which the loss of talented and skilled individuals affects the production of new inventors, we therefore extend our baseline model of eq. 1 (as estimated in column 3 of Table 1) as follows:

$$y_{it}^{New} = \alpha + \beta_1 \mathbb{1}(d_i^{WWI} > med) \times Post + \beta_2 \mathbb{1}(HC_i) \times Post + FE + \epsilon_{it}, \tag{4}$$

where $\mathbb{1}(HC_i)$ is an indicator variable for human capital that takes the value one when the parish experiences any deaths of individuals that have skills and characteristics that are associated with the generation of new ideas. We will define this indicator in different ways using the information contained in the mortality records. To match the nature of these variables, we include total WWI deaths in this regression using an indicator for above median value. The coefficient of interest is β_2, which measures the effect on innovation of the loss of these skills, conditional on the parish being a high or low death location.

Results are reported in Table 4. In column 1, we define $\mathbb{1}(HC_i)$ as an indicator for the parish having lost at least one individual belonging to the social elite, defined as those who studied at Oxford or Cambridge, or an elite school. The coefficient is negative and significant and indicates that the loss of a social elite reduces the likelihood of subsequent patenting by more than 3 percentage points. In columns 2 and 3, we examine the effect of deaths of officers. In column 2, we replace the social elite indicator with an indicator for death of any officer, finding again a negative effect. In column 3, we divide officers into two groups, engineers and other officers. Results indicate that the loss of an officer with an engineering background has a much larger negative effect on subsequent patenting than the loss of other types of officers, for which the effect is small and insignificant.

In column 4, we add back our indicator for losing a social elite and show that both social-elite and engineer deaths matter. The loss of these groups thus has effects on subsequent patenting over and above the loss of other men. The estimate on engineers is in line with recent evidence from the UK and the Americas showing that engineers are key drivers of innovation and growth and produce more patents than other types of innovators (Hanlon, 2022; Maloney and Valencia Caicedo, 2022). In column 5, we address the concern that the results might merely reflect pre-war differences in the local supply of engineers and white-collar workers by adding interactions between the $Post$ dummy and (i) the share of adult men in engineering occupations in 1911 and (ii) the share of adult men in white-collar occupations in 1911. The estimates are unchanged, and the coefficient on overall WWI mortality remains precisely estimated. This suggests that our mortality measure captures both the loss of skilled men that we observe in the data and unobserved variation in the loss of talented men.

7.2. Heterogeneity and the mitigating role of institutions

Previous work demonstrates that the presence of other inventors, transport links, communications technologies, and institutions can drive the generation of new ideas (see e.g., Akcigit, Grigsby and Nicholas 2017*b*; Bell et al. 2019*b*; Andrews 2023). Local access to these resources in the period prior to WWI could potentially mitigate the effects of war deaths on innovation. Specifically, if these ecosystem factors can substitute for human-capital inputs in the production of new ideas, we may expect to find that the effect of war deaths is smaller in places where they are more accessible. We conduct heterogeneity analysis to test for this using four county-level accessibility measures and our parish level data. In all cases, we separate out counties with above-median access to an ecosystem factor, and test for differential effects relative to the overall average effect by adding a triple interaction term to our baseline

TABLE 4

WWI DEATHS AND PATENTING – OLS ESTIMATES – IMPORTANCE OF ELITES

	Any patent by new inventor				
	(1)	(2)	(3)	(4)	(5)
I(Social elite) x Post	-0.033***			-0.016*	-0.017*
	(0.009)			(0.010)	(0.010)
I(Officer) x Post		-0.010**			
		(0.004)			
I(Engineer off.) x Post			-0.050***	-0.047***	-0.044***
			(0.008)	(0.008)	(0.008)
I(Non eng off) x Post			-0.006		
			(0.004)		
$I(d^{WWI} > med)\times$ Post	-0.032***	-0.031***	-0.024***	-0.025***	-0.025***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Mean dep. var.	0.19	0.19	0.19	0.19	0.19
R2	0.45	0.45	0.45	0.45	0.45
Observations	162004	162004	162004	162004	160933
1911 Occupation x Post controls	N	N	N	N	Y

Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by "new" inventors (patentees who did not appear in the sample earlier) resident of parish i during 5-year period t. Social elite is an indicator for losing one or more elite school graduates. Similarly, officer is an indicator for losing one or more officers. Engineer is an indicator for losing one or more officers with engineering background. Parish, time, and county-time fixed effects are included in all specifications. Column 5 additionally controls for $Post$ interacted with i) share of adult men in engineering occupations in 1911 and ii) share of adult men in white collar occupations in 1911. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

specification. Note that the county-time effects in this specification absorb the two-way interaction between above median accessibility and post WWI so we are effectively conducting comparisons strictly within places with comparable baseline accessibility to each ecosystem factor. In addition, to absorb the effects of density in general we also control for post times above-median accessibility measure constructed using distance to churches.

Results are reported in Table 5. In column 1, we find that the effect of war deaths is significantly smaller – approximately half as large in absolute terms – in counties with above-median access to engineering role models (proxied by members of engineering institutes). In column 2, we examine proximity to universities and find a similar pattern. However, in columns 3 and 4 we find there is small or no effect of proximity to rail stations and post offices. Notably, across all columns and as expected, we find that the mean of the dependent variable is considerable larger in places with better access to ecosystem factors, consistent with the idea that these factors are positively associated with patenting. Overall, these results indicate that the effects of war deaths are in fact considerably smaller in some places, and in particular suggests that physical proximity to universities and networks of experts can mitigate the effects of shocks, at least to some degree.

To conclude, this section shows that the impact of war deaths on innovation is amplified by the loss of talented and skilled individuals, and mitigated by the presence of institutions

TABLE 5
WWI DEATHS AND PATENTING – OLS ESTIMATES – HETEROGENEITY

	Any patent by new inventor			
	(1)	(2)	(3)	(4)
$Log(d^{WWI})\times$ Post WWI	-0.026***	-0.028***	-0.024***	-0.023***
	(0.002)	(0.002)	(0.002)	(0.002)
x I(Access > med)	0.006**	0.008***	0.002	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)
Access factor	Engineers	Univer.	Stations	Post Off.
Mean dep.var.(Access < med)	0.12	0.13	0.12	0.16
Mean dep.var.(Access > med)	0.25	0.24	0.26	0.22
R2	0.45	0.45	0.45	0.45
Observations	162004	162004	162004	162004

Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by "new" inventors (patentees who did not appear in the sample earlier) resident of parish i during 5-year period t, by pre-WWI characteristics defined at the district level. Parish, time, and county-time fixed effects are included in all specifications. In all cases we additionally control for the interaction between post and above median access to churches. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

and ecosystem factors that can substitute for or sustain human capital in the production of new ideas. These results suggest that the loss of talented and technically skilled servicemen had persistent local effects: communities lost not only potential future inventors and role models, but also part of the human-capital base that supports the production of new ideas. While this evidence is indirect, it is consistent with the idea that the loss of talented men in the war affected the production of talented individuals in the next generation, by reducing the local supply of innovators and rendering these places less attractive for new inventors.

8. Robustness checks

8.1. Strong parallel trends

As discussed in Section 4, difference-in-differences analysis using a continuous treatment variable requires assuming a stronger version of the parallel-trends assumption (Callaway, Goodman-Bacon and Sant'Anna, 2024), which cannot be directly tested. To assess its plausibility in our setting, we construct an empirical exercise that mimics as closely as possible the theoretical comparison across all possible mortality levels, before and after the treatment. Specifically, to produce Figure A.5 in the Appendix, we divide the sample based on deciles of the distribution of $log(d_i^{WWI})$ and estimate event studies specifications in which we compare every decile of $log(d_i^{WWI})$ with each other (one comparison at a time) and using a binary treatment indicator (i.e., being in the higher decile of the two). Then, for each regression, we formally test for joint statistical significance of the pre-war interaction coefficients and save the p-values. Figure A.5 presents the distribution of p-values obtained performing all the regressions. Only one of the attempted regressions shows a statistically significant imbalance in pre-treatment coefficients ($p = 0.048$). We conclude that the strong parallel trends assumption is likely to hold in our setting.

8.2. IV results

In this subsection, we discuss an alternative instrumental variable strategy that does not rely on any of the parallel trends assumptions needed for the diff-in-diffs approach. Specifically, we instrument the log of WWI deaths with the log of predicted deaths based on battalion-level mortality rates. The instrument is constructed by aggregating battalion-level mortality to the parish level, using the share of a parish's servicemen assigned to each battalion as weights. The intuition behind this approach is that during WWI, servicemen were assigned to battalions – fighting units roughly 1,000 strong – in a sequential manner based on enlistment order (Bet-El, 2009; Carozzi, Pinchbeck and Repetto, 2023). This sequential assignment is plausibly exogenous, as it is driven by the order of arrival rather than by underlying parish characteristics. Consequently, serving in one battalion rather than in the next is unlikely to be correlated with pre-existing determinants of innovation at the parish level.[32]

Specifically, indexing battalions with b and parishes with i, we instrument $\log(d_i^{WWI})$ with

$$z_i = \log\left(m_i \sum_{b=1}^{B} \alpha_{ib}\tilde{\delta}_{ib}\right),$$

where m_i is mobilisation from parish i; $\alpha_{ib} = m_{ib}/m_i$ is the share of parish i's mobilised who served in battalion b (with m_{ib} mobilised from i in b); d_b and m_b are deaths and mobilisation in battalion b; and $\tilde{\delta}_{ib} = \frac{d_b - d_{ib}}{m_b - m_{ib}}$ is battalion b's death rate excluding parish i (leave-out mean), with d_{ib} denoting deaths from parish i in battalion b.

This instrument has a shift-share structure, with shares α_{ib} and shocks δ_b, estimated by the leave-out means $\tilde{\delta}_{ib}$. Because, in this setting, variation in mortality across battalions is likely to be driven mostly by where they were deployed and by the fortunes of war, we follow the approach that assumes shocks are exogenous (Borusyak, Hull and Jaravel, 2022). Formally, our identification assumption is that our measures of battalion-level mortalities are conditionally uncorrelated with other parish-level determinants of patenting activity. [33]

Borusyak, Hull and Jaravel (2022) show that in shift-share designs with exogenous shocks, the IV orthogonality condition can be formulated either as requiring that the instrument is uncorrelated with parish-level unobservable determinants of the outcome or, equivalently, as imposing that the shocks – here, battalion-level mortality rates – are uncorrelated with shock-level unobservables.

[32]Construction of the instrument is possible only for battalions and parishes that appear in both the CWGC-FWR data and the FamilySearch data. As a consequence, the sample size is reduced by about one third in the IV results. For additional technical details on the instrument's construction, we refer the reader to Carozzi, Pinchbeck and Repetto (2023).

[33]This assumption is weaker than requiring that random assignment of servicemen to battalions. In fact, the necessary and sufficient condition for instrument exogeneity is requiring the death rates δ_{ib} not to be systematically related to any other pre-WWI parish-level characteristic that determine present or future innovation.

In Figure A.10, we provide evidence in favour of this orthogonality assumption using either approach. In panel A, we regress the instrument on several parish-level pre-WWI characteristics.[34] Most coefficient estimates are close to zero and statistically insignificant at conventional levels, indicating that the instrument is not correlated with observable characteristics that could affect patenting activity. In panel B of Figure A.10, we conduct a similar exercise after aggregating the data at the battalion level, using the shares of servicemen from each parish serving in a battalion as weights (Borusyak, Hull and Jaravel, 2022).[35] This amounts to regressing battalion-level mortality shocks (δ_{ib}) on the economic, demographic or geographical characteristics of the community of origin of soldiers that make up these battalions. Once again, we find evidence in favour of the orthogonality assumption. Battalion-level death rates are uncorrelated with pre-determined parish level characteristics.

We confirm the relevance of the instrument for our parish-level analysis in Panel A of Table 6, which shows large first-stage coefficients across all four specifications that mirror those presented in Table 1. F-statistics are well above 1000 in all specifications, suggesting the instrument is strong throughout.

The identification result in Borusyak, Hull and Jaravel (2022) also requires that the number of shocks be sufficiently large and that they be sufficiently dispersed in terms of their average exposure. We follow their recommendation and report the inverse of the Herfindahl Index of shock-level average exposure as a measure of the effective sample size. In our dataset this equals 141, indicating that shocks are well dispersed and that the asymptotic results in that paper apply.

Panel B of Table 6 presents the main parish-level IV estimates. Across the four specifications, the estimated effects of WWI deaths are very similar to the corresponding OLS estimates reported in Table 1, but slightly larger in magnitude. The event-study results, shown in Figure 8, are also qualitatively similar to the OLS counterpart in Figure 3. In particular, pre-trend coefficients are close to zero and statistically insignificant, further supporting the assumption that the instrument is uncorrelated with pre-WWI determinants of innovation. Post-WWI coefficients indicate an immediate drop in innovative activity during the war, a recovery in the interwar period, and another reduction during WW2 followed by a modest rebound. We conclude that endogeneity of WWI deaths does not appear to be a primary concern for our main results. The instrumental variable estimates for the inventor-level analysis also confirm the OLS Difference-in-Differences results and are reported in Table A.5 in the Appendix.

8.3. Additional robustness checks

We next summarise additional robustness checks that address several potential concerns. First, the results in Table 1 and Figure 3 focus on a binary measure of innovation, because most of the variation in patenting is between parishes that patent and those that do not.

[34] All specifications include the logarithms of 1911 population and WWI mobilisation, and historic county dummies.

[35] This aggregation is performed using the *ssaggregate* command in Stata 17.

Table 6

WWI deaths and patenting – parish-level IV estimates

	$Log(d^{WWI})$ × Post	$Log(d^{WWI})$ × Post	$Log(d^{WWI})$ × Post	$Log(d^{WWI})$ × Post	$Log(d^{WWI})$ × Post
A. First-stage					
$z\times$ Post	0.925***	0.922***	0.909***	0.804***	0.550***
	(0.007)	(0.007)	(0.007)	(0.009)	(0.013)
F-stat	19327	18222	15686	8234	1668
Observations	111406	111406	111406	111406	110765

	(1) Any patent	(2) Any patent	(3) Any patent	(4) Any patent	(5) Any patent
B. IV (second stage)					
$Log(d^{WWI})\times$ Post	-0.014***	-0.014***	-0.015***	-0.012***	-0.013**
	(0.002)	(0.002)	(0.002)	(0.003)	(0.006)
$Log(d^{County})\times$ Post		0.000			
		(0.003)			
Mean dep.var.	0.31	0.31	0.31	0.31	0.31
Observations	111406	111406	111406	111406	111406
Parish FE	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y
County × Time	N	N	Y	Y	Y
Controls × Linear trend	N	N	N	Y	N
County × Time $\times Q_n$ Pop	N	N	N	N	Y

Notes: Panel A: IV first-stage estimates of the log number of WWI deaths on the instrument, both interacted with an indicator for time periods after 1915-19. Panel B: IV estimates of the effect of WWI deaths on the probability that at least one patent is registered by a resident of parish i during 5-year period t. Specifications follow those in Table 1. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

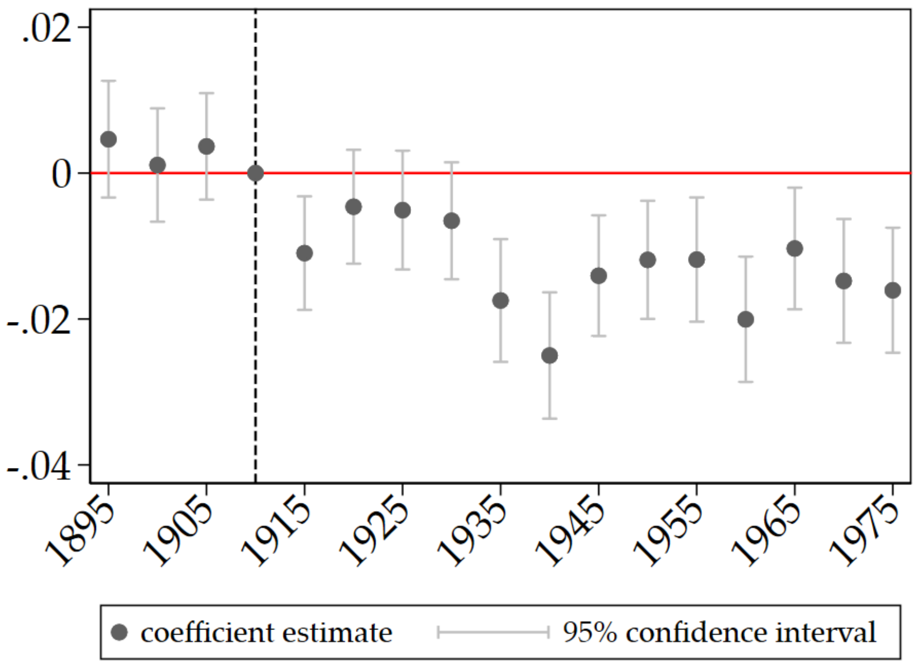
However, our parish-level findings are not confined to the extensive margin. In Figure A.11 we estimate a series of regressions where the dependent variable is an indicator for whether the parish patent rate (number of patents per capita, using the 1911 Census population) exceeds various percentiles of the overall distribution. Across almost all percentiles, higher WWI mortality reduces the probability that a parish reaches higher patenting rates.[36] Notably, the negative effects on being above the 80th or 90th percentile are larger than the effect at the extensive margin, suggesting that WWI mortality had a sizeable impact not just on whether parishes patent at all, but also on the intensity of innovation (the intensive margin).

Second, Table A.2, Panel A, shows that our baseline results are robust to alternative outcome measures, including the total number of patents and patents per 1,000 inhabitants (based on the 1911 Census population). Panel B replaces d^{WWI} with the parish-level mortal-

[36]For percentiles below the 77th, these coefficients are equivalent to those in column 2 of Table 1.

Figure 8

WWI deaths and patenting – Event-study – IV



Notes: IV estimates of the effect of WWI deaths on the probability that at least one patent is registered by a resident of parish i during 5-year period t. Estimates include time and parish fixed effects and county-time fixed effects. 95% confidence intervals constructed using standard errors clustered at the parish level.

ity rate, defined as the ratio of d^{WWI} to the 1911 population. Panel C instead defines treatment as an indicator for mortality in the parish being above the median, $I(d^{WWI} > med)$, and estimates a binary difference-in-differences specification. Table A.2 also reports Poisson pseudo maximum likelihood estimates, following Silva and Tenreyro (2006), which are well suited to count outcomes with a large share of zeros (Chen and Roth, 2023). Across all specifications, the estimates remain consistent with our baseline results using the binary outcome.

Third, our baseline treatment uses the log of WWI deaths, which excludes parishes with zero fatalities. Table A.6 and Table A.7 re-define treatment as $\log(1 + \text{deaths})$, thereby retaining zero-death parishes (and inventors active there). The results are unchanged.

Fourth, the main analysis aggregates outcomes to 5-year periods. Figures A.12 (parish-level) and A.13 (inventor-level) replicate the event-study specifications using two-year intervals, and the qualitative pattern mirrors the baseline.

Lastly, in Table A.8 and Table A.9, we restrict the sample to periods before World War II (WWII) to rule out that our results are driven by the effect of WWI mortality on WWII mortality (Carozzi, Pinchbeck and Repetto, 2023). Consistent with our dynamic specifications, which show that the effect of WWI mortality on innovation materialises shortly after the war, the results are unchanged under this restriction.

9. Conclusions

This paper studies how WWI military mortality shocks affected innovation in Britain over the long run. Linking parish-level deaths to patent and inventor data, we find that communities with higher mortality experienced persistent declines in inventive activity. These declines are larger for breakthrough and technically advanced innovation, indicating that the shock affected not only the quantity but also the quality of invention.

This evidence is consistent with a human-capital supply channel. The effects are stronger when losses include highly skilled individuals, such as engineers and elite schools alumni, and mitigated by favourable pre-existing ecosystems. Our findings speak to the broader literature on labour scarcity and technological change. Rather than challenging evidence that labour scarcity can stimulate labour-saving innovation at broader geographic scale, our results highlight a distinct margin: localised losses in inventor human capital can depress innovative capacity for decades.

At the inventor level, we find lower post-war productivity among exposed inventors, especially for those working in sectors that rely more heavily on technical expertise and complex knowledge and those who remain in place. Mobility and co-authorship attenuate these effects, while sector switching does not, pointing to the importance of local complementarities and innovation networks.

References

Abramitzky, Ran, Adeline Delavande, and Luis Vasconcelos. 2011. "Marrying up: the role of sex ratio in assortative matching." *American Economic Journal: Applied Economics*, 3(3): 124–57.

Acemoglu, Daron. 2002. "Directed Technical Change." *Review of Economic Studies*, 69(4): 781–809.

Acemoglu, Daron. 2007. "Equilibrium bias of technology." *Econometrica*, 75(5): 1371–1409.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60(2): 323–351.

Akcigit, Ufuk, John Grigsby, and Tom Nicholas. 2017*a*. "Immigration and the rise of American ingenuity." *American Economic Review*, 107(5): 327–331.

Akcigit, Ufuk, John Grigsby, and Tom Nicholas. 2017*b*. "The rise of american ingenuity: Innovation and inventors of the golden age." National Bureau of Economic Research.

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi. 2018. "Dancing with the stars: Innovation through interactions." National Bureau of Economic Research.

Andersson, David, Mounir Karadja, and Erik Prawitz. 2022. "Mass migration and technological change." *Journal of the European Economic Association*, 20(5): 1859–1896.

Andrews, Michael J. 2023. "How do institutions of higher education affect local invention? Evidence from the establishment of US colleges." *American Economic Journal: Economic Policy*, 15(2): 1–41.

Andrews, Michael J, and Alexander Whalley. 2022. "150 years of the geography of innovation." *Regional Science and Urban Economics*, 94: 103627.

Audretsch, David B, and Maryann P Feldman. 1996. "R and D spillovers and the geography of innovation and production." *The American Economic Review*, 86(3): 630–640.

Bailey, Roy E, Timothy J Hatton, and Kris Inwood. 2023. "Surviving the Deluge: British servicemen in World War I." *Economics & Human Biology*, 49: 101216.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen. 2019*a*. "Joseph Schumpeter Lecture, EEA Annual Congress 2017: Do Tax Cuts Produce more Einsteins? The Impacts of Financial Incentives VerSus Exposure to Innovation on the Supply of Inventors." *Journal of the European Economic Association*, 17(3): 651–677.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen. 2019*b*. "Who becomes an inventor in America? The importance of exposure to innovation." *The Quarterly Journal of Economics*, 134(2): 647–713.

Bergeaud, Antonin, and Cyril Verluise. 2024. "A new dataset to study a century of innovation in Europe and in the US." *Research Policy*, 53(1): 104903.

Bergeaud, Antonin, Jean-Baptiste Chaniot, and Clément Malgouyres. 2025. "Escaping Labor Scarcity: Innovation and Human Capital after WW1 in France." Working Paper.

Bet-El, Ilana R. 2009. *Conscripts: forgotten men of the Great War.* The History Press.

Boehnke, Jörn, and Victor Gay. 2020. "The Missing Men World War I and Female Labor Force Participation." *Journal of Human Resources*, 0419–10151R1.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-experimental shift-share research designs." *The Review of Economic Studies*, 89(1): 181–213.

Brainerd, Elizabeth. 2017. "The lasting effect of sex ratio imbalance on marriage and family: Evidence from World War II in Russia." *Review of Economics and Statistics*, 99(2): 229–242.

Brakman, Steven, Harry Garretsen, and Marc Schramm. 2004. "The strategic bombing of German cities during World War II and its impact on city growth." *Journal of Economic Geography*, 4(2): 201–218.

BWO. 1922. "Statistics of the Military Effort of the British Empire During the Great War, 1914-1920."

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. 2024. "Difference-in-differences with a continuous treatment." National Bureau of Economic Research.

Carey, Gordon Vero. 1921. *The war list of the University of Cambridge: 1914-1918.* CUP Archive.

Carozzi, Felipe, Edward Pinchbeck, and Luca Repetto. 2023. "Scars of War: The Legacy of WWI Deaths on Civic Capital and Combat Motivation." CEPR, CEPR Press, Paris and London. https://cepr.org/publications/dp18343.

Chen, Jiafeng, and Jonathan Roth. 2023. "Logs with Zeros? Some Problems and Solutions*." *The Quarterly Journal of Economics*, qjad054.

Chupilkin, Maxim, and Zsoka Koczan. 2022. *The economic consequences of war: Estimates using synthetic controls.* European Bank for Reconstruction and Development.

Commonwealth War Graves Commission. 2023. "Online Casualty Database." [Online; accessed 1-Jan-2021].

Correia, Sergio. 2015. "Singletons, cluster-robust standard errors and fixed effects: A bad mix." *Technical Note, Duke University*, 7.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin. 2020. "Fast Poisson estimation with high-dimensional fixed effects." *The Stata Journal*, 20(1): 95–115.

Craig, Edwin Stewart. 1920. *Oxford University roll of service.* Clarendon Press.

Davis, D. R., and D. E. Weinstein. 2002. "Bones, bombs, and break points: The geography of economic activity." *American Economic Review*, 92(5): 1269–1289.

Edgerton, David EH, and Sally M Horrocks. 1994. "British industrial research and development before 1945 1." *The Economic History Review*, 47(2): 213–238.

European Patent Office. 2023. "PATSTAT Global – Autumn 2023." European Patent Office, Vienna, Austria.

FamilySearch. 2023. "United Kingdom World War I Service Records." [Online; accessed 1-Jan-2021].

Forces War Records. 2023. "Online Database." [Online; accessed 1-Jan-2021].

Franck, Raphaël. 2024. "Labor scarcity, technology adoption and innovation: evidence from the cholera pandemics in 19th century France." *Journal of Economic Growth*, 1–41.

Glaeser, Edward L, Hedi D Kallal, Jose A Scheinkman, and Andrei Shleifer. 1992. "Growth in cities." *Journal of political economy*, 100(6): 1126–1152.

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

Habakkuk, H. J. 1962. *American and British Technology in the Nineteenth Century: The Search for Labour-Saving Inventions.* Cambridge:Cambridge University Press.

Hanlon, W. Walker. 2015. "Necessity Is the Mother of Invention: Input Supplies and Directed Technical Change." *Econometrica*, 83(1): 67–100.

Hanlon, W Walker. 2022. "The rise of the engineer: Inventing the professional inventor during the Industrial Revolution." National Bureau of Economic Research.

Hicks, John R. 1932. *The Theory of Wages.* London:Macmillan.

James, Brigadier EA. 2012. *British Regiments 1914-1918.* Andrews UK Limited.

Jaravel, Xavier, Neviana Petkova, and Alex Bell. 2018. "Team-specific capital and innovation." *American Economic Review*, 108(4-5): 1034–1073.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. 2021. "Measuring technological innovation over the long run." *American Economic Review: Insights*, 3(3): 303–320.

Lin, Erin. 2022. "How war changes land: Soil fertility, unexploded bombs, and the underdevelopment of Cambodia." *American Journal of Political Science*, 66(1): 222–237.

Maloney, William F, and Felipe Valencia Caicedo. 2022. "Engineering growth." *Journal of the European Economic Association*, 20(4): 1554–1594.

Minnesota Population Center. 2019. "Integrated Public Use Microdata Series, International: Version 7.2 [dataset]."

Moretti, Enrico. 2021. "The effect of high-tech clusters on the productivity of top inventors." *American Economic Review*, 111(10): 3328–3375.

Nagaoka, Sadao, Kazuyuki Motohashi, and Akira Goto. 2010. "Patent statistics as an innovation indicator." In *Handbook of the Economics of Innovation*. Vol. 2, 1083–1127. Elsevier.

Neffke, Frank M. H. 2019. "The value of complementary co-workers." *Science Advances*, 5(12): eaax3370.

Nicholas, Tom. 2011. "Independent invention during the rise of the corporate economy in Britain and Japan 1." *The Economic History Review*, 64(3): 995–1023.

Riaño, Juan Felipe, and Felipe Valencia Caicedo. 2024. "Collateral damage: The legacy of the secret war in Laos." *The Economic Journal*, ueae004.

Romer, Paul M. 1990. "Endogenous technological change." *Journal of political Economy*, 98(5, Part 2): S71–S102.

Schürer, Kevin, and Edward Higgs. 2014. "Integrated Census Microdata (I-CeM), 1851-1911 [data collection]." *UK Data Service. SN*, 7481.

Shrosbree, Colin. 1988. *Public Schools and Private Education: The Clarendon Commission, 1861–64, and the Public Schools Acts.* Manchester:Manchester University Press.

Silva, JMC Santos, and Silvana Tenreyro. 2006. "The log of gravity." *The Review of Economics and statistics*, 88(4): 641–658.

Smil, Vaclav. 2005. *Creating the twentieth century: Technical innovations of 1867-1914 and their lasting impact.* Oxford University Press.

University of Portsmouth. 2011. "Great Britain Historical GIS Project." [Online; accessed 1-Feb-2022].

Voth, Hans-Joachim. 2021. "Persistence–myth and mystery." In *The handbook of historical economics*. 243–267. Elsevier.

Voth, Hans-Joachim, Bruno Caprettini, and Alex Trew. 2023. "Fighting for growth: Labor scarcity and technological progress during the British industrial revolution." *CEPR Discussion Papers*, , (17881).

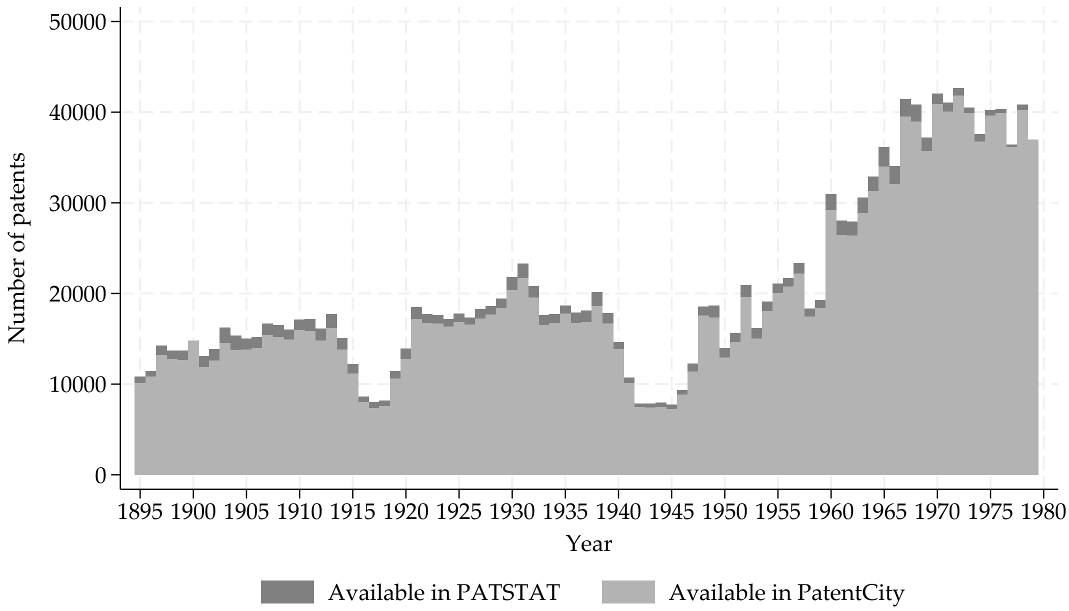
Waldinger, Fabian. 2016. "Bombs, brains, and science: The role of human and physical capital for the creation of scientific knowledge." *Review of Economics and Statistics*, 98(5): 811–831.

Winter, Jay M. 1977. “Britain’s ’Lost Generation’of the First World War.” *Population Studies*, 31(3): 449–466.

Online Appendix

A. Additional Figures and Tables

FIGURE A.1
PATENTCITY AND PATSTAT DATA



Notes: This figure reports, for each year in our sample, the number of patents registered at the Great Britain patent office and reported in PatentCity (light gray) and in PATSTAT (dark gray), respectively.

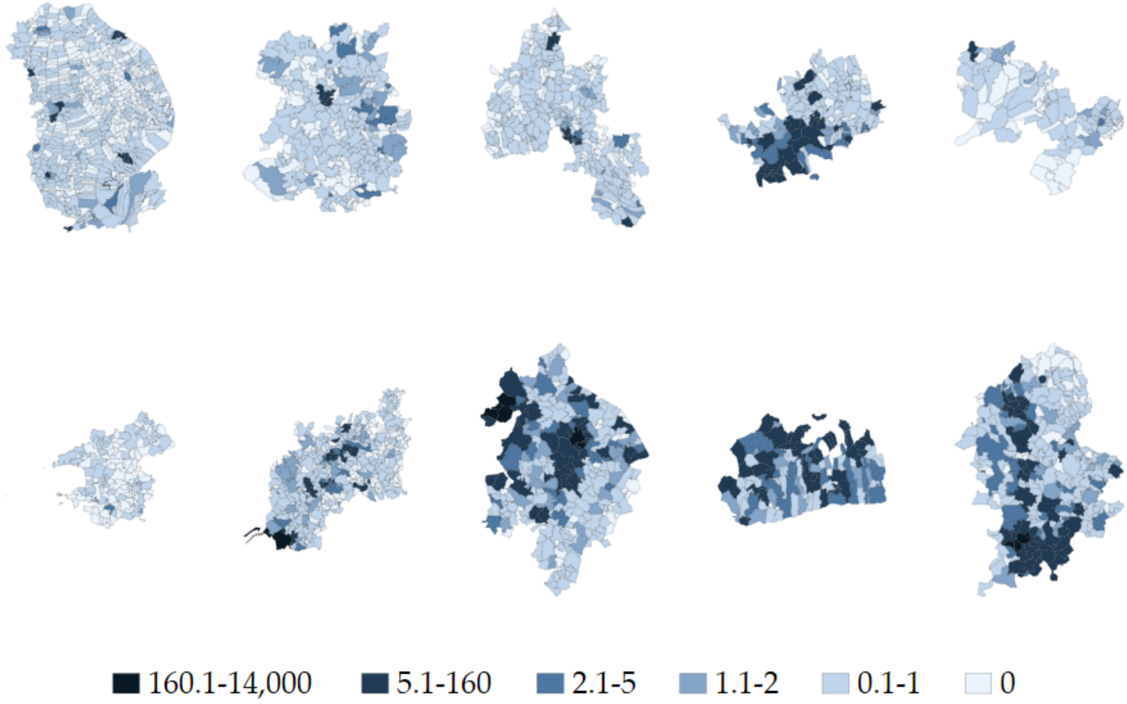
Table A.1

Descriptive statistics

	Mean	Std. dev.	Min	Max
A. Parish-level data				
Population 1911	3,332.30	69218.13	3	7004730
Area (sq. km)	14.25	19.65	0	1314
Population density 1911	252.87	839.97	1	23507
Share of households with servants 1911	0.14	0.09	0	1
Male ratio 1911	0.49	0.07	0	1
N. Mobilised in WWI	284.92	6532.09	0	640759
N. Mobilised/Population in WWI (%)	5.53	4.89	0	58
N. Dead in WWI	54.06	1144.76	0	113271
N. Dead/Population WWI (%)	1.05	1.58	0	67
At least one WWI dead (dummy)	0.88	0.32	0	1
At least one WWI dead had elite educ. (dummy)	0.07	0.25	0	1
At least one WWI dead was engineer (dummy)	0.11	0.31	0	1
Ever registered a patent (1895-1979)	0.66	0.47	0	1
Observations	10807			
B. Parish-level panel data (1895-1979)				
Any patent registered (over 5-year period)	0.24	0.43	0	1
N. patents registered (over 5-year period)	4.13	206.47	0	29501
N. patents reg. by existing inv. (over 5-year period)	1.86	97.39	0	17123
N. patents reg. by first-time inv. (over 5-year period)	1.12	45.62	0	7863
Observations	183719			
C. Inventor-level data				
Co-authored first patent (dummy)	0.33	0.47	0	1
Highly cited first patent (dummy)	0.12	0.33	0	1
Breakthrough first patent (dummy)	0.18	0.38	0	1
Engineer first patent (dummy)	0.38	0.49	0	1
Observations	54642			

Notes: Panel A: Descriptive statistics at the (grouped) parish level. Data are either from the 1911 Census or from the WWI war records, as indicated. Panel B: Descriptive statistics from the parish-level panel, where each observation is a parish observed over a 5-year period. The panel covers all 5-year periods in 1895-1979. Any patent is an indicator for having at least one patent being registered by patentee residing in the parish in a given 5-year period. Similarly, we define n. patents registered as the average number of patents registered in the parish over a given 5-year period, n. patents by existing inventors (i.e., those who registered at least one patent in the pre-War period), n. patents by first-time inventors (those who appear in the data for the first time). Panel C: Descriptive statistics from the inventor-level dataset. Co-authored, highly cited, breakthrough, and engineer first patent are indicators for the first patent being co-authored, highly cited, or breakthrough. Similarly, engineer first patent is an indicator for the patentee registering as an engineer at the time of the first patent.

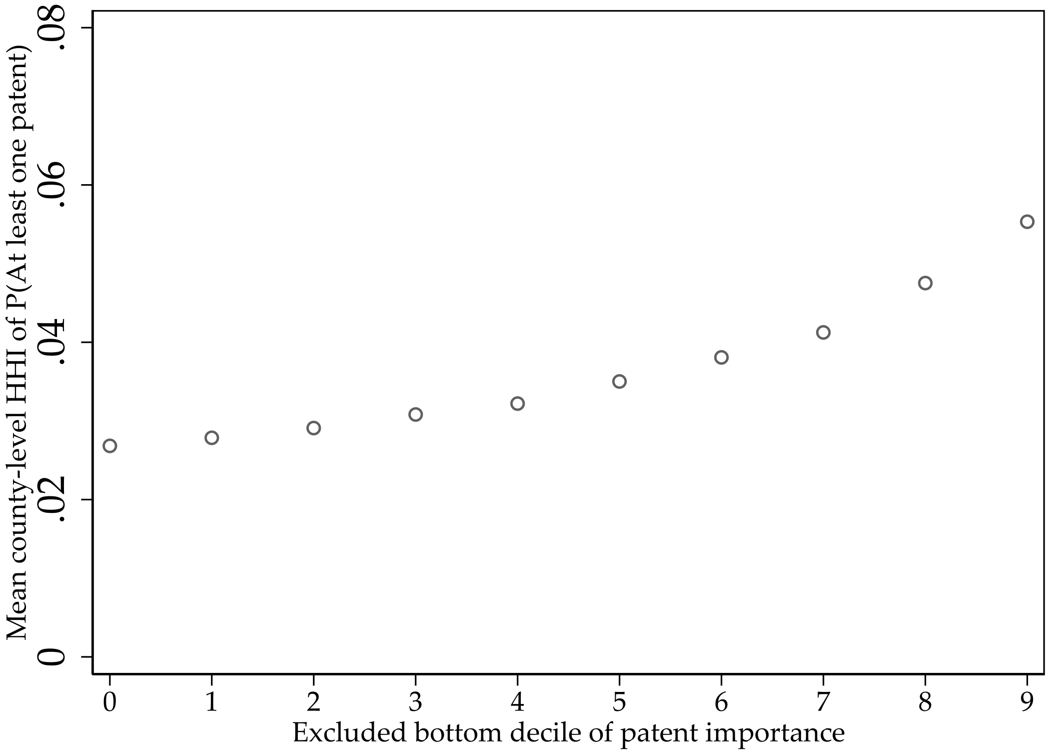
Figure A.2
Patents registered across ten randomly drawn counties in England and Wales



Notes: Historical (grouped) parishes for ten randomly selected counties in England and Wales. The figure shows the number of patents registered over a 5-year period, averaged over all periods in 1895-1979.

Figure A.3

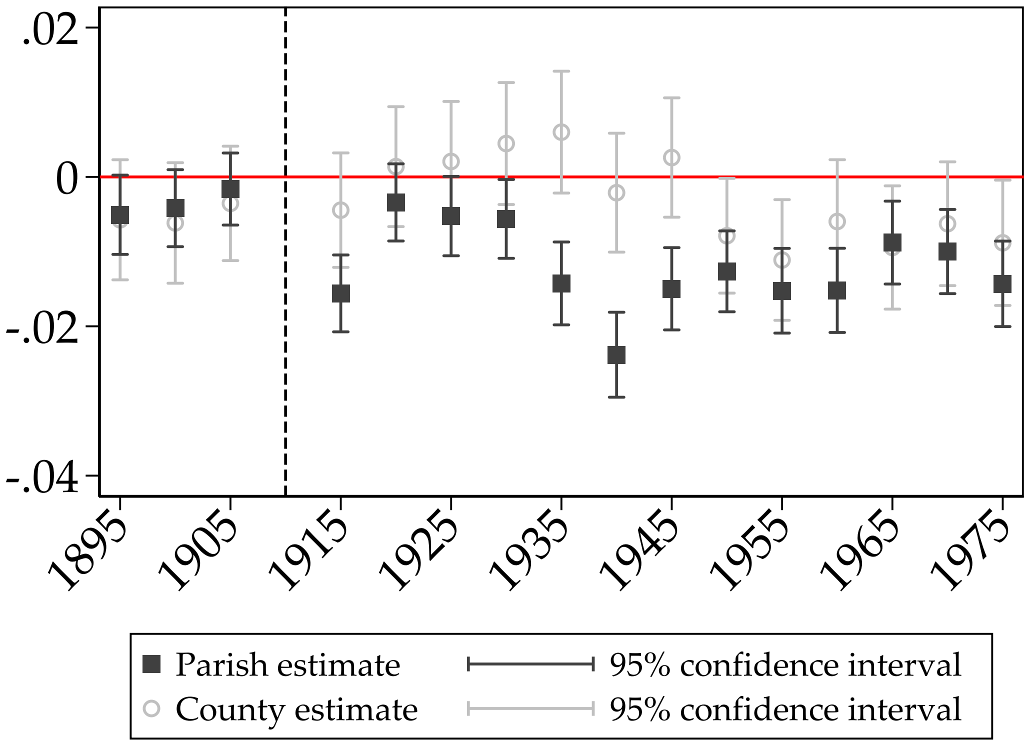
Spatial concentration of patenting activity across parishes of the same county



Notes: This figure reports the average Herfindahl–Hirschman Index (HHI) of the spatial concentration of patenting activity across parishes within counties, averaged over all counties in the sample. We first compute, for each parish, the historical average of our main dependent variable – an indicator equal to one if at least one patent is registered by an inventor residing in parish i during 5-year period t – averaged across all periods in the sample. The HHI is then calculated using this historical average distribution of patenting activity across parishes within each county. The analysis focuses on relatively more important patents, as measured by the patent importance index proposed by Kelly et al. (2021). Specifically, patents are ranked within publication cohorts, and we progressively restrict the sample to patents belonging to higher deciles of the importance distribution. Each coefficient in the figure corresponds to a different HHI calculation based on these restrictions: the left-most coefficient uses all patents, while the right-most coefficient measures the spatial concentration of the top 10 percent most innovative patents published during 5-year period t.

Figure A.4

County and Parish level deaths – Event-study results



Notes: OLS estimates of the effect of WWI deaths at parish and county scales on the probability that at least one patent is registered by a resident of parish i during 5-year period t. Estimates from equation 2, including time and parish fixed effects. 95% confidence intervals constructed using standard errors clustered at the parish level.

TABLE A.2

WWI DEATHS AND PATENTING – OTHER SPECIFICATIONS

	(1) Any patent	(2) N. patents	(3) Patents p.c.
A. Log deaths			
$log(d^{WWI})\times$ Post WWI	-0.01***	-0.10***	-0.02***
	(0.001)	(0.010)	(0.003)
Observations	162208	121398	162208
B. Death rate p.c.			
$d^{WW1}p.c.\times$ Post WWI	-0.46***	-10.84***	-0.58**
	(0.153)	(2.111)	(0.273)
Observations	183719	130511	183264
C. Deaths above median			
$I(d^{WW1} > med)\times$ Post WWI	-0.02***	-0.41***	-0.05***
	(0.004)	(0.075)	(0.008)
Observations	183719	130511	183264

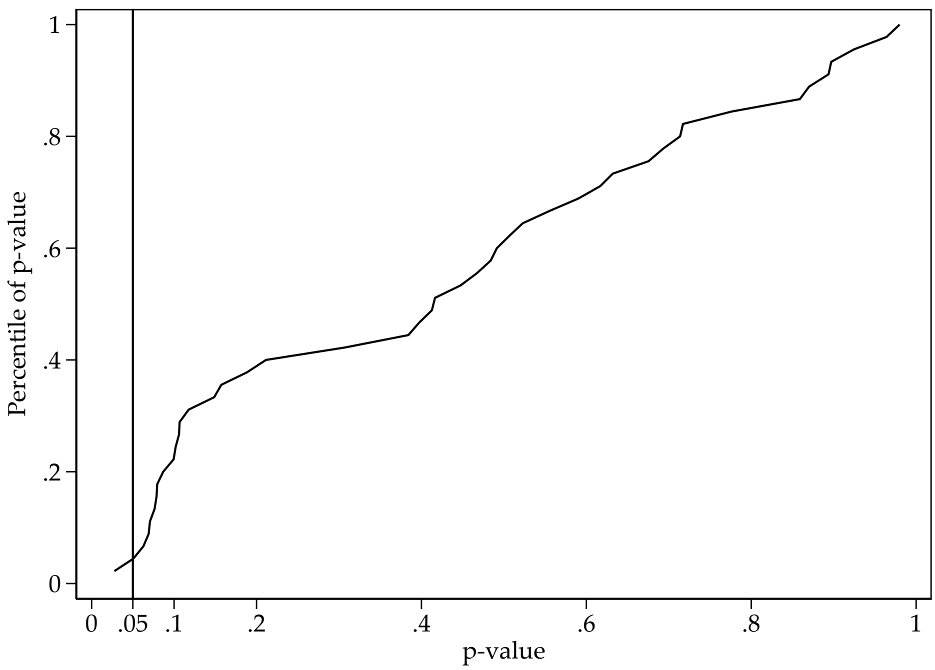
Notes: Each cell in this table represents a separate regression. Estimates in columns 1 and 3 are obtained by OLS. Estimates in column 2 are obtained by poisson pseudo maximum likelihood estimation using the command *ppmlhdfe* in Stata (Correia, Guimarães and Zylkin, 2020). All regressions control for parish fixed effects and county-time fixed effects. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively. The number of observations is lower throughout column 2 because *ppmlhdfe* drops singletons and observations separated by fixed effects (Correia, 2015). Patents p.c. are patents per capita * 1000. Because this variable is highly skewed, with a few large outliers, we winsorise it – that is, replace values above the 95% percentile with the 95% percentile value. Trimming them instead yields very similar results.

Table A.3
WWI deaths and patenting – Inventor-level estimates – excluding WWI deaths

	Not killed in WWI	Not killed in WWI	Not killed in WWI	Active after WWI
$Log(d^{WWI})\times$ Post WWI	-0.002***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Match on	Surn.+3 dgts.	Surn.+2 dgts	Surn.+1 dgt	–
Mean dep.var.	0.13	0.14	0.15	0.23
R2	0.58	0.57	0.55	0.43
Observations	466480	423351	351917	244664

Notes: OLS estimates of the effect of WWI deaths on an indicator equal to one if inventor j from parish i registers a patent during 5-year period t. The sample is restricted to patentees who registered at least one patent before WWI. Columns 1–3 present results from equation 3, restricting the sample to patentees who do not appear in the CWGC dataset of WWI fatalities under different matching criteria: column 1 requires a match on surname and the first 3 name initials; column 2, surname and the first 2 initials; column 3, the first initial to match exactly. As the matching criteria are relaxed, more patentees are excluded. Column 4 restricts the sample to active patentees (those who registered at least one patent in the post-WWI period). Inventor, time, and year of first patent by time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

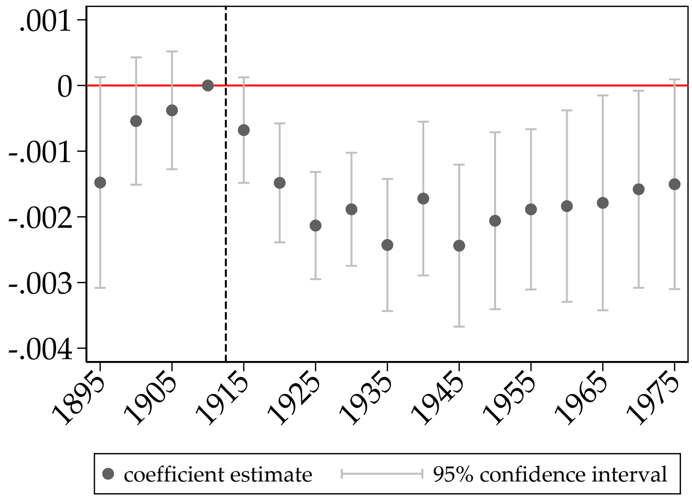
Figure A.5
Parallel trends - additional evidence

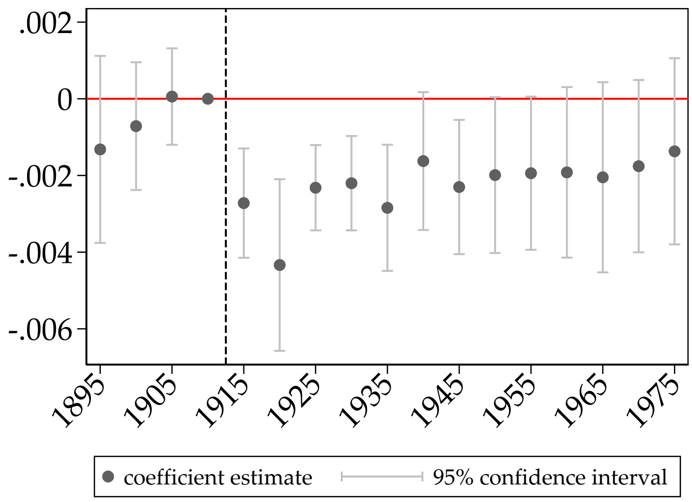


Notes: The figure plots the empirical CDF of p-values from pairwise "strong" parallel-trends tests across deciles of $\log$(number dead). For each admissible pair of deciles (i, z) (with $i < z$ and $i+z \leq 10$), we estimate an event-study regression of the outcome on $\mathbf{1}\{\text{decile} = z\} \times \mathbf{1}\{t\}$ indicators (omitted year 1910), with parish and county-time fixed effects and standard errors robust to clustering at the parish level. We test $H_0 : \beta_{1895} + \beta_{1900} + \beta_{1905} = 0$ and plot the distribution of the resulting p-values; the vertical line indicates $p = 0.05$.

Figure A.6

Inventor-level event study

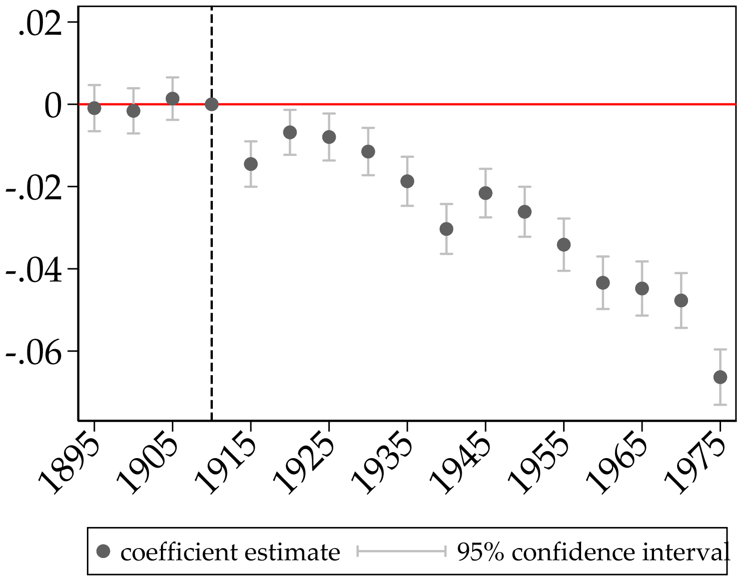




Notes: Inventor-level OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered during 5-year period t. Estimates from equation 3. Individual, time, and year of first patent by time fixed effects are included in all specifications. 95% confidence intervals constructed using standard errors clustered at the parish level. Upper panel: all inventors. Lower panel: inventors that are not killed in WWI (defined as those we could not match to the CWGC dataset of WWI fatalities using full name and surname).

Figure A.7

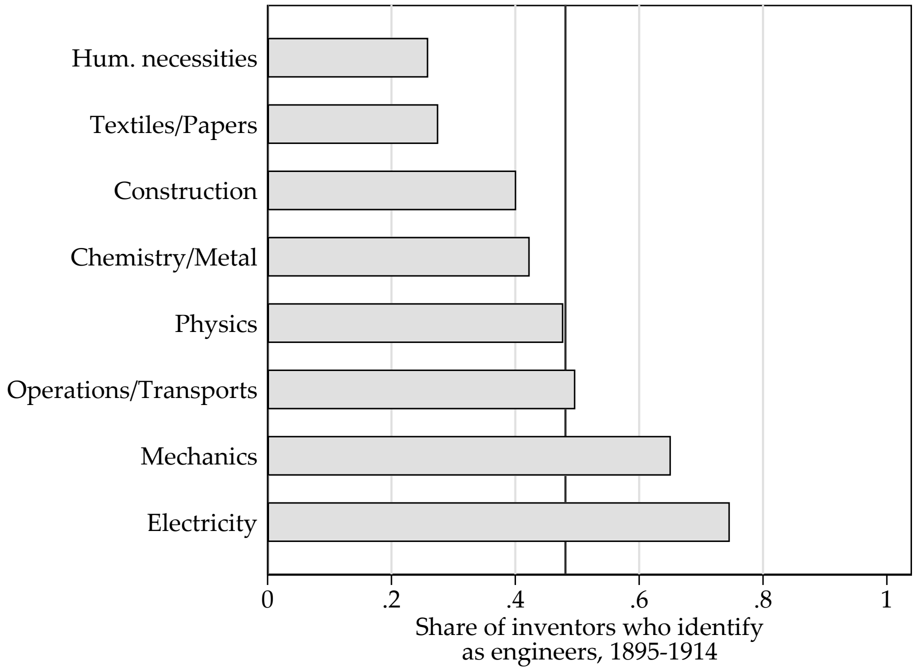
WWI deaths and patenting – Excluding company patents



Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by an individual from the parish during a 5-year period (patents registered by companies are excluded). The mean of this outcome is 0.192. Estimates from equation 2, including time and parish fixed effects and county-time fixed effects. 95% confidence intervals constructed using standard errors clustered at the parish level.

Figure A.8

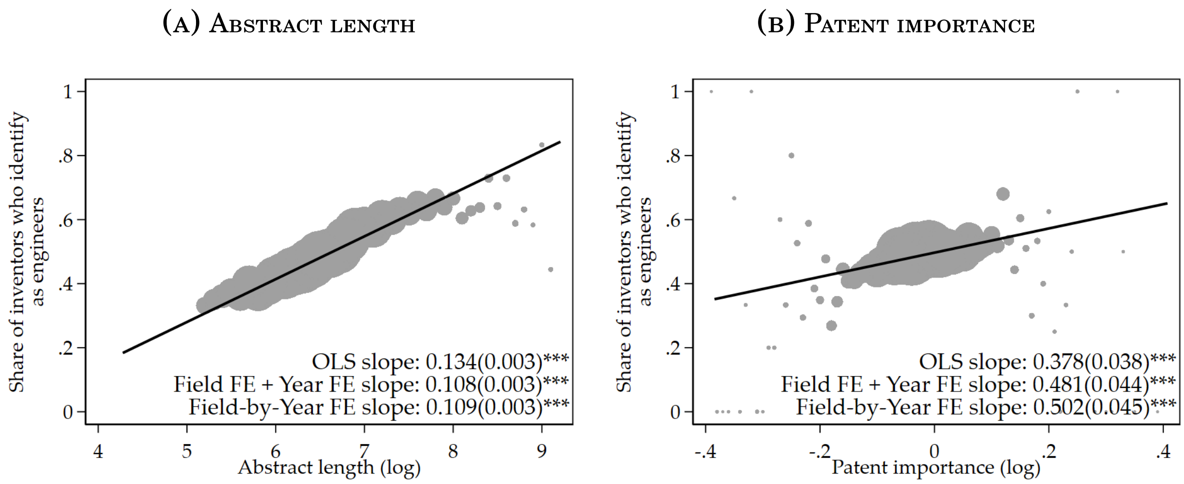
Predominance of engineers among inventors by field



Notes: This figure reports the incidence of engineers among inventors across different fields of specialisation in the pre-WWI period. For each major technological sector, we compute the share of inventors who report their occupation as being "engineer". The figure illustrates the extent to which engineering expertise was predominant among British inventors of the period across various technological domains.

Figure A.9

Engineers, patent complexity and importance



Notes: Scatter plots of the abstract length (panel A) and patent importance (panel B – measured following Kelly et al. 2021) against the share of authors who were engineers. Estimated OLS slopes (with or without year and field $\times$ year fixed effects) are also reported, with robust s.e. in parentheses.

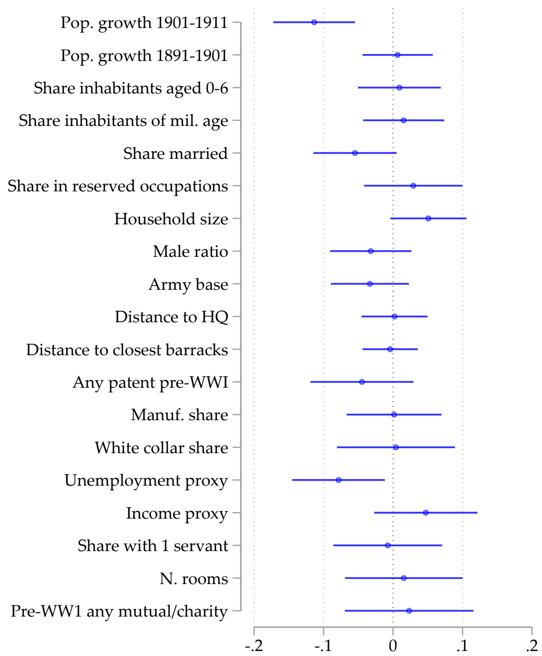
Table A.4

Inventor-level estimates – sectors, movers, and co-authorship

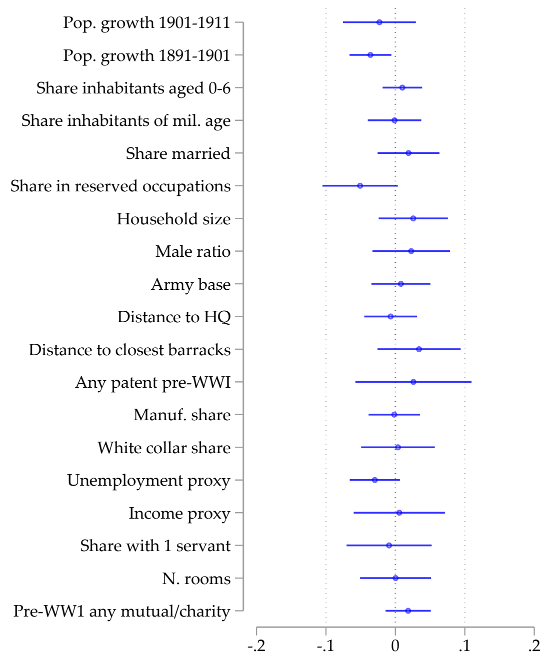
		Same vs. new field		Same vs. new parish		Authorship	
	Any patent	Any patent same field	Any patent new field	Any patent same parish	Any patent new parish	Any single-auth. patent	Any co-authored patent
$Log(d^{WWI})\times$ Post	-0.001***	-0.001***	-0.002***	-0.003***	0.004***	-0.002**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Mean dep. var.	0.117	0.084	0.049	0.083	0.043	0.083	0.048
R2	0.60	0.74	0.23	0.75	0.23	0.45	0.27
Observations	924477	924477	924477	924477	924477	924477	924477

Notes: Inventor-level OLS estimates of the effect of WWI deaths on patenting from eq. 3. Column 1 reports the baseline estimate (analogous to column 1 of Table 3) for comparison. In column 2 (resp., 3), the dependent variable is an indicator taking the value 1 if inventor j registers at least one patent during 5-year period t within the same sector (resp., within a different sector) as the one the inventor first patented in. In column 4 (resp., 5), the dependent variable is an indicator taking the value 1 if inventor j registers at least one patent during 5-year period t while residing in the same parish (resp., while residing in a different parish) as the one in which they were residing upon registering their first patent. Finally, in column 6 (resp., 7), the dependent variable is an indicator taking the value 1 if inventor j registers at least one single-authored (resp., co-authored) patent during 5-year period t. Notice that inventor j in the same 5-year period t can produce patents within different fields, residing in multiple locations, and with different co-authorship teams. Individual, time, and year of first patent by time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

Figure A.10
Instrumental Variable Balancing Checks



(a) Parish-level



(b) Battalion-level

Notes: Panel A: OLS estimates of individual regressions of the instrument z_i on different variables, with 95% confidence intervals. All outcomes are standardised to have mean zero and unit standard deviation. All specifications control for the log of 1911 population, the log of WWI mobilisation, and historic county fixed effects. Panel B: OLS estimates of individual regressions of the battalion-level death rate δ_b on different variables. All variables have been aggregated at the battalion level using the *ssaggregate* command in Stata 17 (residualised using log of population and log of mobilisation) following Borusyak, Hull and Jaravel (2022). Population growth rates are trimmed at the 1st and 99th percentile, while household size, the number of rooms per person, and area are trimmed at the 99th percentile. Share of inhabitants of military and share of married men are calculated restricting to individuals aged 14-35 at the time of the 1911 Census. Income proxy is the first principal component of white collar share, unemployment, share with 1 servant, and n. rooms per person. Standard errors are clustered at the historic county level (panel A) or regiment level (panel B).

Table A.5

WWI deaths and patenting – inventor-level IV estimates

	Any patent	Any patent	Any breakthr. patent	Any highly cited patent
$Log(d^{WWI})\times$ Post WWI	-0.001***	-0.002***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Sample	Full	Not killed in WWI	Full	Full
Mean dep.var.	0.12	0.13	0.03	0.02
R2	0.00	0.00	0.00	0.00
Observations	897770	476068	897770	897770

Notes: IV estimates of the effect of WWI deaths (instrumented using predicted deaths as described in Section 8.2) on an indicator equal to one if inventor j from parish i registers a patent during 5-year period t. The sample is restricted to patentees who registered at least one patent in the pre-War period. Column 1 presents results from equation 3 for the full sample, whereas in column 2 the sample is restricted to individuals who do not appear in the CWGC dataset of WWI fatalities, matched using the initial of the first name and surname. In columns 3 and 4, the outcomes are, respectively, an indicator for the inventor registering a "breakthrough" patent or a highly-cited patent (see text for details). Individual, time, and year of first patent by time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

Table A.6

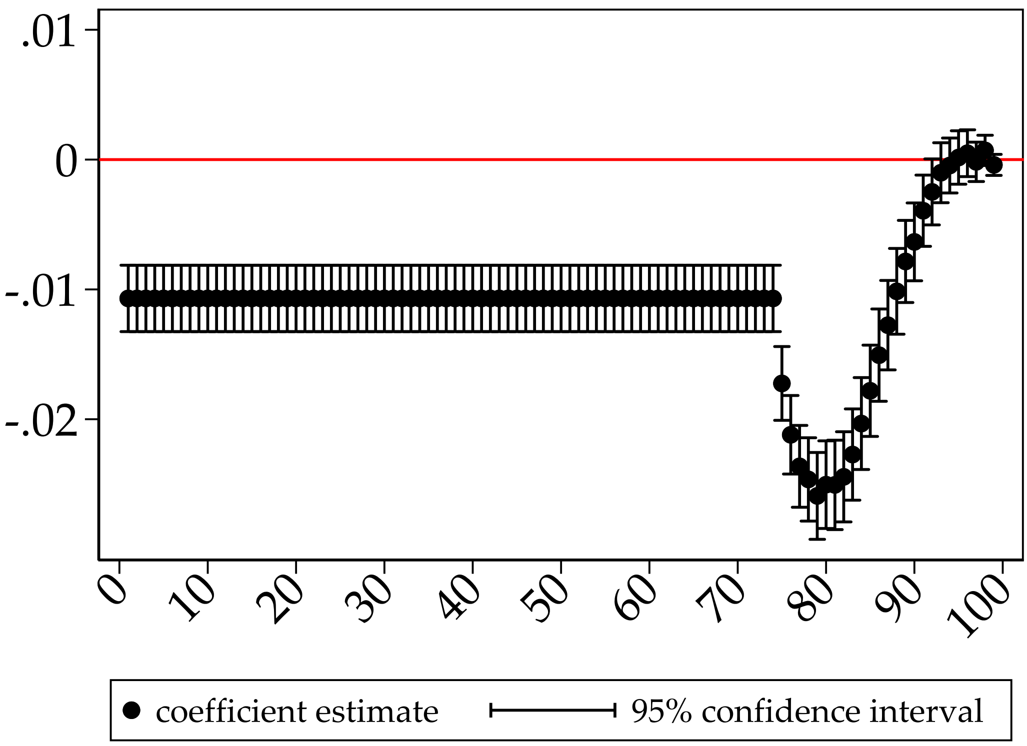
WWI deaths and patenting – Parish-level estimates including parishes without any WWI deaths

	(1) Any patent	(2) Any patent	(3) Any patent	(4) Any patent
$Log(1+d^{WWI})\times$ Post WWI	-0.008***	-0.013***	-0.009***	-0.009***
	(0.001)	(0.001)	(0.002)	(0.002)
Mean dep.var.	0.24	0.24	0.24	0.24
R2	0.48	0.49	0.49	0.49
Observations	182333	182333	182333	182333
Parish FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
County $\times$ Time	N	Y	Y	Y
Controls $\times$ Linear trend	N	N	Y	Y
Population (log)	N	N	N	Y

Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by a resident of parish i during 5-year period t. The treatment variable in this regression is log(1+number of deaths). Column 2 adds county-time fixed effects. Column 3 adds controls interacted with a linear trend. Column 4 also controls for the log of population (interpolated). Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

FIGURE A.11

WWI DEATHS AND PATENTING – OLS ESTIMATES – PATENT RATE ABOVE EACH PERCENTILE OF DISTRIBUTION OF PATENT RATES



Notes: OLS estimates of the effect of WWI deaths on the probability that the number of patents registered by residents of parish i during 5-year period t, divided by the population of parish i at the 1911 Census is above each percentile of the distribution of patent rates in our sample. Each coefficient is a different regression, where the dependent variable refers to the percentile specified on the horizontal axis. Estimates are based on equation 1, including time and parish fixed effects, and county-time fixed effects. 95% confidence intervals constructed using standard errors clustered at the parish level.

TABLE A.7

WWI DEATHS AND PATENTING – INVENTOR-LEVEL ESTIMATES INCLUDING PARISHES WITHOUT ANY WWI DEATHS

	Any patent	Any patent	Any breakthr. patent	Any highly cited patent
$Log(1 + d^{WWI})\times$ Post WWI	-0.001***	-0.002***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Sample	Full	Not killed in WWI	Full	Full
Mean dep.var.	0.12	0.13	0.03	0.02
R2	0.00	0.00	0.00	0.00
Observations	928897	492133	928897	928897

Notes: OLS estimates of the effect of WWI deaths on an indicator equal to one if inventor j from parish i registers a patent during 5-year period t. The sample is restricted to patentees who registered at least one patent in the pre-War period. The treatment variable in this regression is log(1+number of deaths), which implies that inventors active in parishes without any WWI deaths are not dropped from the sample. Inventor and time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

Table A.8

WWI deaths and patenting – Parish-level estimates removing post-WWII period

	(1) Any patent	(2) Any patent	(3) Any patent	(4) Any patent
$Log(d^{WWI}) \times$ Post WWI	-0.006***	-0.008***	-0.006***	-0.006***
	(0.001)	(0.001)	(0.002)	(0.002)
Mean dep.var.	0.26	0.26	0.26	0.26
R2	0.52	0.53	0.53	0.53
Observations	85958	85958	85958	85958
Parish FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
County $\times$ Time	N	Y	Y	Y
Controls $\times$ Linear trend	N	N	Y	Y
Population (log)	N	N	N	Y

Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by a resident of parish i during 5-year period t. Sample period is restricted to 1895–1939. Column 2 adds county-time fixed effects. Column 3 adds controls interacted with a linear trend. Column 4 also controls for the log of population (interpolated). Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

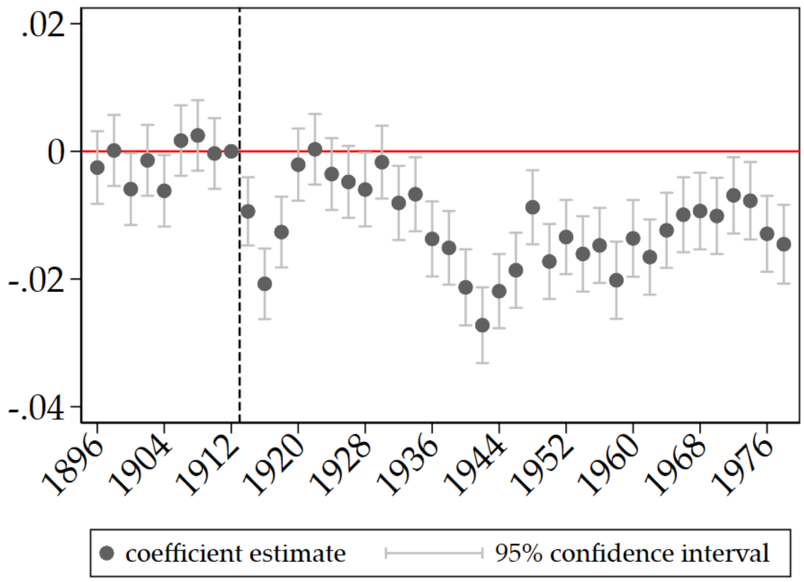
Table A.9

WWI deaths and patenting – Inventor-level estimates removing post-WWII period

	Any patent	Any patent	Any breakthr. patent	Any highly cited patent
$Log(d^{WWI}) \times$ Post WWI	-0.001***	-0.002***	-0.000**	-0.001**
	(0.000)	(0.001)	(0.000)	(0.000)
Sample	Full	Not killed in WWI	Full	Full
Mean dep.var.	0.21	0.23	0.04	0.04
R2	0.65	0.63	0.25	0.22
Observations	489456	259191	489456	489456

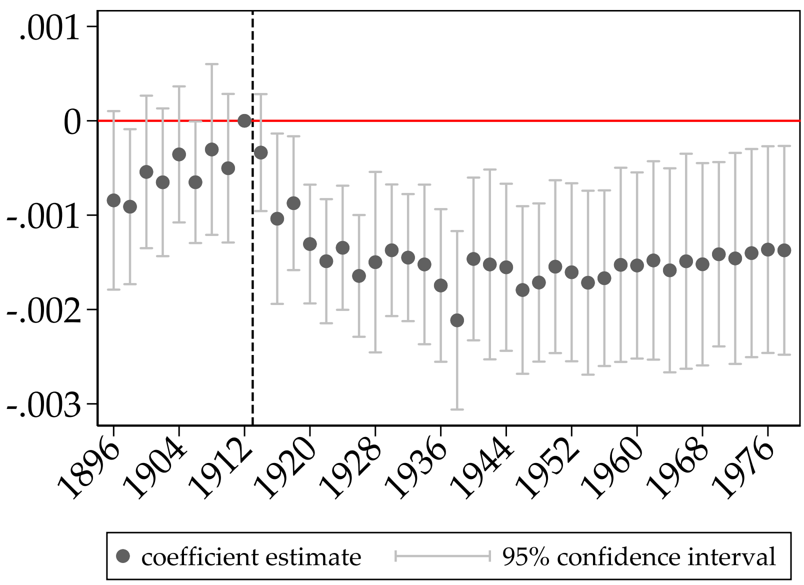
Notes: OLS estimates of the effect of WWI deaths on an indicator equal to one if inventor j from parish i registers a patent during 5-year period t. The sample is restricted to patentees who registered at least one patent in the pre-War period. Sample period is then restricted to 1895-1939. Individual, time, and year of first patent by time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

Figure A.12
WWI deaths and patenting – Bi-yearly data



Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by a parish resident during a 2-year period. Estimates are based on equation 2 and include time, parish, and county-time fixed effects. 95% confidence intervals constructed using standard errors clustered at the parish level.

Figure A.13
WWI deaths and patenting – Bi-yearly data - Inventor-level



Notes: Inventor-level OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by an inventor during a 2-year period. Estimates are based on equation 3. Individual, time, and year of first patent by time fixed effects are included. 95% confidence intervals constructed using standard errors clustered at the parish of first patent level.

B. Data

B.1. Military records

Data on British service personnel killed in WWI was obtained from the Commonwealth War Graves Commission (CWGC) (Commonwealth War Graves Commission, 2023), an intergovernmental organisation dedicated to marking, recording and maintaining the graves, memorials and memories of the men and women of the Commonwealth forces who died in both World Wars. Open data from this organisation contains information on names, time of death, rank, fighting unit, honours (e.g., gallantry medals), age and a string from which we can extract the location of origin of dead soldiers. Data on locations is augmented using information from Forces War Records (FWR), a military genealogy specialist website (Forces War Records, 2023). To implement this merge, we searched FWR for all soldiers in CWGC, and record place of birth and residence whenever available. After merging the datasets we perform a number of cleaning operations and restrict attention to soldiers from England and Wales who were killed between 1 August 1914 and 31 December 1918. We then geolocate soldiers to 1911 parish of origin by a combination of matching location strings to parish names and batch geolocation. When assigning soldiers to parishes, we prioritise the FWR data and residential location such that we only use birthplace when residence is missing or cannot be matched to a parish. The geolocation procedure assigns parishes to 73% (585,371) of the soldiers killed.

We next use ancillary data to identify members of the social elite in our CWGC-FWR data. Our primary method to do so relies on the Oxford University Roll of Service (Craig, 1920), and The War List of the University of Cambridge 1914-1918 (Carey, 1921). These documents name all faculty, enrolled students, and alumni of these pre-eminent universities that served during the war. We digitise these documents and match individuals listed to soldiers recorded in the processed CWGC-FWR data using surnames, initials, and rank. We successfully match 3305 individuals (23%) from the Oxford and 3260 (23%) from the Cambridge lists to our soldier-level fatalities dataset in this way.[37] As this definition of elites is highly selective, we then specify an alternative measure of social elites by using the information in the rank field of the CWGC-FWR data to construct a dummy that proxies for a killed soldier being an officer (either commissioned or non-commissioned).[38]

Data on 4,135,026 war records of soldiers mobilised during WWI is obtained from FamilySearch, a not-for-profit organisation which offers on-line access to large genealogical datasets (FamilySearch, 2023). FamilySearch draws its information from the British Army Service Records for 1914 to 1920. These records contain information on enrolled soldiers including names, place of residence, birthplace, age at the time of enlistment, year and unit in which the soldier was enlisted.[39] When cleaning and processing this information, we use as reference the Table of Organisation of each regiment as detailed in James (2012).

[37]The match rate reflects the high mortality in these groups, well above the mortality in the general population of servicemen.

[38]The ranks recorded in this field are highly heterogeneous for officers but not for privates, so we define officers as those whose rank does not coincide with a set of predetermined strings (e.g. "Private", "Rifleman").

[39]Digitised versions of these records can be consulted at `www.ancestry.co.uk`. The FamilySearch collection,

B.2. Further data sources

Individual-level information on the English and Welsh population before the Great War is obtained from the 1911 Census of population. The data we use originates from Schürer and Higgs (2014) and is distributed by IPUMS (Minnesota Population Center, 2019). We use this data both at the individual level and to construct aggregates at the parish level. From this source, we obtain information on several income proxies including the number of servants and the number of rooms per household. We obtain aggregate area-level information for Census years 1901-1931, as well as digital maps for parishes, districts and constituencies from "A Vision of Britain through Time" (VoB), an online library of spatial data created by the Geography Department at the University of Portsmouth (University of Portsmouth, 2011). The parish-level population counts in the VoB data come from the *Census Reports* that were published following each Census. There are known to be discrepancies between the population counts in this source and the more recently published micro-data, for example because not all records have survived or there is ambiguity in the true parish in the individual level records. Consequently, in general we use the counts from the *Census Reports* where available. Further, to minimise discrepancies we also implement the corrections to assigned parishes in the 1911 micro-data using the look-up tables published on the I-CeM website.[40]

We obtain membership lists of four engineering trade bodies in pre WWI years – the Institution for Mechanical Engineers, 1896; Iron and Steel Institute, 1888; Institution of Automobile Engineers, 1910; Institution of Electrical Engineers, 1887 – from the website Grace's Guide, a list of pre-WWI universities from Wikipedia, a list of rail stations in 1910 from a GIS first constructed by Jordi-Marti Henneberg, and a georeferenced list of all post offices in 1900 from the GB1900 Gazetteer. We geocode the first two sources using batch geolocation of address information. We then use these sources to construct a series of separate county-level ecosystem accessibility measures. Counties are large – there are 52 in England and Wales – and nest parishes. To construct each measure, we first compute the linear distance between each parish and the nearest ecosystem factor (e.g., a university), then obtain the average distance to that factor in each county by taking the weighted average of the parish level distances, using as weights 1911 parish populations. We then take the natural log of this weighted average distance.

B.3. Spatial Units of Analysis and Reconciliation

Our main analysis is based on a 1911 parish-level dataset covering England and Wales. We take 1911 as our reference year because it was the last Census conducted before the onset of the Great War in 1914. The civil parishes we use in our analysis are administrative units

which includes the extracted data, is called "United Kingdom, World War I Service Records, 1914-1920". The original sources of this information are the "Burnt documents" (record code WO 363) and the "Unburnt collection" (record code WO 364), which are kept in the National Archives at Kew in London. The Burnt Documents are roughly 2.5 million records on WWI soldiers which survived the fire resulting from an incendiary bomb hitting the War Office Record Store in 1940. The Unburnt Collection is made of soldier information obtained from pension claims. This collection was stored separately in 1942 and, therefore, did not suffer the fate of many of the Burnt Documents.

[40]Available at https://www.essex.ac.uk/-/media/newparids11.txt?la=en, accessed on May 5, 2023.

corresponding to the lowest level of local government in the United Kingdom. Civil parishes evolved from ecclesiastical parishes during the 19^{th} century, but by 1880 had no religious or ecclesiastical duties. In 1911, the territories of England and Wales were divided into 14,664 parishes, of which 13,404 in England and 1,260 in Wales. We drop all parishes that had zero population in 1911 – usually parcels of empty land in remote rural areas – and 10 additional parishes that have repeated names within the same county.

Parishes are nested within local government districts, of which there were 1,861 in 1911, and in turn within 52 counties. Parish boundaries change over time and in some cases variables are only available at other (higher) levels of aggregation. In order to aggregate or re-weight information to common boundaries we use a spatial matching procedure based on the assumption of uniform population distribution within parishes. Because our main spatial units (parishes) are relatively small (10 sq. km on average) and parish boundaries are often quite stable in the 30-year period we study, we expect the measurement error induced by making this assumption to be limited. We selectively group parishes using a semi-automated approach that groups those with similar names and that lie in close proximity. For example, we group High Abbotside with Low Abbotside and we also group Ledbury Urban with Ledbury Rural. In addition to this, we group a small number of suburban parishes with the corresponding city parish. For example, we group the suburb of Sculcoates with Hull. Reflecting that soldiers from the London area often report their location as being London, we also group together London parishes based on the historical conurbation definition available from Vision of Britain. Finally, we exclude 26 parishes, of which 25 are parcels of empty land and the other is unnamed. After restrictions and grouping, our final parish set encompasses 10,807 parishes.

Our data on 1911 parishes come from two different sources: the 1911 Census micro-data from I-CeM and the *Census Reports* from VoB. These sources use different parish codes and contain a slightly different set of parishes, so we create a mapping file and reconcile the data before conducting analysis.

B.4. Geolocation Procedure, Measurement Error, and Validation

Our empirical analysis requires geolocating soldiers based on information on their place of birth and residence from the sources described above. Here we provide details of this procedure.

The CWGC data on soldiers killed during WWI includes 796,601 records, from which, because our analysis focuses on England and Wales, we remove servicemen born in Scotland, Ireland, and abroad. We then extract information for residence or birthplace (or both) from either the birthplace and residence fields in FWR or the "additional information" string included in the CWGC source.

Geolocation of WWI fatalities proceeds by combining a) direct string matches with parish names based on data from FWR on historic county and location of birthplace/residence, b) direct string matching as above but based on the CWGC additional information field, and c) latitudes and longitudes obtained from a batch geolocating service to which we input the

FWR locations. For the batch geolocation process, we use a service provided by the company OpenCageGeo, which is based on OpenStreetMap and is available across platforms.

The data on parish of origin (birthplace or residence) of mobilised men in WWI – obtained from FamilySearch – has a slightly different structure and, therefore, we use a different procedure from the one used for CWGC/FWR data.[41] To match the FamilySearch records to an individual parish we combine: a) a direct string match with parish names for records that have both an historic county and a location, b) direct string matching with parish names for records that include no county information (only match to parishes with unique names), c) hand matching of a fraction of remaining records carried out by identifying locations via GoogleMaps. We are able to geolocate just over 2.6 million of these records.

When using this data together with the CWGC information on deaths to construct our instrument, we further exclude 2.05 million records for which the battalion is missing. Finally, we drop 57,795 entries that are duplicates in terms of all variables, and 72,131 records dated before 1905 or after 1920. Of the remaining sample, we then drop 735,768 individuals that could not be geolocated, as well as 28,578 from regiments with zero or negligible mortality, such as the Hussars. Finally, to ensure we have enough observations to construct the shares serving in each battalion, we drop 38,281 soldiers from battalions with less than 100 servicemen in the data.

Owing to measurement error in the geolocation process and the incompleteness of the FamilySearch records, some parishes exhibit mobilisation or WWI death counts that are unusually large relative to their population. To ensure that these potential outliers do not drive our results, we identify all parishes in which the per capita number of mobilised or of deaths lies above the 99^{th} percentile of the respective distribution. We then replace these figures with imputed values obtained by multiplying the 1911 parish population by the corresponding district-level death or mobilisation rate.[42] Results are robust to not applying this correction.

Various validation checks that support our geolocation approaches are provided in Carozzi, Pinchbeck and Repetto (2023).

B.5. Patents data

We draw on two principal sources of patent micro-data: PATSTAT and PatentCity.

PATSTAT is the European Patent Office's global patent database, providing standardised bibliographic data on over 100 million patents collected from more than 100 authorities worldwide. For our purposes, we focus on the subset of patents filed at the UK Intellectual Property Office between 1852 and 1979, encompassing information on inventors, applicants, technical classifications, titles and – where available – textual abstracts (European Patent Office, 2023).

PatentCity is a historical database of British patents covering the years 1895-1979 and

[41] For example, the batch geocoding procedure that we used and validated when using FWR data on locations for killed soldiers yields very poor results when used with the FamilySearch strings.

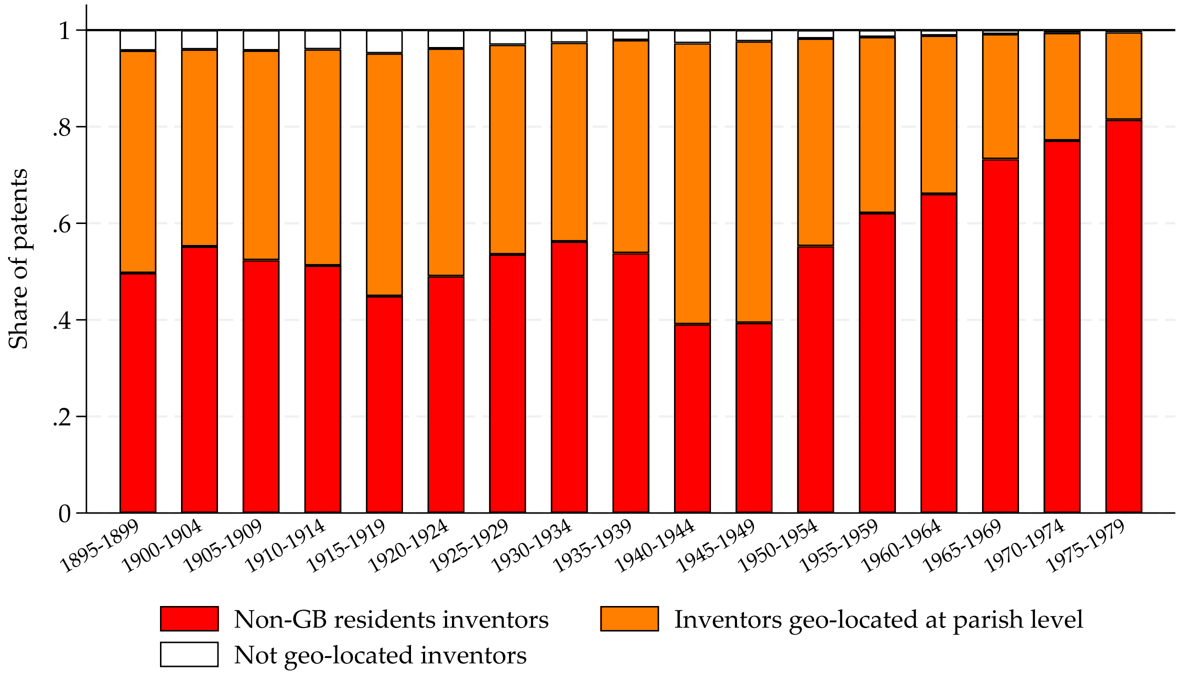
[42] We use the historical county when district and grouped parish boundary coincide.

specifically designed for spatial analysis of inventors' locations (Bergeaud and Verluise, 2024).[43] For our purposes, we focus on the 3.8 million records registered at the UK Intellectual Property Office. For about 1.2 million of these, at least one inventor can be matched to a location within the United Kingdom at the parish, district, or city level. Of these, we are able to assign a parish in England or Wales to 1.13 million of them. The combination of PATSTAT and PatentCity provides a comprehensive panel of British patents, allowing us to observe inventor identities and their location, as well as information on patent title, abstract, and sector.

B.5.1. Assignment of inventors to parishes

FIGURE B.1

INVENTORS GEO-LOCATED AT PARISH LEVEL



Notes: This figure reports, for each 5-year period in our sample, the share of patents available in PatentCity that are granted to inventors not residing in Great Britain (in red), the share of patents produced by inventors residing in Great Britain that we are able to geo-locate at the parish level (in orange), and the share of inventors residing in Great Britain for which the precision of the information available in PatentCity does not allow a geo-location to England and Wales parishes. Our estimation samples are based on inventor-patent pairs that fall in the orange category.

We assign each inventor–patent tuple to a parish using latitude and longitude coordinates provided by PatentCity, which reflect PatentCity's internal geocoding of inventor addresses. We retain observations with sufficiently precise geolocation quality – those geolocated to a city, district, street, house-number, or postal-code – and restrict to patentees located in Great

[43] In the original dataset there are patents dating as early as 1784 and as late as 2021. However, coverage before 1895 is incomplete, and the source used after 1979 is different. For these reasons, we restrict to the period 1895-1979.

Britain. These coordinates are then spatially matched to historical 1911 parish boundaries via polygon intersection using QGIS, discarding non-matching points.

More specifically, we discard 56 percent of patents because the patentee is located outside the United Kingdom and 2 percent because the georeferenced information provided in PatentCity is not sufficiently detailed or the coordinates thereby reported do not match with the map of 1911 parishes of England and Wales. Figure B.1 reports the share of patents registered by foreign inventors, the share of patents for which we are able to assign a parish to at least one of the inventors, and the share of patents that we are not able to geolocate, for each 5-year window in our sample.

B.5.2. Inventor-level link between PatentCity and PATSTAT

PATSTAT and PatentCity have a common patent identifier, so we can merge the two sources at the patent level. However, they do not share a common inventor identifier, hence a match between the two sources at the inventor level is not immediate. This match is instrumental for constructing the individual-level panel of inventors because the PATSTAT name and identifier allow us to trace inventors over time. When the patent is single-authored, we merge inventor-level records by the exact patent identifier across the two sources. When the patent is multi-authored, we link the two sources at the inventor level using a fuzzy matching procedure.

To this end, we follow a three-step approach. First, within each patent we generate the full cross-product of PatentCity and PATSTAT inventor entries, producing a long-format list of admissible inventor–inventor candidate pairs (the two sources share a unique patent identifier, so patents can be merged exactly). Second, we clean name strings and then use Stata's *matchit* command to compute fuzzy similarity scores between each PatentCity name and each candidate PATSTAT name within the same patent. For each PatentCity inventor, we select the PATSTAT inventor with the highest similarity score, requiring a minimum score of 0.25; lower-scoring pairs are left unmatched. Third, for matched records, we import the cleaned PATSTAT inventor name and assign it a unique inventor ID based on PATSTAT's numeric inventor identifier. When the same PATSTAT inventor is matched multiple times within a patent, we keep only the highest-scoring match to enforce a one-to-one mapping within each patent. The resulting dataset provides a harmonised inventor-level linkage that combines PatentCity's geolocation and contextual information with PATSTAT's standardised names and numeric inventor identifiers.

Based on these criteria, we match PatentCity and PATSTAT inventors for approximately 750,000 inventor–patent tuples in our main PatentCity data (more than 92 percent of all inventor–patent tuples). We find all PatentCity inventors in PATSTAT for 551,063 patents; for the remaining 49,153 patents, at least one inventor is unmatched.

To construct the inventor-level panel, we exclude patents identified by PATSTAT as registered by companies (non-individual inventors) and we keep only inventors that registered at least one patent in the pre-war period (1895-1914). These restrictions leave us with 54,642 unique individuals, which comprise our final sample.

C. Construction of the patent importance measure

This measure follows Kelly et al. (2021).

C.1. Construction of the Textual Representation

The construction of the patent-importance measure starts from the universe of patents in PATSTAT registered at the Great Britain patent office (with no further restrictions to maximise predictive power) and keeps only records with an abstract. This yields about 1.6 million patents. We transform each abstract into a TF–IDF vector using `scikit-learn`.[44] Each vector is then normalised to unit L^2 norm so that cosine similarity between any two patents equals the dot product of their normalised TF–IDF vectors.

C.2. Similarity Windows and Mean Cosine Similarity

For each patent k filed in year t, we compute its similarity to two sets of comparator patents. The backward set consists of all patents filed during the 5 years preceding t, that is $(t-5, \ldots, t-1)$, while the forward set consists of all patents filed during the ten years after t, that is $(t+1, \ldots, t+10)$. For each of these windows, we compute the mean cosine similarity between patent k and the patents in the corresponding set. Because the TF–IDF vectors are normalised, this step reduces to computing dot products between the vector for patent k and the vectors of all patents in the relevant window.

To make this computation feasible at scale, we follow the centroid-based approach introduced by Kelly et al. (2021). For each year, we construct the sum of the normalised vectors of all patents falling in the backward window and, separately, the sum of the normalised vectors of all patents falling in the forward window. The mean similarity of patent k to a window is then obtained by taking the dot product of the vector for k with the corresponding sum vector and dividing by the number of patents in the window.

C.3. Importance Measure

Following the definition in Kelly et al. (2021), we define the technological importance of patent k as the ratio between its similarity to future patents and its similarity to prior patents,

$$\text{Importance}_k = \frac{\text{ForwardSimilarity}_k}{\text{BackwardSimilarity}_k}$$

This measure is high when the textual content of patent k is more similar to language that subsequently appears in later innovations than to the terminology used in earlier ones, indicating that the ideas embodied in k anticipate future developments in the technological frontier.

[44]The vocabulary is restricted to the 1,000 most frequent terms in the corpus and English stopwords are removed.

D. CPC prediction algorithms

This section provides details on the Natural Language Processing (NLP) models used to predict the one-digit Cooperative Patent Classification (CPC) code from patent titles.

We use the patent's title as the input for our NLP algorithms. For each exercise, we restrict attention to patents for which a unique one-digit CPC code is observed in PATSTAT. We divide these patents randomly into a training sample (80%) and a held-out testing sample (20%), using a fixed random seed to guarantee replicability. The training sample is used to estimate the relationship between textual features extracted from the patent title and the CPC field, while the testing sample provides an unbiased evaluation of out-of-sample predictive accuracy.[45] As the name suggests, the training sample is used to "teach" the algorithm about the relationship between the text of each patent's title and the one-digit CPC code.

We estimate four supervised classifiers commonly used in text classification: *Naive Bayes*, *Decision Tree*, *Random Forest*, and *Support Vector Machine* (SVM). All models are trained on the same TF–IDF representation of the cleaned patent titles, where text is lowercased, punctuation is removed, English stopwords are dropped, and terms are weighted using term-frequency inverse-document-frequency with a maximum vocabulary size of 5,000 features.[46] Each model is then used to predict the CPC field for the patents in the testing sample, and its performance is evaluated by comparing predicted and observed CPC codes.

Lastly, we utilise the same algorithm to predict the one-digit CPC code also for patents published before 1910 (for which a CPC code is not available in our sources) and for patents that used to report multiple CPC codes (summing up to ≈ 1.3 million patents). Among the algorithms, the SVM achieves the highest predictive accuracy ($\approx$ 85%), and therefore we use the SVM predictions as our preferred measure of the most likely field.

For one-digit CPC codes, we present and discuss three sets of diagnostics based on the testing sample for each of the four classifiers. First, in Figure D.1, we compare the marginal distribution of predicted CPC fields with the true empirical distribution of fields in the test sample. For each classifier, Figure D.1 plots overlaid histograms showing that the predicted and observed distributions are broadly similar, although the degree of alignment varies across algorithms. In all panels, red bars reflect the distribution of one-digit CPC codes as they are observed in the training sample while blue bars report the distribution of predicted one-digit CPC codes in the test sample. All classifiers perform well according to this metric, perhaps with the exception of Naive Bayes, which tends to overestimate the share of patents in sector B (Performing operations and transports) and to underestimate most other fields.

Second, we examine the full pattern of misclassifications by computing, for each observed

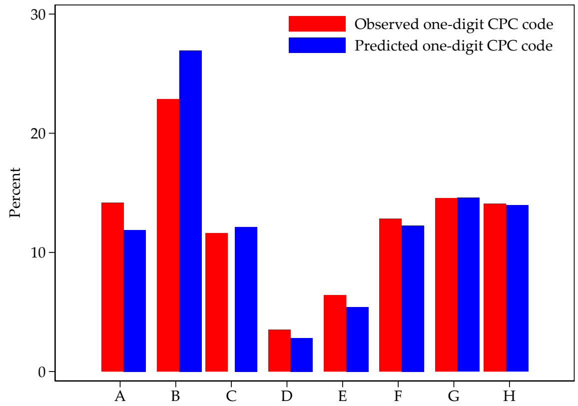
[45] Notice that this exercise is performed on all patents registered at the Great Britain patent office and available on PATSTAT, with no restriction about publication year and/or geo-reference to parishes in PatentCity. We follow this route in order to maximise the accuracy of the prediction algorithm. The total number of patents used is 3,420,207.

[46] These transformations are implemented in Python using the `nltk` and `scikit-learn` libraries.

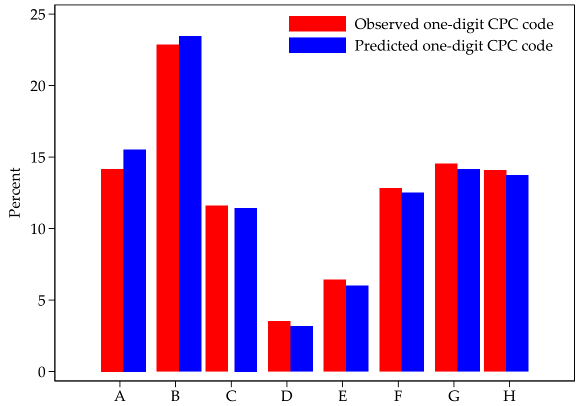
Figure D.1

Comparison of observed and predicted CPC distributions across classifiers

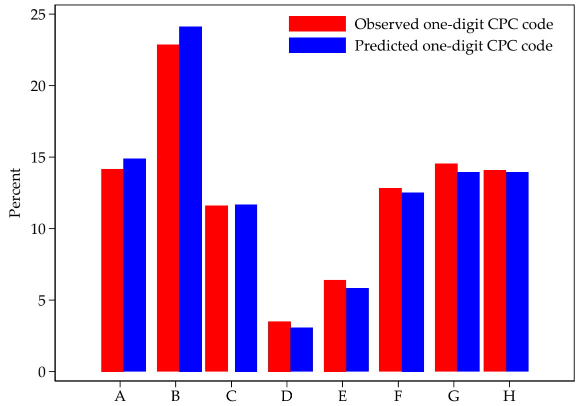
(a) Naive Bayes



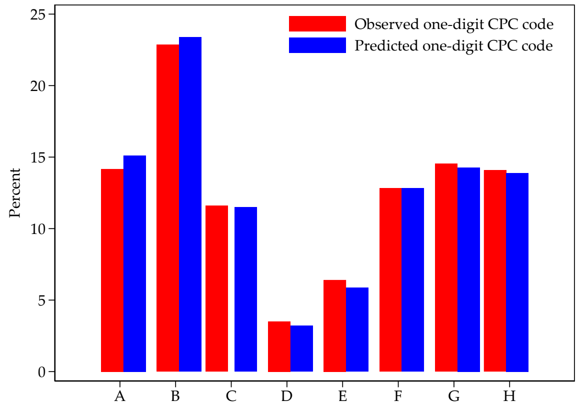
(b) Decision Tree



(c) Random Forest



(d) Support Vector Machine



Notes: Each panel compares the empirical distribution of observed one-digit CPC codes (red) with the distribution predicted by the corresponding machine-learning classifier (blue), computed on the held-out test sample. A closer alignment indicates better predictive performance.

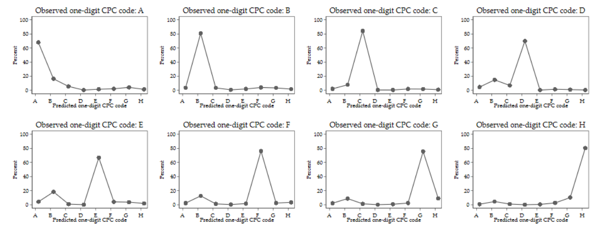
field, the share of patents that are predicted into each possible field. For every algorithm, Figure D.2 reports a panel of eight plots (one per CPC field), which can be interpreted as a visual confusion matrix. In a well-performing classifier, each plot displays a dominant peak at the correct CPC category. The evidence presented in Figure D.2 is reassuring about the validity of the prediction exercise, as for all CPC codes and classifiers we consistently see a spike aligned with the observed field. Similarly to the results presented in Figure D.1, the Naive Bayes classifier visibly underperforms compared to the other alternatives.

Third, Tables D.1–D.4 report the detailed classification performance for each of the four supervised models considered in the validation exercise: Naive Bayes, Decision Tree, Random Forest, and Support Vector Machine (SVM). For every one-digit CPC field (A–H), we compute two standard metrics used in multi-class text classification: *recall* and *precision*. Recall is defined as $\Pr(k \in \hat{F} \mid k \in F)$ and measures the share of patents k truly belonging to field F that the algorithm correctly assigns to that field. Precision is defined as $\Pr(k \in F \mid k \in \hat{F})$ and measures the share of patents predicted to be in field F that

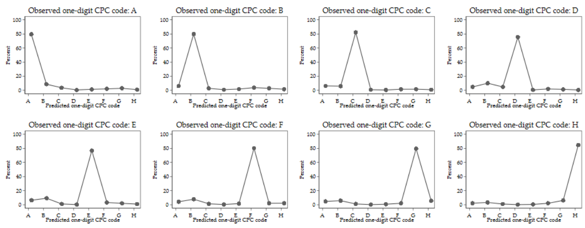
Figure D.2

Conditional distribution of predicted CPC codes by classifier

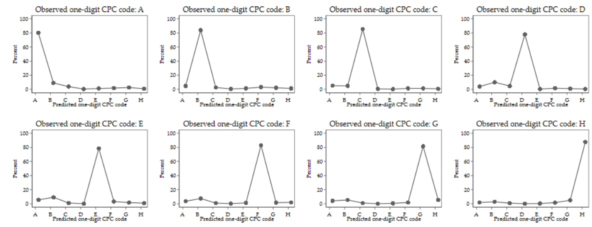
(a) Naive Bayes



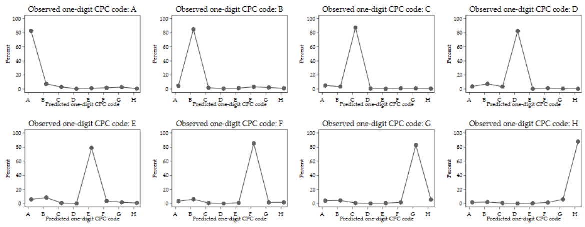
(b) Decision Tree



(c) Random Forest



(d) Support Vector Machine



Notes: Each panel summarises, for a given classifier, the conditional distribution of predicted one-digit CPC codes given the observed CPC field. Within each subfigure, the eight small plots correspond to patents whose true CPC is $A, \ldots, H$; on the horizontal axis we report the predicted CPC code, and on the vertical axis the share (in percent) of patents in that true field that are assigned to each predicted field. Thus, for a patent with true field F, the plotted values trace out the distribution of predicted fields $\hat{F}$. For each classifier, these shares are computed as the number of patents with true field F and predicted field $\hat{F}$, divided by the total number of patents with true field F, multiplied by 100. A well-performing classifier displays a dominant peak and relatively low mass on other entries.

are in fact observed in that field. These metrics summarise, respectively, how often each model misses patents belonging to a field (low recall) and how often it over-assigns patents to a field (low precision). Each table also reports a final row showing the simple average of recall and precision across fields, which describes the model's balanced performance across the CPC spectrum. Lastly, the final row reports the overall *multi-class accuracy*, computed as the share of patents in the held-out test sample for which the predicted one-digit CPC code exactly matches the observed CPC code. This metric serves as our primary criterion for algorithm selection. The Support Vector Machine achieves the highest multi-class accuracy among all classifiers.

TABLE D.1

FIELD-LEVEL RECALL AND PRECISION: NAIVE BAYES

CPC field	Recall	Precision
A	0.681	0.812
B	0.809	0.686
C	0.842	0.805
D	0.700	0.872
E	0.668	0.791
F	0.762	0.797
G	0.756	0.753
H	0.805	0.813
Average	0.753	0.791
Accuracy (multi-class)	0.767	

Notes: Rows A–H report recall and precision for each one-digit CPC field. Recall is defined as $\Pr(k \in \hat{F} \mid k \in F)$ and precision as $\Pr(k \in F \mid k \in \hat{F})$. The "Average" row reports averages of recall and precision across fields. The final row reports the overall *multi-class accuracy*, defined as the share of patents in the test sample for which the predicted CPC code equals the observed one.

TABLE D.2

FIELD-LEVEL RECALL AND PRECISION: DECISION TREE

CPC field	Recall	Precision
A	0.796	0.727
B	0.799	0.779
C	0.822	0.834
D	0.756	0.839
E	0.768	0.819
F	0.803	0.822
G	0.796	0.818
H	0.847	0.869
Average	0.798	0.813
Accuracy (multi-class)	0.805	

Notes: Rows A–H report recall and precision for each one-digit CPC field. Recall is defined as $\Pr(k \in \hat{F} \mid k \in F)$ and precision as $\Pr(k \in F \mid k \in \hat{F})$. The "Average" row reports averages of recall and precision across fields. The final row reports the overall *multi-class accuracy*, defined as the share of patents in the test sample for which the predicted CPC code equals the observed one.

Table D.3
Field-level recall and precision: Random Forest

CPC field	Recall	Precision
A	0.802	0.764
B	0.840	0.796
C	0.855	0.849
D	0.778	0.892
E	0.785	0.862
F	0.831	0.850
G	0.815	0.851
H	0.876	0.885
Average	0.823	0.844
Accuracy (multi-class)	0.831	

Notes: Rows A–H report recall and precision for each one-digit CPC field. Recall is defined as $\Pr(k \in \hat{F} \mid k \in F)$ and precision as $\Pr(k \in F \mid k \in \hat{F})$. The "Average" row reports averages of recall and precision across fields. The final row reports the overall *multi-class accuracy*, defined as the share of patents in the test sample for which the predicted CPC code equals the observed one.

Table D.4
Field-level recall and precision: Support Vector Machine

CPC field	Recall	Precision
A	0.826	0.774
B	0.850	0.831
C	0.873	0.882
D	0.823	0.902
E	0.789	0.863
F	0.854	0.854
G	0.830	0.846
H	0.878	0.891
Average	0.840	0.856
Accuracy (multi-class)	0.846	

Notes: Rows A–H report recall and precision for each one-digit CPC field. Recall is defined as $\Pr(k \in \hat{F} \mid k \in F)$ and precision as $\Pr(k \in F \mid k \in \hat{F})$. The "Average" row reports averages of recall and precision across fields. The final row reports the overall *multi-class accuracy*, defined as the share of patents in the test sample for which the predicted CPC code equals the observed one.