Tracking the Economy through Firm Creation: Evidence from Real-Time Administrative Data

Authors:
Yannis Galanakis[x]
King's College London
Anthony Savagar[y]
Bank of England

Date:
November 2025

The Productivity Institute
Working Paper No.061

[x] Economic Statistics Centre of Excellence and The Productivity Institute
[y] University of Kent and Economic Statistics Centre of Excellence

Key words
Business Dynamism, Real-Time Indicators, Administrative Data, Economic Monitoring, Economic Measurement

JEL codes
C81, D24, L11, L25, L26, O47

Authors' contacts:
yannis.galanakis@kcl.ac.uk
asavagar@gmail.com

Acknowledgements
We thank Ambrogio Cesa-Bianchi, Rebecca Riley, Javier Miranda, Neeltje Van Horen, Misa Tanaka and Petr Sedláček for their comments. We thank conference and seminar participants at: Bank of England, ESCoE Manchester 2024, CAED Penn State 2024, UNSW-ESCoE Conference on Economic Measurement 2024, NIESR Brownbag Seminar, Workshop on firm survival at the University of Westminster and Productivity Research Conference 2025 by TPI. We acknowledge financial support from ESRC grant ES/V002740/1.

Suggested citation

Y.Galanakis, A. Savagar (2025) *Tracking the Economy through Firm Creation: Evidence from Real-Time Administrative Data*. Working Paper No. 061, The Productivity Institute.

The Productivity Institute is an organisation that works across academia, business and policy to better understand, measure and enable productivity across the UK. It is funded by the Economic and Social Research Council (grant number ES/V002740/1).

The Productivity Institute is headquartered at Alliance Manchester Business School, The University of Manchester, Booth Street West, Manchester, M15 6PB. More information can be found on The Productivity Institute's website. Contact us at theproductivityinstitute@manchester.ac.uk

Abstract

We introduce a novel real-time dataset—Companies House Real-Time (CHRT)— that captures daily firm creation and dissolution activity for the full population of UK-registered companies. CHRT tracks official business demography statistics but is available months earlier, providing timely disaggregated indicators. We demonstrate that firm entry is a leading indicator of GDP and employment. Using a structural vector autoregression (SVAR), we find that a one-standard-deviation increase in firm entry raises GDP by 0.1–0.2% over the following year and generates persistent gains in employment and productivity. These results highlight the value of real-time administrative data for macroeconomic monitoring, and underscore the importance of business formation as a margin of adjustment during economic fluctuations. Our findings suggest that firm entry data should be integrated into early-warning systems and policy frameworks, particularly in times of crisis or structural change.

Introduction

This paper demonstrates that real-time firm creation data offer a timely and policy-relevant measure of macroeconomic conditions. Using a novel high-frequency dataset based on UK company registrations, we show that firm entry is a leading indicator of official business statistics and macroeconomic aggregates—and that entry shocks have persistent causal effects on output, employment, and productivity. A one-standard deviation increase in the firm entry rate leads to a 0.1–0.2% rise in GDP, with effects persisting for 8 to 12 quarters. These results highlight firm formation as an important transmission mechanism for macroeconomy.

A growing body of evidence documents that rates of firm entry, growth, and exit have slowed in the United States (Akcigit and Ates 2023), Europe (Biondi, Inferrera, Mertens, and Miranda 2023), and across the OECD (Calvino, Criscuolo, and Verlhac 2020). However, much of this evidence is based on datasets with substantial publication lags, limiting their usefulness for monitoring turning points in economic activity or informing real-time policy decisions. The absence of timely indicators of firm creation and reallocation has left policymakers without early-warning signals during economic shocks.

This paper addresses that gap by leveraging a novel source of high-frequency administrative data. In the UK, all firms are required to register with Companies House, the official registrar of companies. Over the last decade, the registration process has moved entirely online, allowing for near-instantaneous collection of detailed firm-level information. The UK has emerged as a global leader in corporate data transparency: Companies House data are released publicly and at daily frequency, with broad coverage across sectors and regions.[1]

We construct the *Companies House Real-Time* (CHRT) dataset, a new administrative dataset that captures real-time firm entry and exit in the UK economy. CHRT covers all incorporated (limited liability) firms and is available at daily frequency by industry and region. It allows for disaggregated and high-frequency monitoring of business dynamism using legally defined events—incorporation and dissolution. A companion public dashboard displaying weekly CHRT indicators by sector and region is available at: https://asavagar.shinyapps.io/BusinessDynamicsDashboard/. This tool enables near-instantaneous monitoring of UK business formation and can support policy analysis and crisis response.

We make three contributions. First, we develop and validate a novel real-time dataset on firm creation. We show that it tracks official Companies House statistics released with a delay, making it suitable for early monitoring of business activity. Second, we demonstrate that our measure of firm entry is a leading indicator of other official data—

[1] OpenCorporates ranks the UK first in its global Open Company Data Index. See https://opencorporates.com/registers.

specifically, tax-based business demography statistics—and also anticipates macroeconomic aggregates such as GDP and employment. Third, we estimate a structural VAR to show that shocks to firm entry have persistent causal effects on employment and output.

Our approach offers a contrast to existing real-time indicators of UK economic activity—such as card transactions, shipping volumes, and mobility data—which predominantly reflect short-term consumption patterns.[2] While these measures are useful for capturing immediate fluctuations in behaviour, they provide limited insight into structural economic changes or forward-looking investment dynamics. CHRT firm creation data reflects supply-side capacity formation. It provides a legally grounded, universal coverage dataset that precedes official tax-based records by several months. Firm creation represents a commitment of capital, time, and regulatory effort. Incorporation reflects expectations about future profitability, cost conditions, and demand. As such, it is a fundamentally forward-looking economic decision that captures changes in the supply side of the economy. For example, during the COVID-19 pandemic, CHRT data revealed a sharp and sustained rebound in firm registrations by mid-2020—particularly in online retail, delivery logistics, and digital services—well before the recovery in GDP and employment became visible in official statistics (Bahaj, Piton, and Savagar 2024). This anticipatory signal was absent in standard real-time indicators, underscoring the unique macroeconomic value of firm creation data.

Furthermore, unlike other real-time indicators, firm creation is a fundamental variable determining production capacity in various economic theories. In recent dynamic general equilibrium models with entry frictions and sunk costs – such as those by Bilbiie, Ghironi, and Melitz (2012), Lewis and Poilly (2012), and Sterk, Sedláček, and Pugsley (2021) – entry is a state variable that evolves endogenously and drives both cyclical fluctuations and long-run growth. Similarly, in models of endogenous growth with creative destruction, new firms enable innovation (Acemoglu, Akcigit, Alp, Bloom, and Kerr 2018). From this perspective, real-time monitoring of firm creation provides insights not only into current business sentiment, but into the future path of output, employment, and productivity.

A recent example illustrates the advantages of real-time firm creation data for policy analysis. During the COVID-19 pandemic, Office for National Statistics (ONS) mobility and spending indicators dropped rapidly in March 2020, reflecting the lockdown-induced collapse in economic activity. However, CHRT data showed a sharp and sustained rebound in firm creation by mid-2020, particularly in online services, delivery logistics, and digital consultancy (Bahaj, Piton, and Savagar 2024). This rebound preceded the broader recovery in output and employment and highlighted a structural reallocation

[2]See the ONS real-time indicators dashboard: https://www.ons.gov.uk/economy/economicoutputandproductivity/output/bulletins/economicactivityandsocialchangeintheukrealtimeindicators/24july2025#latest-indicators-at-a-glance

within the economy. Unlike other real-time indicators, firm creation data captured the forward-looking entrepreneurial response and helped anticipate post-COVID economic restructuring. A similar analysis showing the long-run employment effects of firm creation is presented in Benedetti-Fasil, Sedláček, and Sterk (2022).

Related Literature

This paper contributes to two strands of research: the macroeconomic effects of firm entry and real-time measurement of business dynamism.

Macroeconomic effects of firm entry Our work is related to Gourio, Messer, and Siemer (2016), who estimate the effect of firm entry on US GDP using annual state-level data and local projections. Our contribution differs by focusing on novel, high-frequency, entry data which enables timely monitoring for policy, including over the COVID-19 period. We review their results relative to ours in Section 3.5.

Tian (2018) documents the business cycle properties of firm entry and exit in the US using lower frequency annual data. Our analysis supports these findings but through a higher-frequency, real-time, lens. Various other papers develop structural models to understand how firm entry transmits other shocks or acts as a shock itself. Our predictions on the transmission of entry shocks to output and employment are consistent with evidence in Lewis and A. Stevens (2015) and Lewis and Winkler (2017). Sedláček (2020) makes the point that falls in firm entry during recessions are a key determinant of future employment, ahead of other general equilibrium responses, whilst Sedláček and Sterk (2017) explain that the conditions on startup are vital for future outcomes. All these structural modelling papers, and others (Gutiérrez, Jones, and Philippon 2021), include exercises on the effects of entry cost shocks, but a persistent challenge is identification of entry shocks, since entry responds endogenously to macroeconomic factors such as business cycles, monetary and fiscal policy, and tax incentives. A central contribution of our work is to make progress on this identification.

Real-time business formation data There is growing interest in using real-time data to monitor firm creation and broader economic activity. In the U.S., Bayard, Dinlersoz, Dunne, Haltiwanger, Miranda, and J. Stevens (2018) introduce the Business Formation Statistics (BFS), which are based on Employer Identification Number (EIN) applications. BFS is an indirect measure of firm creation based on payroll data and registration with the tax authority (IRS). In contrast, Companies House data reflect the legal act of incorporation, offering clean, immediate, measurement with precise documentation on postcode and 5-digit sector.

Recent BFS-based research includes Dinlersoz, Dunne, Haltiwanger, and Penciakova

(2021), who project employer transitions during COVID-19 using a linear probability model, and Haltiwanger and Decker (2023), who explore whether the pandemic-era surge in entry marks a reversal of declining U.S. business dynamism. Asturias, Dinlersoz, Haltiwanger, and Hutchinson (2023) show that real-time registrations in the U.S. can serve as leading indicators of macroeconomic activity. We contribute cross-country evidence by presenting similar descriptive statistics for the UK, and extend the analysis by estimating a structural VAR to recover causal effects of entry shocks.

Roadmap

The remainder of this article is structured as follows. Section 1 describes the collection process, main features, and limitations of our data. Further, we compare our data to official government statistics. Section 2 shows that our data is a leading indicator of official statistics. Section 3 uses a three-variable structural VAR model with Cholesky decomposition as identification. Section 4 concludes with a summary of our results.

1 Data

1.1 Companies House Real-Time data

We construct a novel, high-frequency panel dataset – referred to as the Companies House Real-Time (CHRT) data – based on publicly available records from Companies House. Companies House is the UK official company registrar of all incorporated limited liability entities. Drawing on monthly register snapshots and the Companies House API tool, the CHRT dataset provides a census of the universe of incorporated companies in the UK since 2012. The dataset allows for granular tracking of firm-level dynamics, including entries, exits, and changes in industrial or geographic classification; thereby offering a real-time view of the corporate demography in the UK.

1.1.1 Data collection

To assemble the CHRT dataset, we integrate multiple data products disseminated by Companies House. Appendix A describes in more detail our data compilation process.

Active and new companies Our primary source for the stock of active firms, as well as for identifying new entrants, is the static register of companies published periodically by Companies House. These registers include all 'live' companies, i.e. with a non-dissolved status as of the reference date. They have been released monthly since the second quarter of 2020, with less frequent coverage before that point. Firms are considered active unless formally dissolved, even if they are currently in the process of dissolution (typically

within a three-month window). This definition diverges from the operational definitions of company activity employed by other UK statistical authorities such as HM Revenue and Customs (HMRC) and the Office for National Statistics (ONS) Inter-Departmental Business Register (IDBR). We use the date of register release to monitor the stock of active companies.

We identify new firms, i.e. entrants, in each period by tracking incorporations listed in the latest register that do not appear in the preceding one. Incorporation dates provided in the register allow us to verify and timestamp these entry events precisely. By comparing subsequent snapshots of the register, we further track post-entry transitions in firms' Standard Industrial Classification (SIC) codes and geographic locations, enabling us to monitor sectoral reclassification and spatial mobility within the active firm population.

Dissolutions Since the static registers may retain firms that are in the process of dissolution for up to three months, we complement the register data with dissolution records retrieved via the Companies House API. This API provides time-stamped information on the exact date of dissolution, as well as the company's last known geographic location and SIC code before exit. We define firm exit strictly based on the official dissolution date recorded in the API, rather than the date at which a firm is removed from the register. This approach ensures temporal accuracy in identifying exits and mitigates misclassification biases that arise from the lag between a firm's cessation of legal existence and its administrative removal from the register.

Accounts data The administrative definition of 'activity' employed by Companies House – namely, the inclusion of any non-dissolved company in the register – may overstate the extent of economically meaningful firm presence. Many such entities may be legally extant but economically dormant, lacking operational activity, employment, or revenue generation. To address this limitation, and for robustness purposes, we supplement our dataset with monthly filings of company accounts, as submitted electronically to Companies House. These filings are available only for companies that file online, a group which constitutes approximately 75% of the roughly 2.2 million annual account submissions expected by Companies House.

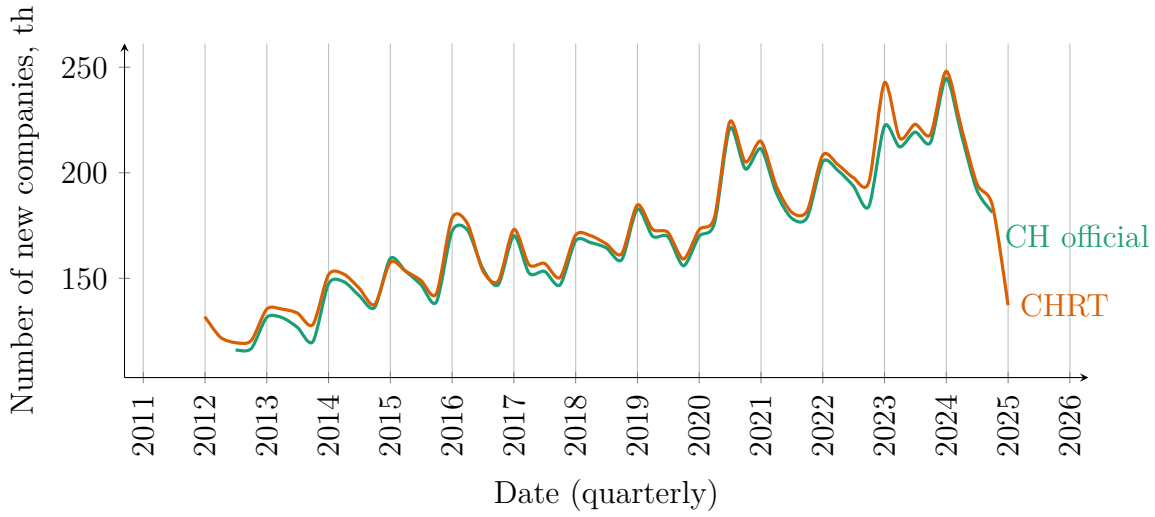
In line with criteria used by the UK Department for Business and Trade, we further define a firm as economically active if (a) it has submitted non-dormant accounts in the past two years and (b) exhibits changes in reported current assets across consecutive filings. We look at the current assets against other balance sheet information as micro companies are mandated to report.

1.1.2 Data features and limitations

Coverage The CHRT dataset provides comprehensive coverage of all companies incorporated in the United Kingdom, regardless of the nationality of their owner and across all sectors and locations of the economy. Firms are classified according to the UK Standard Industrial Classification (SIC) system, with codes available at the five-digit level. In addition to standard SIC categories, Companies House includes designations that facilitate the identification of firms with limited or no economic activity. Specifically, firms may be flagged as non-trading (SIC code 74990) or dormant, either through a special SIC code (99999) or by submitting dormant accounts[3]. Geographic coverage is similarly granular: firms are geolocated using full UK postcode information, enabling precise spatial analysis at a fine level of aggregation.

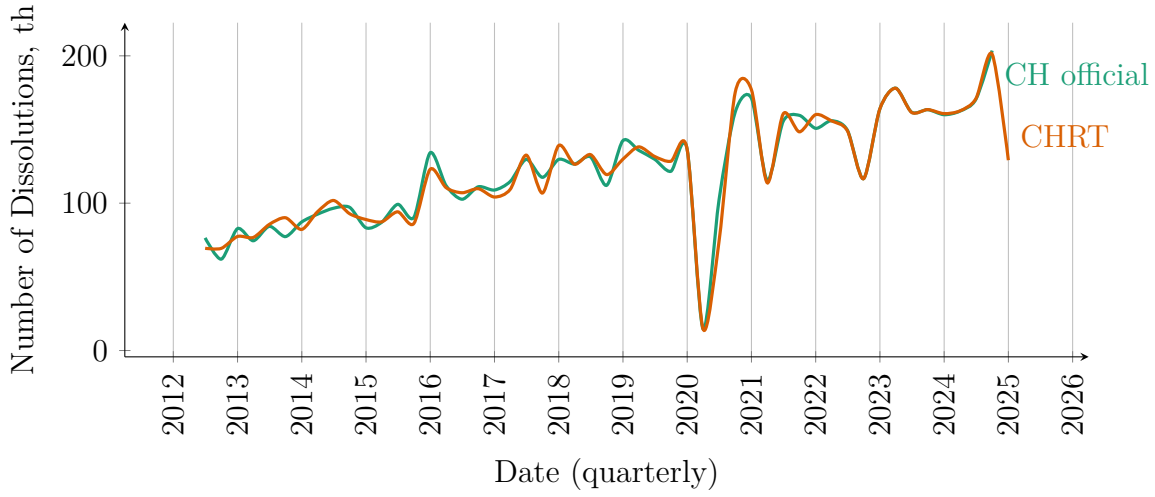
External validity Companies House releases their own statistics at a quarterly frequency with a time lag. We compare our real-time construction with their release in figure 1. Our data matches their data closely for both incorporations (figure 1a) and dissolution (figure 1b). In appendix, we compare the stock of active companies, too.

[3]There is no legal requirement for dormant firms to update their SIC codes or to file dormant accounts, potentially leading to underestimation of dormant firms in our data.



(a) CH Offical v CHRT Incorporations

Note: Our timing convention is, for example, data for the second quarter of 2015 are represented by the date 2015-04-01. The grey vertical lines represent the first quarter of the year, for example, 2024-01-01 represents the first three months of 2024. CH official statistics exclude limited liability partnerships from figures.



(b) CH Official v CHRT Dissolutions

Note: Our timing convention is, for example, data for the second quarter of 2015 are represented by the date 2015-04-01. CH official statistics exclude limited liability partnerships from figures. From 2022/23 onwards, CH official statistics count dissolutions when The Gazette notice is published, as opposed to when a company is dissolved on the Companies House system. These events usually take place approximately one week apart. Companies House updates dissolution as they are actioned in near real-time through the Rest API. Any differences between the CH official and CHRT are due to company restoration, which takes a company out of the dissolution status.

Figure 1: CH Offical Statistics v CHRT

Notes: CHRT data from 2012-07-01 – 2025-03-31.
Source: Authors' calculations based on CHRT data and Companies House Official Statistics (CH).

Limitations While the CHRT dataset offers near-universal coverage of UK-incorporated companies, several limitations must be acknowledged. First, the register excludes non-incorporated business forms such as sole traders, partnerships, and self-employed individuals not registered for VAT or PAYE. These legal form of businesses collectively account for approximately half of the UK business population. The economic behavior of these

entities differs markedly from incorporated firms. For example, self-employment increases in recessions ('entrepreneurship out of necessity'), while firm creation is procyclical (excluding the pandemic period).

Second, industry classifications are self-reported by firms. They appear to be less reliably populated prior to 2017, potentially limiting the accuracy of sectoral analyses in earlier years. Third, the frequency of register snapshots varies over time: while monthly data are available from the second quarter of 2020 onward, earlier coverage is more intermittent. The biggest gap is between 2014/5. This affects the precision of entry rates and our firm survival analysis.

Finally, the high incidence of multiple company registrations at a single postcode. This is driven by factors such as virtual offices, formation agents, and administrative addresses (e.g., HMRC or company secretarial services) introducing distortions in geographic analysis. These clustering effects can bias local-level estimates of firm births, deaths, and survival rates, especially when used in conjunction with ONS spatial statistics[4].

1.1.3 Measures of interest

Entry/Exit rate Our main measure of business dynamism is the company entry rate[5]. This indicator is widely used in the literature. Its measurement is easy and helpful for cross-country comparisons. Following Calvino, Criscuolo, and Verlhac (2020), we define the entry rate (ER) as the number of new (or entrant, E) firms over the number of active (N, i.e. new and incumbent) firms in the same period t, or

$$\mathrm{ER}_t = \frac{E_t}{N_t} \tag{1}$$

The equivalent exit rate (ExR) looks at the number of the dissolved firms (D) over the number of active (N) firms in the same period t, or

$$\mathrm{ExR}_t = \frac{D_t}{N_t} \tag{2}$$

Given the granularity of our data, we measure business dynamism both at a quarterly and annual level, depending on the exercise and the level of any external data availability.

[4]For further details, see ONS analysis https://www.ons.gov.uk/businessindustryandtrade/business/activitysizeandlocation/articles/multiplebusinessregistrationsatasinglepostcodeuk/2023.

[5]Usually, the literature uses the job reallocation rate. We do not observe the employment at the time of company creation. We could only observe it once companies submit their first accounts to Companies House online. Since only 75% of companies submit accounts online, a measure that depends on employment would be biased.

1.2 ONS Business Demography

The Office for National Statistics (ONS) maintains the Inter-Departmental Business Register (IDBR), the core administrative source used for producing official UK business demography statistics.[6] The IDBR includes businesses that meet specific administrative criteria: either registration for Value Added Tax (VAT), indicating a turnover above the VAT threshold (currently £85,000), or registration for Pay As You Earn (PAYE), indicating the presence of at least one paid employee. The ONS publishes quarterly series on additions to (entries) and removals from (exits) the IDBR as part of its Business Demography (BD) statistical release, covering all sectors of the UK economy since 2017.

Unlike the Companies House register, which records all incorporated entities at the point of legal formation, the IDBR focuses on businesses that are actively engaged in economic production as defined by their tax obligations. Moreover, while Companies House is limited to limited liability firms, the IDBR includes a wider range of legal forms, including self-employed individuals and unincorporated businesses. As a result, the IDBR captures a more economically diverse group of firms—but with a significant delay relative to real-time incorporation data.

1.2.1 From Companies House to Tax Authority to ONS

Incorporation with Companies House is the first step in a multi-stage administrative process that determines whether a business ultimately appears on the IDBR. After incorporation, firms must register for tax with HM Revenue and Customs (HMRC), either for VAT or PAYE, depending on whether they expect to exceed the turnover threshold or hire employees. Only when this tax registration is completed and verified by HMRC is the firm eligible for inclusion on the IDBR, following periodic information exchange between HMRC and the ONS.

This sequential process introduces a measurable lag between legal formation and appearance in official business registers. According to Office for National Statistics (2022), approximately 85% of firms that eventually appear on the IDBR do so within 12 months of their initial Companies House registration. However, this lag varies systematically by industry, firm size, and region. Firms with employees or higher expected turnover tend to appear more quickly, while sole traders, dormant entities, or firms that never engage in taxable activity may remain permanently excluded.

Our CHRT dataset captures all incorporations at the moment of legal registration and is therefore more timely than tax-based or register-based statistics. While CHRT includes some firms that will never become economically active, it also allows us to observe entry

[6] Accredited researchers can access anonymised microdata versions of the IDBR through the Business Structure Database (BSD) and Longitudinal Business Database (LBD), both available via secure ONS Research Secure Labs. These datasets are subject to access protocols and are typically released with a substantial time lag.

dynamics in near real time and anticipate the trajectory of official business demography measures months in advance.

1.3 ONS Estimates on Productivity, Employment and Output

To estimate the dynamic effects of firm entry shocks on aggregate employment, productivity, and output, we employ a structural vector autoregression (SVAR) framework. This analysis draws on several high-frequency macroeconomic datasets from the UK Office for National Statistics (ONS). For productivity, we consider the labour productivity series (output per worker), available quarterly from 2012Q3 to 2024Q4[7]. For employment and output, we use monthly PAYE employee counts and Gross Value Added (GVA) estimates[8] from July 2014 until April 2025.

2 Descriptive Statistics

We assess the informational value of the CHRT dataset by evaluating its timeliness relative to official statistics. Real-time administrative data may provide earlier signals of firm dynamics than conventional sources, which are typically released with delay. We pursue two strategies. First, we examine whether CHRT-based indicators lead or lag official statistics from the Office for National Statistics (ONS). Second, we track firm transitions from incorporation to subsequent indicators of economic activity, such as employment or tax registration, to evaluate how quickly CHRT captures real economic engagement.

2.1 CHRT Predicting Official Government Data

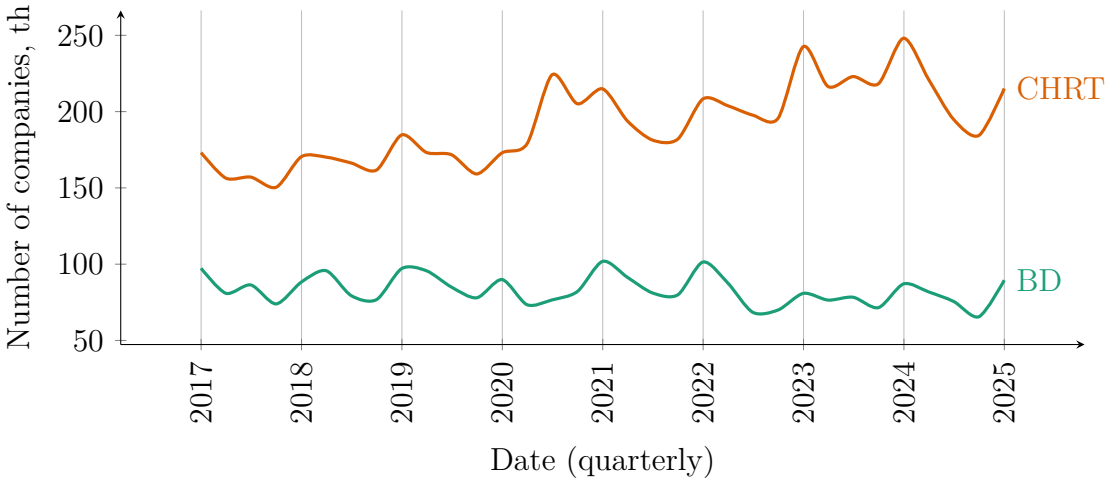
First, we compare the entry and exit between CHRT and ONS BD. The aforementioned eligibility criteria for IDBR generate conceptual differences between BD and CHRT. BD uses IDBR to count the number of enterprises, while CHRT counts registered companies. An enterprise is a *statistical* unit and may include several companies. A registered company is a *legal* unit.

[7]ONS Labour Productivity: Tables 13, 14, 17, and 18, available at https://www.ons.gov.uk/economy/economicoutputandproductivity/productivitymeasures/datasets/outputperworkeruk (retrieved May 15, 2025). From this file we collect GVA quarterly estimates, too.

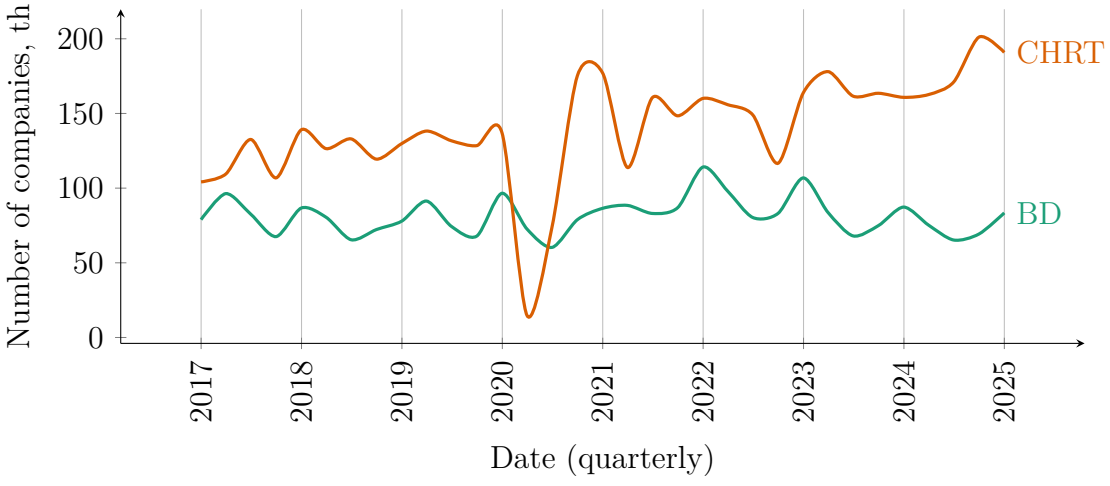
[8]Employment: Table 1 from https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/earningsandemploymentfrompayasyouearnrealtimeinformationuk/june2025 (retrieved June 10, 2025).
GVA: Columns 1 and 2 from https://www.ons.gov.uk/economy/grossdomesticproductgdp/datasets/gdpmonthlyestimateuktimeseriesdataset (retrieved June 12, 2025).

2.1.1 Comparison in levels

Figure 2 compares the firm entry and exit we get from CHRT and the ONS Business Demography (BD). CH always exceeds BD because many company registrations will not grow large enough to register for VAT or PAYE tax. A growing gap between the two lines indicates lower pass-through of company registrations to the tax-based Government register.



(a) Entry



(b) Exit

Figure 2: CHRT vs. ONS Business Demography: entry and exit

Notes: There are conceptual differences between CHRT and BD. CHRT counts for the number of companies registered with Companies House. ONS Business Demography counts the number of enterprises that appear on the IDBR as described in the main text.
Source: Authors' calculations based on CHRT and ONS BD data

2.1.2 Reduced-form VAR

To examine the dynamic relationship between the number of firms in the Business Demography (BD) data and new registrations from Companies House (CH), we estimate

a reduced-form Vector Autoregression (VAR) model. The quarterly data are seasonally adjusted and detrended to ensure stationarity, and the model is specified with four lags.

Figure 3 presents the impulse response function for a one standard deviation shock to CH registrations. The shocks are identified using a Cholesky decomposition, with the variables ordered such that CH registrations are contemporaneously exogenous to the BD firm count. This assumption means that while the number of BD enterprises can respond immediately to a shock in CH registrations, CH registrations do not respond to a BD shock within the same quarter.

The results indicate that a one standard deviation shock to CH registrations is followed by a statistically insignificant contemporaneous increase of approximately 1% in the number of BD firms. However, the response grows and becomes statistically significant two quarters after the initial shock, peaking at an increase of 4%. This finding confirms that new company registrations go on to form legitimate businesses with administrative tax footprints in official statistics.

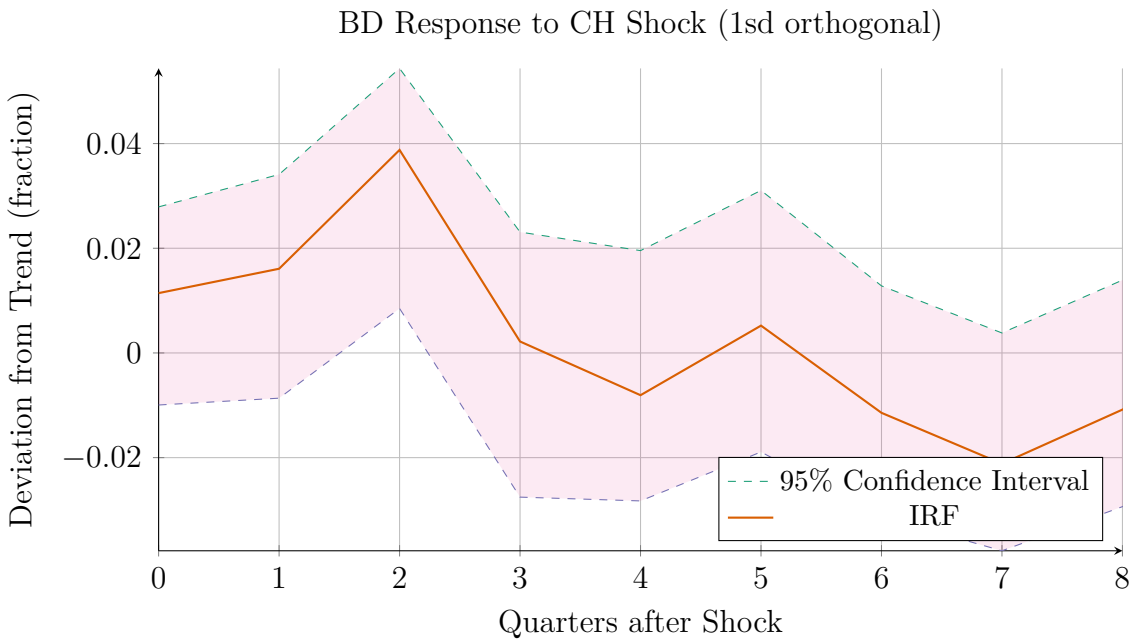


Figure 3: Impulse Response Function from a Reduced-Form VAR

Notes: The plot shows the impulse response of business demography (BD) to a one standard deviation shock in CHRT data, based on a VAR(4) model with a constant. The shock to CHRT is orthogonalised using Cholesky decomposition, ordered such that CH does not respond instantaneously to BD. Both time series are seasonally adjusted, de-trended and stationary. The shaded area is the 95% bootstrap confidence interval generated from 500 iterations. **Source:** Authors' elaboration based on CHRT and ONS Business Demography data.

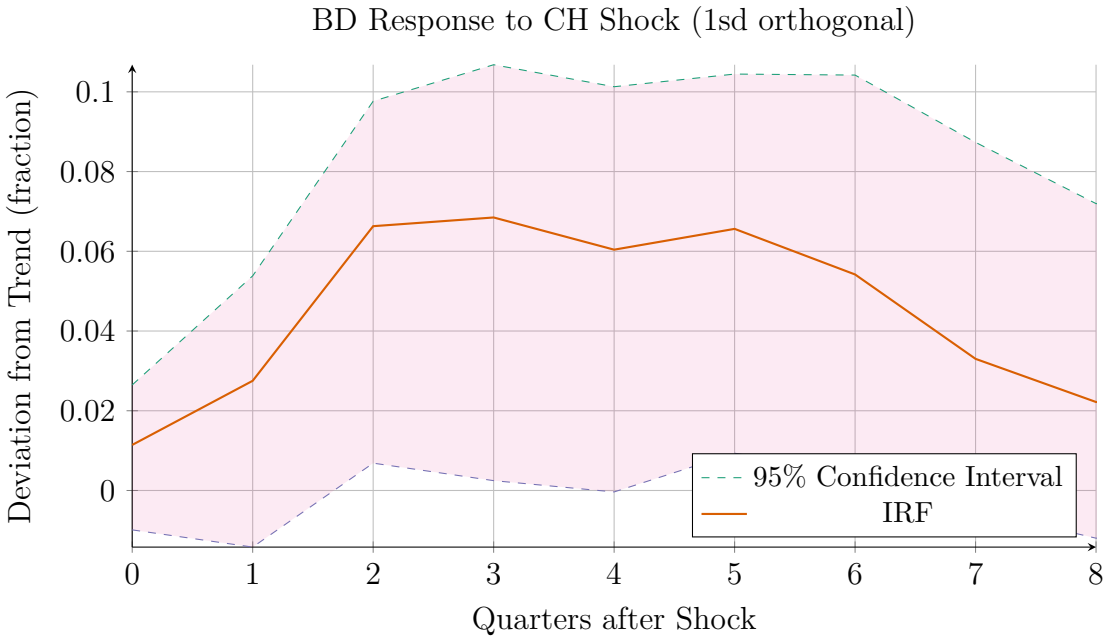


Figure 4: Impulse Response Function from a Reduced-Form VAR (cumulative)

Notes: The plot shows the cumulative impulse response of business demography (BD) to a one standard deviation shock in CHRT data, based on a VAR(4) model with a constant. The shock to CHRT is orthogonalised using Cholesky decomposition, ordered such that CH does not respond instantaneously to BD. Both time series are seasonally adjusted, detrended and stationary. The shaded area is the 95% bootstrap confidence interval generated from 500 iterations. **Source:** Authors' elaboration based on CHRT and ONS Business Demography data.

2.2 Transition from Registration to Employment/Tax footprint

We link company registrations to their subsequent inclusion in the IDBR. This indicates that a new registration has grown to a size where it qualifies for payroll tax (PAYE) or VAT. As described in Section earlier, businesses appear on IDBR when they meet VAT or PAYE thresholds.

Figure 5 plots the share of CH companies that match any IDBR enterprise within 2, 4, 6, 8 and 12 quarters from their incorporation. First, we count the number of companies that have been linked to an enterprise either by satisfying the PAYE or the VAT criterion to join IDBR by year and quarter of incorporation and year and quarter of IDBR presence. An enterprise may have a *simple match* (1-to-1, i.e. one enterprise to one CH company) or a *multiple match* (1-to-many, i.e. one enterprise to multiple CH companies). Here we count the companies with *any* match type. Second, we count the number of new incorporations by year and quarter of incorporation in CHRT data. Third, we take the share of companies that show up in IDBR within 2 quarters (or 6 months), 4 quarters (or 1 year), 6 quarters (1.5 years), 8 quarters (or 2 years) or 12 quarters (3 years). There is some time lag with the latest available data in IDBR; this matching exercise includes data up to the last quarter of 2022. We commence the timeseries from 2012Q3, when the first register in CH becomes available. The first point of each timeseries means the following. For the cohort of companies that incorporated in 2012Q3, 20%, 28%, 37%,

40% and 43% show up in IDBR within 6, 12, 18, 24 and 36 months, respectively.

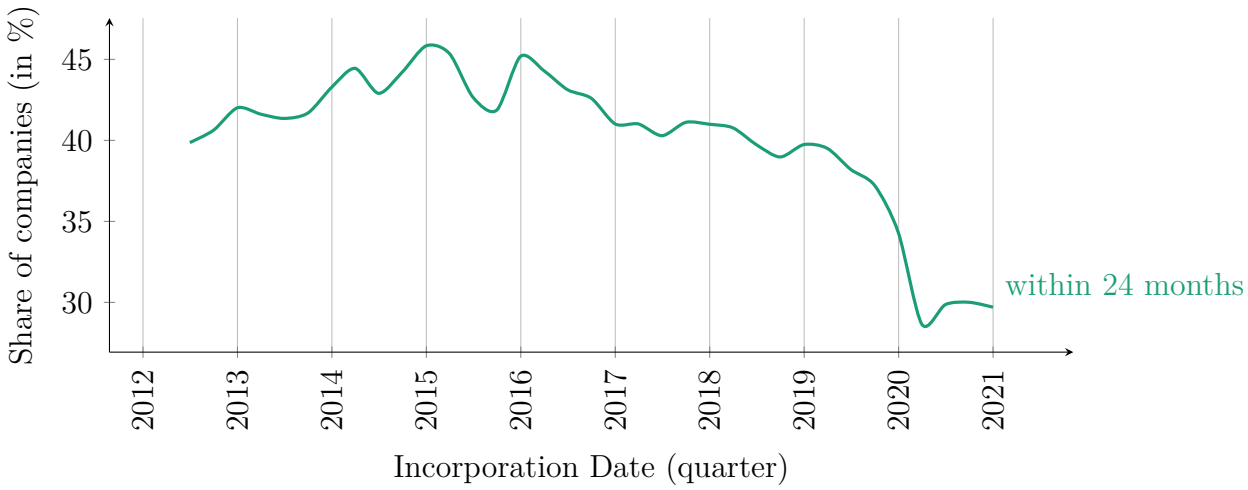


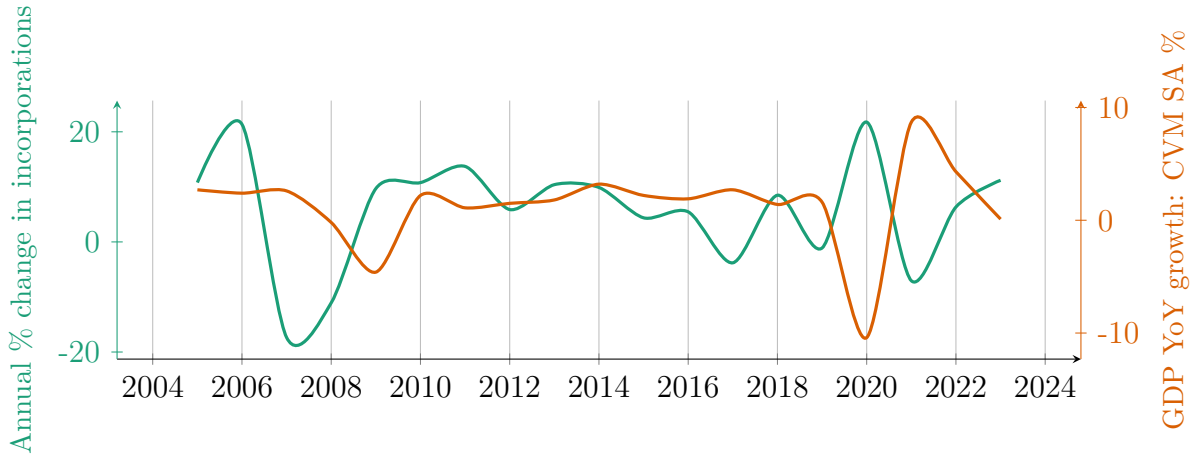
Figure 5: Share of companies that appear in IDBR, 2012Q3–2022Q4

Notes: The numerator is the distinct number of companies that match to an enterprise in IDBR by satisfying either the PAYE or the VAT criterion by year and quarter of incorporation and IDBR year and quarter. The denominator is the distinct number of companies by year and quarter of incorporation. Due to data availability constraints LBD data are up to 2022Q4.
Source: Authors' elaboration based on CHRT and LBD data.

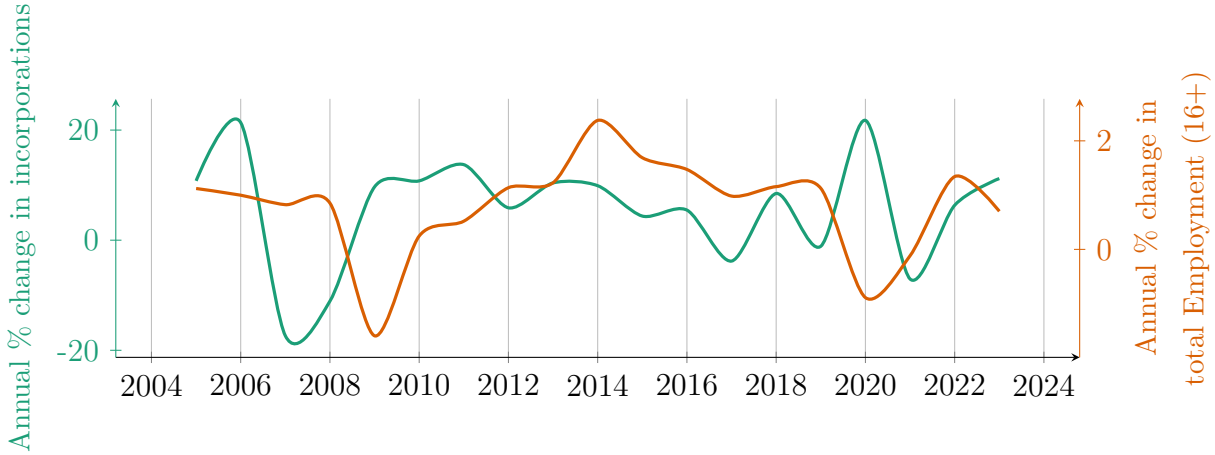
We think the most meaningful statistic is the 2-year horizon. This time threshold is closer to the mandatory first-accounts reporting deadline.

2.3 Firm Creation as a Cyclical Indicator

Before estimating the structural macroeconomic effects of firm creation, we provide descriptive evidence that firm entry is a forward-looking indicator of the UK business cycle. Figure 6 plots the annual percentage change in firm incorporations (from Companies House annual statistics) against GDP growth and total employment growth since 2005.



(a) Incorporations vs. GDP growth



(b) Incorporations vs. total employment growth

Figure 6: Firm creation is a leading indicator of growth

Notes: The left-hand y-axis (green) is the same in both panels. It shows the annual percentage change in incorporations. Number of incorporations are reported in Companies House annual statistical releases and are calculated on a financial year basis. In panel a, the right-hand y-axis shows the year-on-year GDP growth as reported by the ONS. In panel b, the right-hand y-axis shows the annual percentage change in total employment (aged 16+, seasonally adjusted).
Source: Authors' elaboration based on Companies House Annual Reports and ONS data (full source details omitted here for brevity).

The figure shows that firm creation typically declines ahead of recessions. Entry dropped sharply before the 2008 Global Financial Crisis and immediately following the COVID-19 shock. These movements anticipate broader swings in GDP and employment, reinforcing the view that firm creation reflects forward-looking investment and production decisions.

This descriptive evidence motivates our structural analysis in the following sections, where we quantify the causal effects of firm entry on macroeconomic outcomes using a structural VAR framework.

3 The Effect of Firm Entry Shocks on Aggregate Variables

This section investigates the short-run macroeconomic effects of firm entry shocks using a structural vector autoregression (SVAR) framework. Our primary objective is to trace how innovations in firm creation affect key aggregate outcomes—productivity, employment, and GDP—through impulse response analysis.

3.1 Identification Strategy

We identify structural shocks using a recursive ordering via Cholesky decomposition. Based on economic theory and institutional considerations, we order the variables as follows: firm entry, employment or productivity, and GDP. This ordering reflects several assumptions about the contemporaneous relationships in the system. We treat firm entry as the most exogenous variable: firm formation decisions are typically based on forward-looking expectations and are unlikely to respond immediately to macroeconomic aggregates within the same period. In contrast, employment and productivity may respond contemporaneously to entry shocks, while GDP—being the most aggregate measure—is allowed to respond within the same period to both.

Specifically, a structural shock to firm entry may affect both productivity (or employment) and output contemporaneously. However, shocks to GDP or productivity are not assumed to affect firm creation instantaneously, reflecting lags in entrepreneurial decision-making and firm formation. This recursive identification enables us to interpret the impulse responses as the dynamic effects of a structural entry shock on the broader economy.

3.2 SVAR Specification and Estimation Method

We estimate a three-variable SVAR of the form:

$$A_0 \mathbf{y}_t = A_1 \mathbf{y}_{t-1} + A_2 \mathbf{y}_{t-2} + \cdots + A_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{3}$$

where $\mathbf{y}_t = [\text{Entry}_t, \text{Macro}_t, \text{GDP}_t]'$ includes firm entry rate (CHRT), either PAYE employment or labour productivity, and real GDP (ONS). A_0 captures contemporaneous interactions, A_i are autoregressive coefficient matrices, and $\boldsymbol{\varepsilon}_t$ is a vector of orthogonal structural shocks. The reduced-form representation is:

$$\mathbf{y}_t = B_1 \mathbf{y}_{t-1} + B_2 \mathbf{y}_{t-2} + \cdots + B_p \mathbf{y}_{t-p} + \mathbf{u}_t, \tag{4}$$

where $\mathbf{u}_t = A_0^{-1} \boldsymbol{\varepsilon}_t$ and $\Sigma_u = E[\mathbf{u}_t \mathbf{u}_t'] = A_0^{-1}(A_0^{-1})'$.

All variables are first log-transformed and then detrended using Seasonal-Trend decomposition based on Loess (STL). The STL decomposition takes the form $\log(Y_t) = \text{trend}_t + \text{seasonal}_t + \text{remainder}_t$, and we retain only the remainder component for estimation. We set `s.window = "periodic"` (fixed seasonal component), `t.window = 64` (5-year smoothing for monthly data), and apply robust fitting to mitigate outlier influence. This approach removes seasonal and low-frequency components while preserving interpretability in log-levels. A value of 0.01 in the impulse response functions corresponds to a 1% deviation from trend. Compared to alternative filters (e.g., HP, band-pass), STL better accommodates strong seasonal patterns in firm registrations, such as fiscal year-end or holiday effects.

The frequency of analysis depends on data availability. When using PAYE employment, we estimate the SVAR at a monthly frequency with approximately 130 observations. When using productivity measures, we rely on quarterly data, which reduces the sample size. This difference affects the interpretation and statistical power of the results.

3.2.1 Estimation Procedure

The SVAR is estimated using the MATLAB VAR Toolbox 3.0 by Ambrogio Cesa-Bianchi. The estimation follows these steps:

1. Log-transform and STL-detrend each variable to extract the short-run component.
2. Confirm stationarity of remainder components using unit root tests.
3. Select the lag length using the Akaike Information Criterion (AIC).
4. Estimate the reduced-form VAR at the chosen lag.
5. Identify structural shocks using Cholesky decomposition of the residual covariance matrix.
6. Compute impulse response functions with confidence intervals from a wild bootstrap (1,000 replications).

3.3 Results: Entry, Productivity, and Output

Figure 7 presents the impulse responses of firm entry, output per worker, and GDP to an entry shock. A positive shock to firm creation results in a persistent increase in both productivity and output. These effects peak within 2–4 quarters and remain statistically significant for several periods. While the entry shock itself is short-lived, it initiates a sustained improvement in aggregate performance, consistent with the idea that new firms bring innovation and competitive pressure.

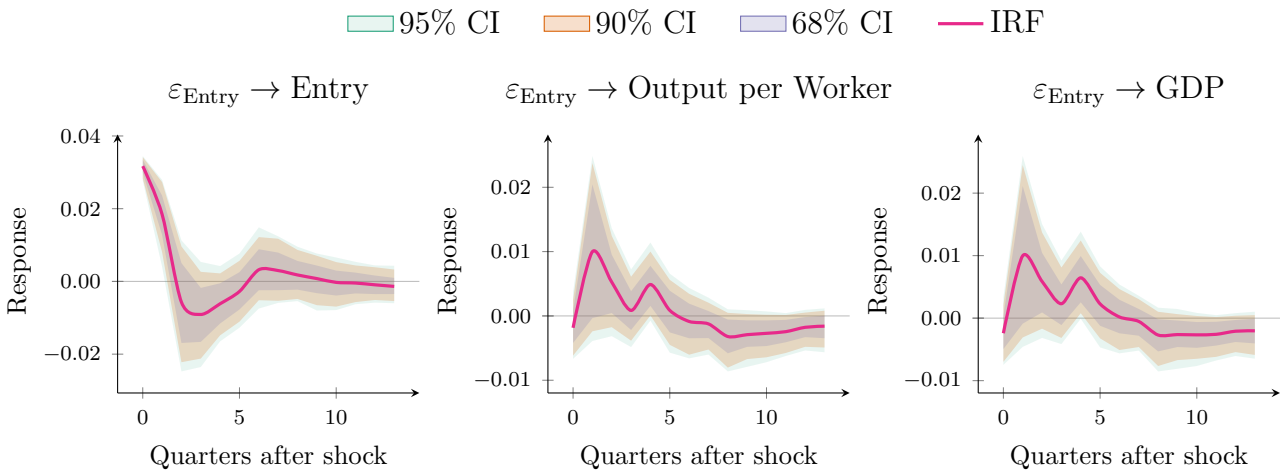


Figure 7: Impulse response functions with 95%, 90%, and 68% confidence intervals (SVAR)

Notes: Confidence intervals based on 1,000 wild bootstrap iterations. The horizon has 14 quarters. Identification via the Cholesky decomposition. Columns represent responses of: firm entry (1), output per worker (2), and GDP (3). Estimates used the MATLAB VAR Toolbox 3.0 by Ambrogio Cesa-Bianchi. The one–standard deviation shock to firm entry equals 0.0692 in period 0, implying that firm entry rises by approximately 7% above its trend level. **Source:** Authors' calculations using CHRT and ONS data.

Figure 8 displays the cumulative impulse response functions for the quarterly VAR specification. The estimates show that, four quarters after the shock, output per worker increases cumulatively by approximately 2% above trend. These persistent cumulative effects demonstrate that a one-standard-deviation (3.2%) increase in firm entry above trend generates substantial and lasting improvements in both productivity and aggregate output.

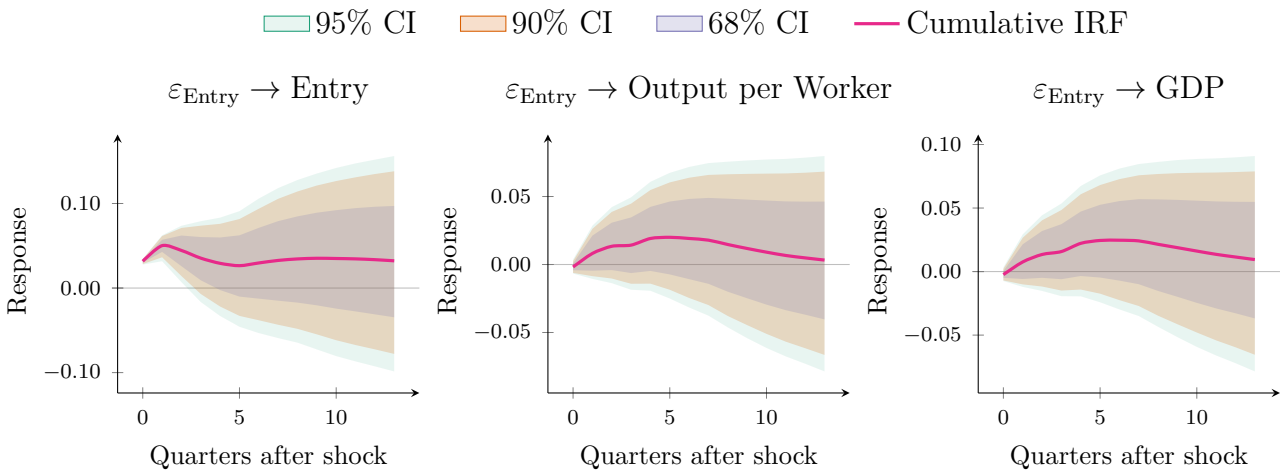


Figure 8: Cumulative Impulse response functions with 95%, 90%, and 68% confidence intervals (SVAR)

Notes: Confidence intervals based on 1,000 wild bootstrap iterations. The horizon has 14 quarters. Identification via the Cholesky decomposition. Columns represent responses of: firm entry (1), output per worker (2), and GDP (3). Estimates used the MATLAB VAR Toolbox 3.0 by Ambrogio Cesa-Bianchi. The one–standard deviation shock to firm entry equals 0.0692 in period 0, implying that firm entry rises by approximately 7% above its trend level. **Source:** Authors' calculations using CHRT and ONS data.

3.4 Results: Entry, Employment, and Output

Figure 9 shows the impulse responses when employment replaces productivity. The estimated entry shock leads to a significant and persistent increase in both employment and GDP. The employment response peaks after approximately 10 months, while GDP responds within 2–4 months. Slightly wider confidence intervals reflect greater short-term volatility in monthly employment series.

Figure 10 shows the cumulative impulse response function (IRF). The estimates indicate that, twelve months after the shock, employment rises 2.1% above its trend level, increasing to 2.8% after twenty-four months. Given the employment trend of 35 million in the UK, these deviations correspond to roughly 735,000 additional jobs after one year and 980,000 after two years. In broad terms, a one-standard-deviation (7%) increase in firm entry above trend is associated with about one million extra jobs after two years.

Together, these results support the view that real-time firm creation data capture meaningful variation in aggregate economic activity. Entry shocks propagate to both the intensive (productivity) and extensive (employment) margins, reinforcing the economic relevance of firm creation as a driver of the business cycle.

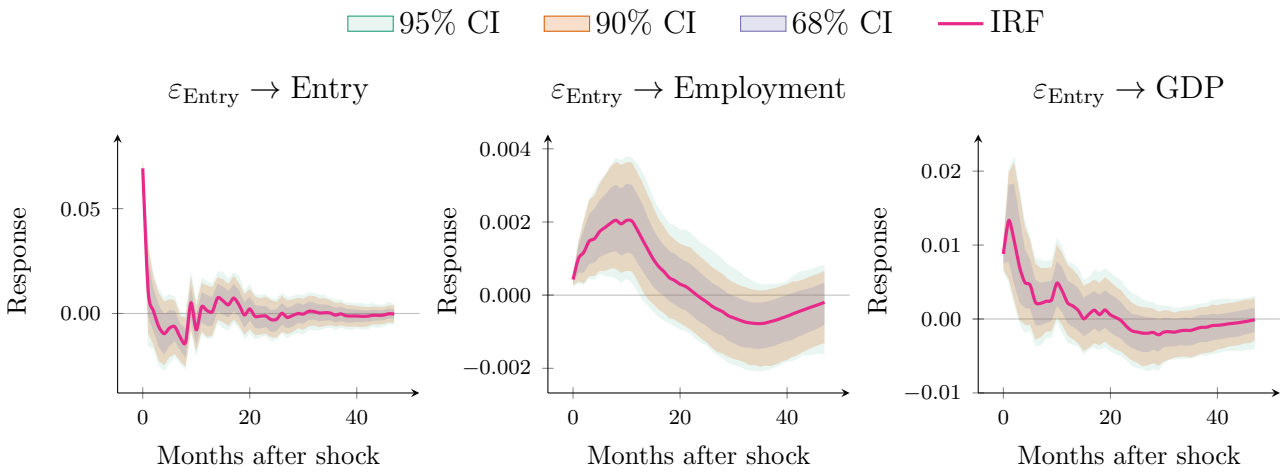


Figure 9: Impulse response functions with 95%, 90%, and 68% confidence intervals (SVAR)

Notes: Confidence intervals based on 1,000 wild bootstrap iterations. The horizon has 48 months. Identification via the Cholesky decomposition. Estimates used the MATLAB VAR Toolbox 3.0 by Ambrogio Cesa-Bianchi. The one–standard deviation shock to firm entry equals 0.0692 in period 0, implying that firm entry rises by approximately 7% above its trend level. **Source:** Authors' calculations using CHRT and ONS data.

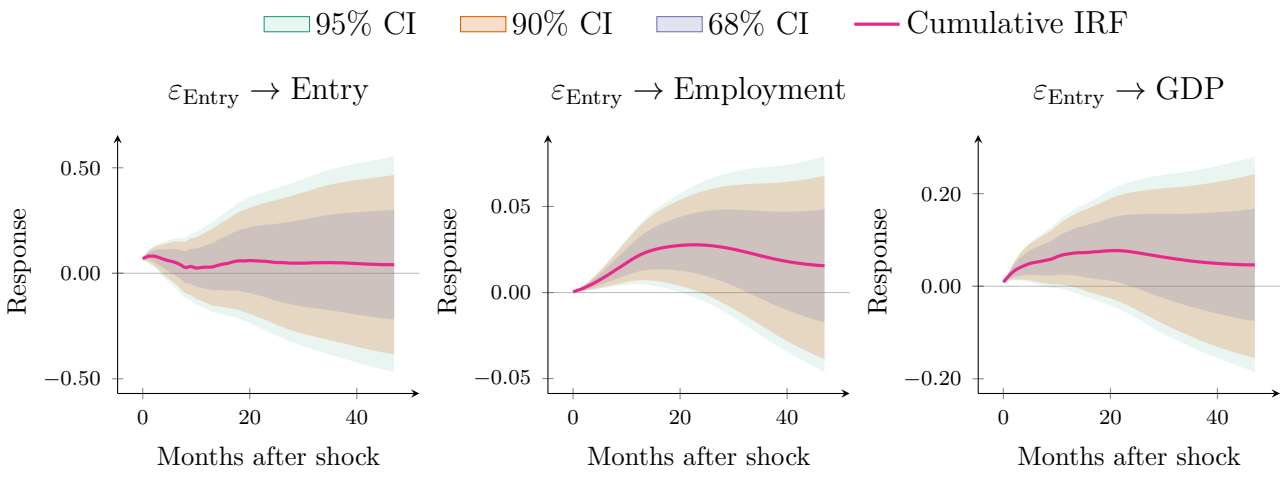


Figure 10: Cumulative Impulse response functions with 95%, 90%, and 68% confidence intervals (SVAR)

Notes: Confidence intervals based on 1,000 wild bootstrap iterations. The horizon has 48 months. Identification via the Cholesky decomposition. Estimates used the MATLAB VAR Toolbox 3.0 by Ambrogio Cesa-Bianchi. The one–standard deviation shock to firm entry equals 0.0692 in period 0, implying that firm entry rises by approximately 7% above its trend level. **Source:** Authors' calculations using CHRT and ONS data.

3.5 Comparison with Existing Estimates

Our results provide short-run estimates of the macroeconomic effects of firm entry shocks, complementing and extending earlier findings from lower-frequency data. Gourio, Messer, and Siemer (2016) estimate that a one-standard-deviation increase in firm entry (around 5%) leads to a cumulative increase in GDP of 0.23% after four years and 1.2% after

twelve years, using annual U.S. state-level data and a local projection framework. While our impulse responses are estimated at monthly and quarterly frequencies, the short-run magnitudes are broadly consistent with their long-run findings. For instance, we find that a firm entry shock raises GDP by 0.1–0.2% over a horizon of 8 to 12 quarters. These comparable magnitudes suggest that the entry-output elasticity is persistent and detectable even at high frequencies, underscoring the economic relevance of firm creation as a business cycle driver.

Our findings are consistent with Lewis (2009), who shows that firm entry rises following expansionary macroeconomic shocks. Lewis uses historical U.S. data and sign restrictions with a focus on monetary shocks, both nonetheless studies highlight firm entry as a key margin of macroeconomic adjustment with real effects on output and employment.

4 Conclusions

This paper demonstrates the potential of real-time administrative data for monitoring business dynamism and anticipating macroeconomic developments. We construct and validate a novel dataset—the Companies House Real-Time (CHRT) data—which records firm creation and dissolution activity at high frequency for the entire population of UK-incorporated companies. We show that CHRT entry data closely track official business demography statistics but become available far earlier, making them a valuable leading indicator for policy and forecasting.

Using monthly and quarterly aggregates, we document descriptive statistics in firm creation and estimate a Structural VAR to quantify the macroeconomic effects of entry shocks. We find that increases in firm entry lead to persistent and statistically significant gains in output, employment, and productivity—highlighting the role of firm formation as a driver of short-run economic fluctuations.

While the CHRT dataset is available at a daily frequency and disaggregated by industry and region, our analysis focuses on national-level monthly and quarterly aggregates. It emphasises firm entry due to its cleaner administrative definition compared to dissolutions. Future research could exploit the sectoral and regional dimensions of the data to examine whether certain industries—such as construction or retail—respond more strongly to entry shocks. Additionally, allowing for nonlinearities or interaction effects (e.g., examining whether entry shocks during recessions have different effects) would enrich our understanding of the cyclical role of new firm formation.

Beyond academic analysis, the CHRT data offer practical value for real-time policy monitoring. Our public dashboard or early-release indicator based on CHRT provides policymakers with a timely measure of business sentiment and capacity formation, particularly during crisis episodes. Real-time firm creation data should be integrated into

macroeconomic surveillance systems as a forward-looking signal of both structural and cyclical change in the economy.

References

Acemoglu, Daron, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr (2018). "Innovation, reallocation, and growth". In: *American Economic Review* 108.11, pp. 3450–3491.

Akcigit, Ufuk and Sina T Ates (2023). "What happened to US business dynamism?" In: *Journal of Political Economy* 131.8, pp. 2059–2124.

Asturias, Jose, Emin Dinlersoz, John Haltiwanger, and Rebecca Hutchinson (May 2023). "Are Business Applications Early Economic Indicators?" In: *AEA Papers and Proceedings* 113, pp. 151–55.

Bahaj, Saleem, Sophie Piton, and Anthony Savagar (2024). "Business Creation during COVID-19". In: *Economic Policy*.

Bayard, Kimberly, Emin Dinlersoz, Timothy Dunne, John Haltiwanger, Javier Miranda, and John Stevens (Mar. 2018). *Early-Stage Business Formation: An Analysis of Applications for Employer Identification Numbers*. Working Paper 24364. National Bureau of Economic Research.

Benedetti-Fasil, Cristiana, Petr Sedláček, and Vincent Sterk (May 2022). "Startups and employment following the COVID-19 pandemic: a calculator". In: *Economic Policy* 37.111, pp. 507–533.

Bilbiie, Florin O, Fabio Ghironi, and Marc J Melitz (2012). "Endogenous entry, product variety, and business cycles". In: *Journal of Political Economy* 120.2, pp. 304–345.

Biondi, Filippo, Sergio Inferrera, Matthias Mertens, and Javier Miranda (2023). *Declining business dynamism in Europe: The role of shocks, market power, and technology*. Tech. rep. Jena Economic Research Papers.

Calvino, Flavio, Chiara Criscuolo, and Rudy Verlhac (2020). "Declining business dynamism". In: *OECD Science, Technology and Industry Policy Papers* 94.

Department for Business and Trade (2023). *Business population estimates*.

Dinlersoz, Emin, Timothy Dunne, John Haltiwanger, and Veronika Penciakova (2021). "Business formation: A tale of two recessions". In: *AEA Papers and Proceedings*. Vol. 111, pp. 253–257.

Gourio, François, Todd Messer, and Michael Siemer (2016). "Firm entry and macroeconomic dynamics: a state-level analysis". In: *American Economic Review* 106.5, pp. 214–218.

Gutiérrez, Germán, Callum Jones, and Thomas Philippon (2021). "Entry Costs and Aggregate Dynamics". In: *Journal of Monetary Economics* 124, S77–S91.

Haltiwanger, John and Ryan Decker (2023). "Surging business formation in the pandemic: Causes and consequences?" In: *Brookings*.

Hopkins, Rhys (Aug. 2022). *Business classifications on the Inter-Departmental Business Register: 2022*.

Lewis, Vivien (2009). "Business cycle evidence on firm entry". In: *Macroeconomic Dynamics* 13.5, pp. 605–624.

Lewis, Vivien and Céline Poilly (2012). "Firm entry, markups and the monetary transmission mechanism". In: *Journal of Monetary Economics* 59.7, pp. 670–685.

Lewis, Vivien and Arnoud Stevens (2015). "Entry and markup dynamics in an estimated business cycle model". In: *European Economic Review* 74, pp. 14–35.

Lewis, Vivien and Roland Winkler (2017). "Government spending, entry, and the consumption crowding-in puzzle". In: *International Economic Review* 58.3, pp. 943–972.

Office for National Statistics (2022). *Registration lags on the UK Inter-Departmental Business Register between 2011 and 2021*. Accessed July 2025.

ONS (Sept. 2022). *Registration lags on the UK Inter-departmental Business Register: between 2011 and 2021*.

Sedláček, Petr (2020). "Lost generations of firms and aggregate labor market dynamics". In: *Journal of Monetary Economics* 111, pp. 16–31.

Sedláček, Petr and Vincent Sterk (Oct. 2017). "The Growth Potential of Startups over the Business Cycle". In: *American Economic Review* 107.10, pp. 3182–3210.

Sterk, Vincent, Petr Sedláček, and Benjamin Pugsley (Feb. 2021). "The Nature of Firm Growth". In: *American Economic Review* 111.2, pp. 547–79.

Tian, Can (2018). "Firm-level entry and exit dynamics over the business cycles". In: *European Economic Review* 102, pp. 298–326.

A Companies House Data Compilation

We construct a census of all companies incorporated in the UK since 2012. This tracks the lifecycle of every incorporated company in the UK in terms of its evolving industrial classification, location, as well as birth and dissolution. To construct the census, we use static snapshots of the Companies House register over regular intervals since 2012. For each company on the register in 2012, we record its status in each subsequent register entry, roughly annually. After initial entry onto the register, subsequent changes in future register releases include postcode and SIC industry classification changes. The census provides insight into the lifecycle of incorporated companies in the UK.

A.1 Overview

Census data are important to track lifecycle mobility or transitions often used in research. This is why we construct a census with publicly available Companies House (CH hereafter) data. This census will include companies incorporated in the UK. It is based on the snapshot "Free Company Data" released periodically from CH since July 2012 and regards all active companies at the time of release. The final dataset includes more than 10m companies that have been incorporated in the UK since that time. Our data output has two files:

1. **CH census panel**
   It includes information on the date of each register release, postcode, and all 5-digit SIC codes by Company Registration Number (CRN).

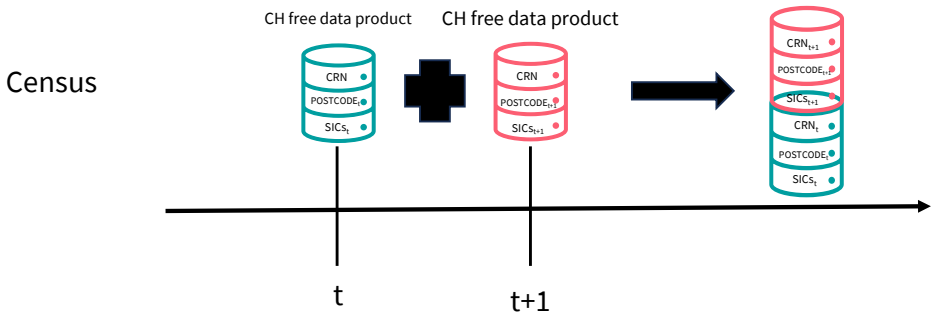


Figure A.1: Visual representation of CH census

Source: Authors' elaboration

2. **CH census flat-file (also known as masterfile)**
   It includes the Incorporation and Dissolution dates of each CRN.

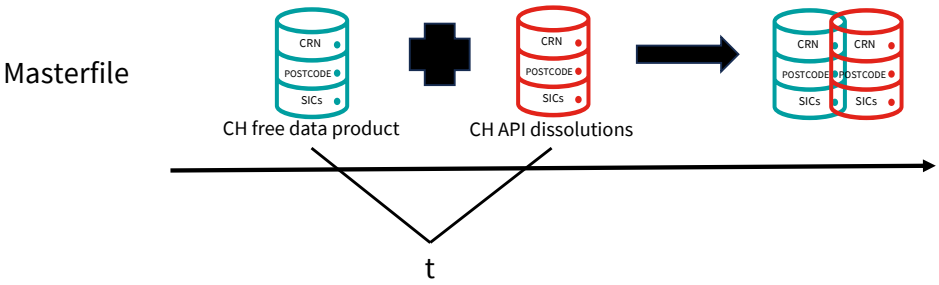


Figure A.2: Visual representation of CH masterfile

Source: Authors' elaboration

The split is necessary for data management purposes. Both outputs allow potential links with other administrative or survey firm-level data that use the same CRN identifier. In the main text we refer to both files as Companies House Real-time (CHRT) data.

Why company census data? Company census data are important for researchers, policymakers, and business owners. Researchers can inform about business dynamics. This census does not only allow an analysis of firm birth and death. It further allows analysis of changes in incumbent companies. Business owners may be informed about their competitors in nearby areas (same postcode and SIC code) and whether they changed the nature of their activity.

Benefits of our CH census data There are three benefits of using this CH census data. First, we can monitor over time changes beyond the company name. Indeed, the CH free bulk data include both the current and the previous Company Name. However, they do not include any previous postcodes or SIC codes. Annually, companies confirm their SIC code when they fill in the Annual Return/Confirmation Statement. Using this API, we can retrieve the information for each company. However, the Annual Return/Confirmation Statement API does not allow a search for multiple companies. Regarding the change of address, we can monitor changes of full address or just postcode by looking at the company filing history. The "filing history API" offers a bit more flexibility in terms of how many companies we can search at once, but is limited to 5,000 entries[9]. This dataset, though, overcomes these limitations. We can track changes of both SIC, primary and secondary, and postcode at the same time, by looking at the different month-year of register release.

Second, any commercial competitor (e.g. Financial Assets Made Easy (FAME) by Bureau van Dijk) does not have a dynamic setting. This is not true for institutions that

[9]Companies House API sets 5,000 entries constraint to efficiently operate its search engine.

buy the "historic FAME" data version. This dataset is not only free but it also monitors the change in company-reported address and nature of business. Users can link this dataset to FAME using the unique CRNs. By this way, FAME can become dynamic.

Finally, can this census dataset offer anything to the Inter-Departmental Business Register (IDBR)? IDBR holds two SIC measures – a "current" measure which can change month-on-month, and a "frozen SIC" that remains static throughout the year. Frozen SIC is updated in January of each year. At this point, the current SIC is carried forward to become the frozen SIC for the following year Hopkins (2022). This means that changes in IDBR are tracked backwards 18 months. The comparative advantage of this dataset is that it goes as back to July 2012, when the first online CH register is stored in UK Government Web Archive.

ONS (2022) analyse lags between appearing on the IDBR and HMRC records. They note that there are dates associated with VAT and PAYE birthdates which represent the date from which a business is liable to register. Companies House birthdates are available for 70.1% of IDBR entries. Companies House registration dates do not necessarily represent a start of economic activity. A business may register for Companies House at any point it wishes to adopt limited liability legal status. For example, a successful business might be known to HMRC many years before it decides to switch to limited liability status. Vice-versa a company may be registered with Companies House long before it qualifies for VAT or PAYE.

This section proceeds as follows: we document the data sources (subsection A.2), the methodology to compile the census (subsection A.3), and some additional descriptive statistics as of 1 May 2024 (subsection A.4). Data limitations can be found at the end (subsection A.6).

A.2 Data source

To construct the company census, we use the CH Free Company Data since July 2012. Available registers were released online through CH website in the past. They are currently stored at the UK Government Web Archive. Table A.1 presents the timing of releases that we take into account. There are 65 instances observed. Before 2017, the gaps in register availability are bigger.

There are two types of registers: registers released on CH website (on website `http://download.companieshouse.gov.uk/en_output.html`) (a) with non-broken and (b) broken links. Both types of registers are available through the National Archives Web Archive. Table A.1 outlines releases by type of links.

A.3 Methodology

This section describes how we download and process data from the Web Archive for both types of registers.

Table A.1: Type of links in UK Government Web Archive

Type of links	Month-year release
Non-broken links	Jul-12, Aug-12, Oct-12, Nov-12, Jan-13, Apr-13, Jul-13, Dec-13, Jan-14, Jun-14,Jul-14, Dec-14, May-16, Aug-17, Oct-17, Jan-18, Jul-18, Jan-19, Jul-19, Jan-20, Jun-20, Jul-20, Sep-20, Oct-20, Nov-20, Dec-20, Jan-21, Feb-21, Mar-21, Apr-21, May-21, Jun-21, Jul-21, Aug-21, Sep-21, Oct-21, Nov-21, Dec-21, Jan-22, Feb-22, Mar-22, Apr-22, May-22, Jun-22, Jul-22, Aug-22, Sep-22, Oct-22, Nov-22, Dec-22, Jan-23, Feb-23, Mar-23, May-23, Jun-23, Jul-23
Broken links, but restored	Mar-15, Oct-15, Oct-16, Sep-17
Broken links, not restored	Mar-18, Sep-19, Mar-20, Apr-20, May-20
Personal file*	Apr-23, Aug-23, Sep-23, Oct-23, Nov-23, Dec-2023, Jan-2024, Feb-2024, Mar-2024, Apr-2024, May-2024

* Data have been downloaded separately before this exercise.
Source: Authors' elaboration based on the links provided on the National Archives website

A.3.1 Registers from non-broken links

Step 1: The script fetches data for different months and years, starting from July 2012 up to December 2023.
Each register includes 55 columns[10]. However, we extract relevant columns and information from the Companies House data, such as company registration number, incorporation date, postcode, and SIC codes (both primary and secondary). Registers before 2016 are available in parts while later they become available as a single file.

Step 2: After downloading each register separately, we construct an additional variable called dataframe month-year. This is the month and year when the register was released. The algorithm extracts the date from the URL link it downloads for each register.

[10]The list of variables can be accessed on https://resources.companieshouse.gov.uk/toolsToHelp/pdf/freeDataProductDataset.pdf.

The final data frame has the following columns: CRN, Incorporation date, Register release date, primary and all secondary 5-digit SIC codes.

A.3.2 Registers from broken links

If one attempts to use the UK Government Web Archive timeline of CH, they realise that some links are broken. This part of the exercise restores access to most of these broken links. Before proceeding with the steps described in subsection 3.1, I download data *by hand*. Why? The links are broken because the URL registered on the web archive is different from the actual URL that stores the data.

Step 0: In more detail, say one wants to download data from the register released in March 2015. Clicking the URL on the current version of the Web Archive gives an error of missing page. To overcome this, first, we copy the link address whose form is: https://webarchive.nationalarchives.gov.uk/ukgwa/20150331172618mp_/http://download.companieshouse.gov.uk/BasicCompanyData-2015-03-01-part1_5.zip on a new browser tab. Second, we delete the part of the date and the suffix of the "part". This results in an *active* link that looks like: https://webarchive.nationalarchives.gov.uk/ukgwa/20150331172618mp_/http://download.companieshouse.gov.uk/BasicCompanyData-part1.zip. Then we repeat the same process for as remaining parts as there are and append them.

Why are links broken? Links are broken because CH at the time used a different format of link for the data. Hence, the transformation is necessary to obtain the data.

To receive a dataframe equivalent to subsection A.3.1, once we have appended all parts, we follow steps 1 and 2 as described in subsection A.3.1.

A.4 Descriptive statistics

A.4.1 Population and cleaning

There are 65 instances, i.e. months-years pairs, in which CH releases a register of active companies online. The size of each register changes over time because of the entry and exit (dissolution) of firms. For data management purposes, this entry of the CH census includes only these 65 instances and not any months in-between[11]. When a company does not include an incorporation date, we impute it using the incorporation date from the lag of the company's appearance in the census. This was necessary for 5 observations only. Further, I drop companies without a postcode or a primary SIC code. After the light cleaning exercise, CH census includes 10,255,282 unique companies (305,151,018

[11] A complete data frame should include all months between July 2012 and December 2023. This would yield a census of size 138 months x 5m entries = 690m observations, on average. This would require a substantially greater system capacity. This is why, 23 only keep the instances observed.

month-year of register release x companies pairs) that have an incorporation date, a valid postcode, and a primary SIC code until the registered released in May 2024.

Figure A.3 reports the number of active companies each month year of register release. Here, we follow the same definition of business activity as in the main text. See appendix **??** for additional definitions.

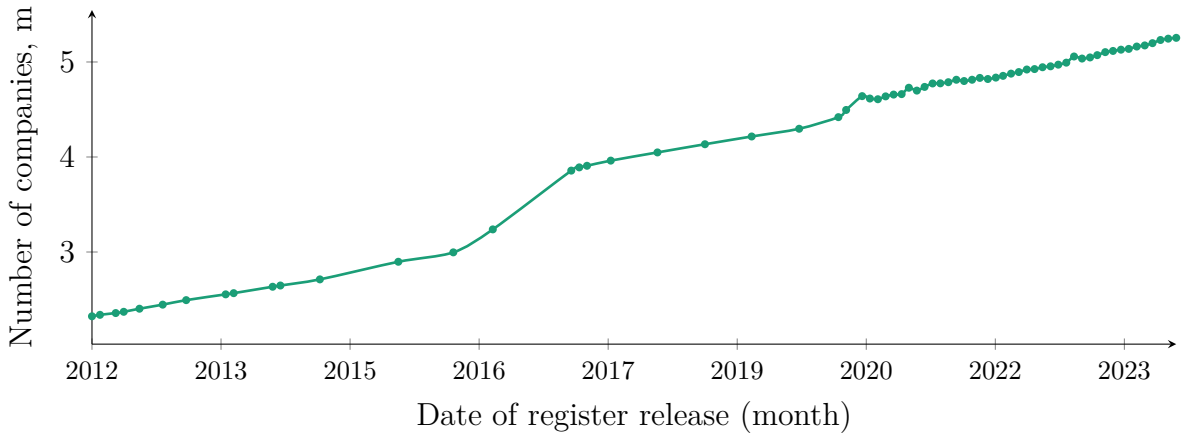


Figure A.3: Number of active companies, by month of register release

Notes: Number of companies in each CH register by date of register release without missing values in incorporation date, postcode, and primary SIC code. Points show the month of a CH register release. In more recent years, we hold almost monthly registers.
Source: Authors' elaboration based on Companies House Census census

Data validity Here we show that our constructed data match the official quarterly statistics release by Companies House. Figure A.4 plots the number of distinct active companies in each register by calendar quarter[12]. It has three lines: (a) the CH official statistics (total numbers on register) at the end of each period; (b) CH census without any cleaning and (c) CH census having dropped registrations with missing values in incorporation date, postcode or any SIC codes. Any differences when calculating the number of companies on the register at the end of the period are due to slight scheduling variations when extracting the data. We choose to visualise the end of the period because these are the publicly available company registrations. Given the differences between the CH official statistics release and CH census cleaned, we note that before 2017 registration recording was poorer.

[12]Note that CH statistics release is not in terms of calendar year, but in fiscal year. In the UK, the fiscal year commences in April.

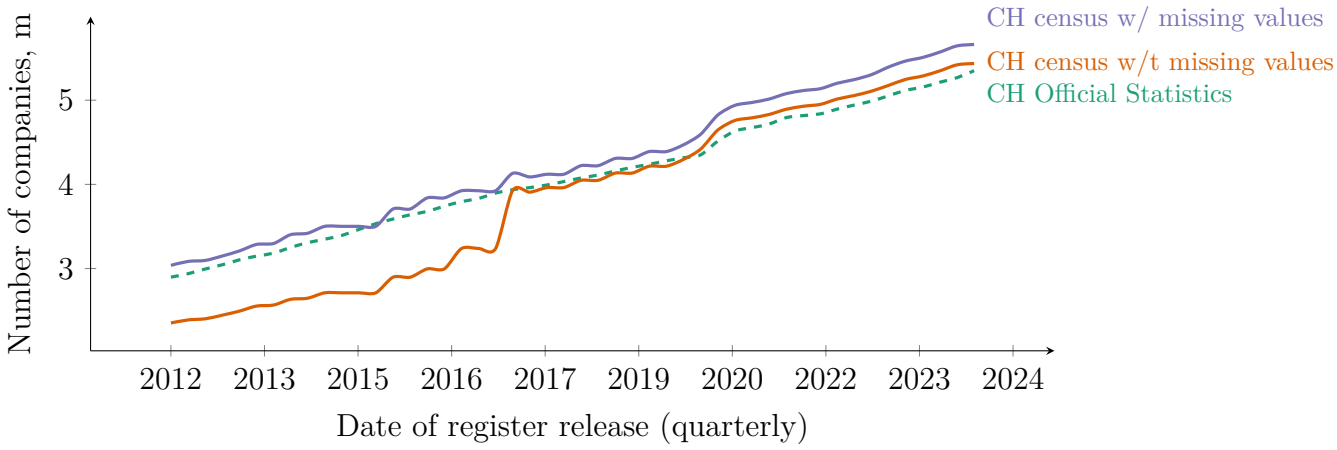


Figure A.4: CH quarterly statistics release vs. census: companies in register

Notes: Dashed line plots the total company number on the register at the end of each quarter as shown on the official quarterly CH statistics release. The dashed line excludes limited liability partnerships. The purple line plots the number of distinct active companies in our compiled census without any cleaning, i.e. including missing values. The orange line plots the number of distinct active companies in our compiled census without any missing values in the incorporation date, postcode, and SIC code(s).
Source: Authors' calculations based on Companies House Official Statistics and census.

A.4.2 Postcode Distributional Descriptive Statistics

This section describes how registered companies are distributed across all postcodes in the dataset, using the latest Companies House snapshot (1 May 2024). We classify postcodes by the number of companies they contain and report both counts and shares of total companies. Most firms are located in postcodes with relatively few companies: 44% are in postcodes with 1–10 firms, while just over one third are concentrated in postcodes with more than 100 firms. The distribution is highly skewed, with a small number of postcodes hosting very large numbers of companies; the most extreme case is postcode WC2H 9JQ, which contains over 91,000 registered firms.

To further characterise the spatial concentration of firms, we group postcodes into three categories: *single-company* (exactly one firm), *semi-single-company* (2–5 firms), and *multiple-company* (more than 5 firms). This classification highlights the prevalence of postcodes with minimal business activity alongside a smaller but significant share of highly concentrated business locations. Figures A.5–A.7 and Table A.2 present these patterns in detail.

Table A.2: Companies per postcode: counts and shares

Number of companies per postcode (grouped)	Companies (count)	Share of total
1–10	2,356,512	0.44
11–20	451,196	0.08
21–30	183,536	0.03
31–40	116,265	0.02
41–50	90,585	0.02
51–60	80,417	0.01
61–70	71,179	0.01
71–80	61,122	0.01
81–90	61,301	0.01
91–100	57,762	0.01
100+	1,856,469	0.34

Notes: Excludes companies without an incorporation date, postcode, or primary 5-digit SIC code. Data from Companies House register (1 May 2024). Postcodes refer to full units (e.g., CT2 7NZ).

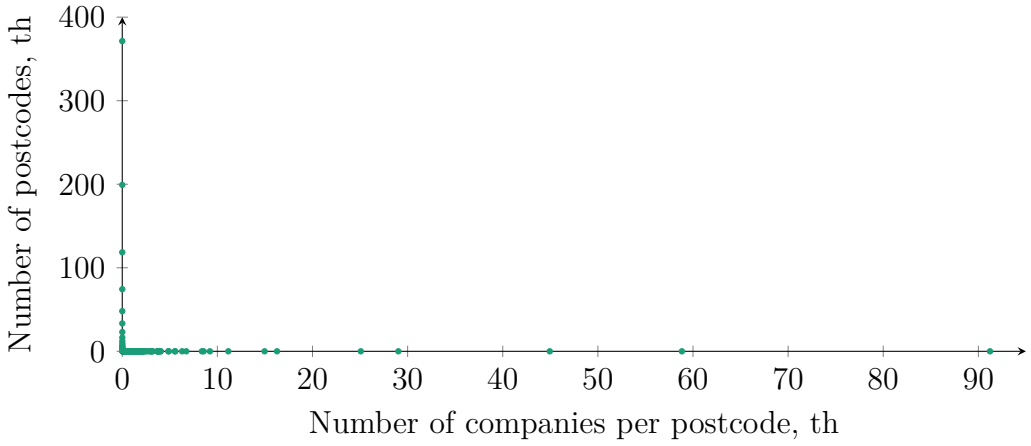


Figure A.5: Postcode-level distribution of company counts

Notes: Each dot shows the number of companies per postcode (x-axis) and the number of such postcodes (y-axis).

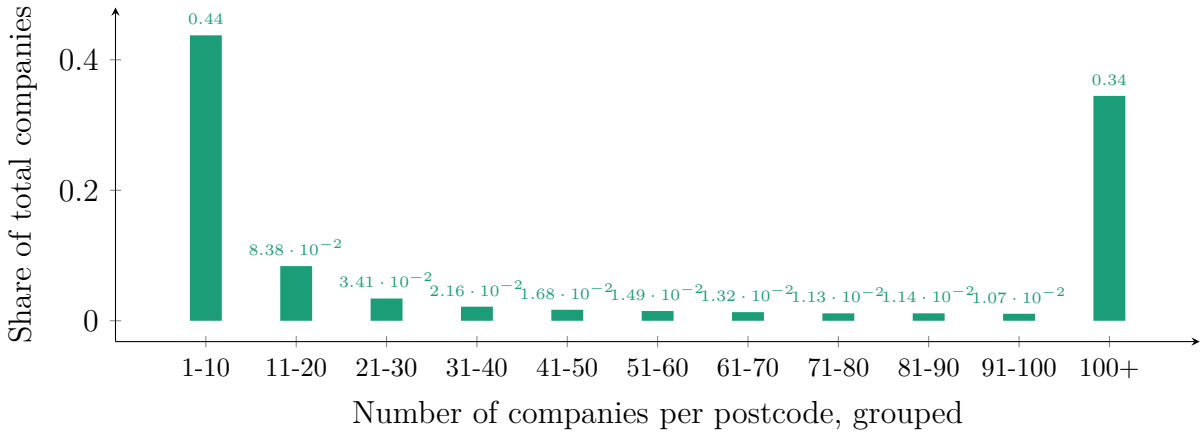


Figure A.6: Share of companies by postcode size group

Notes: Each bar shows the share of total companies in postcodes within the indicated size range.

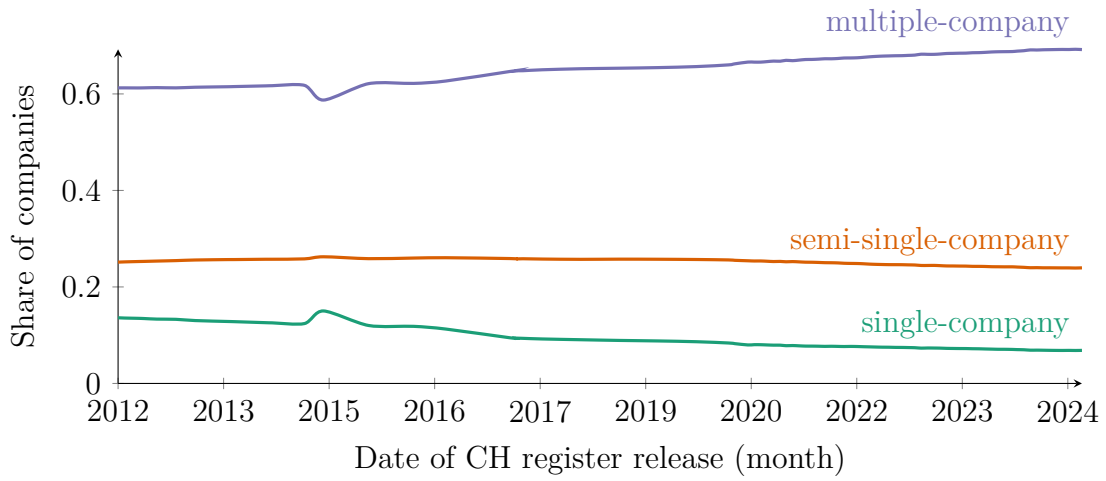
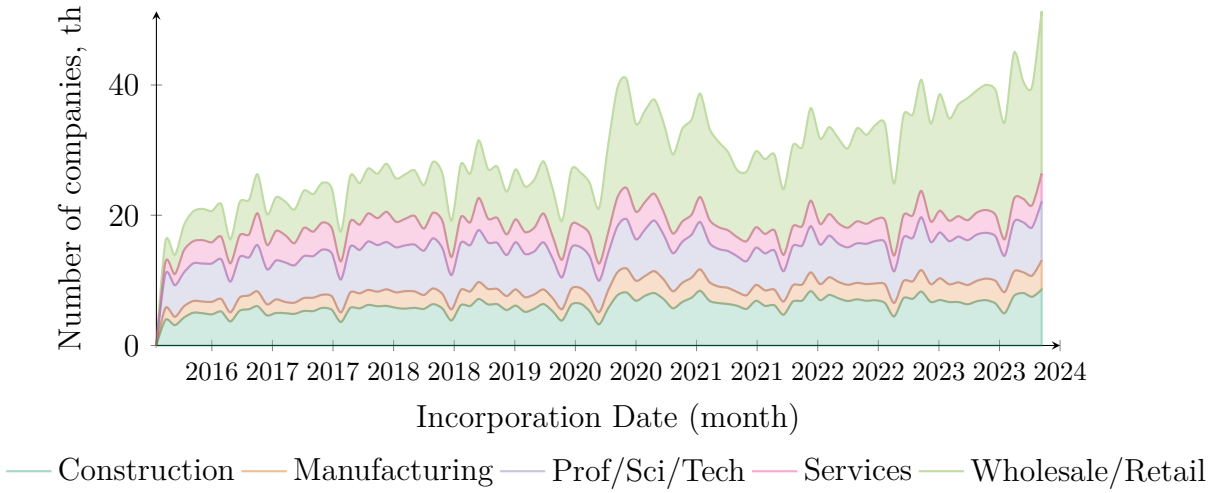


Figure A.7: Share of companies by postcode type

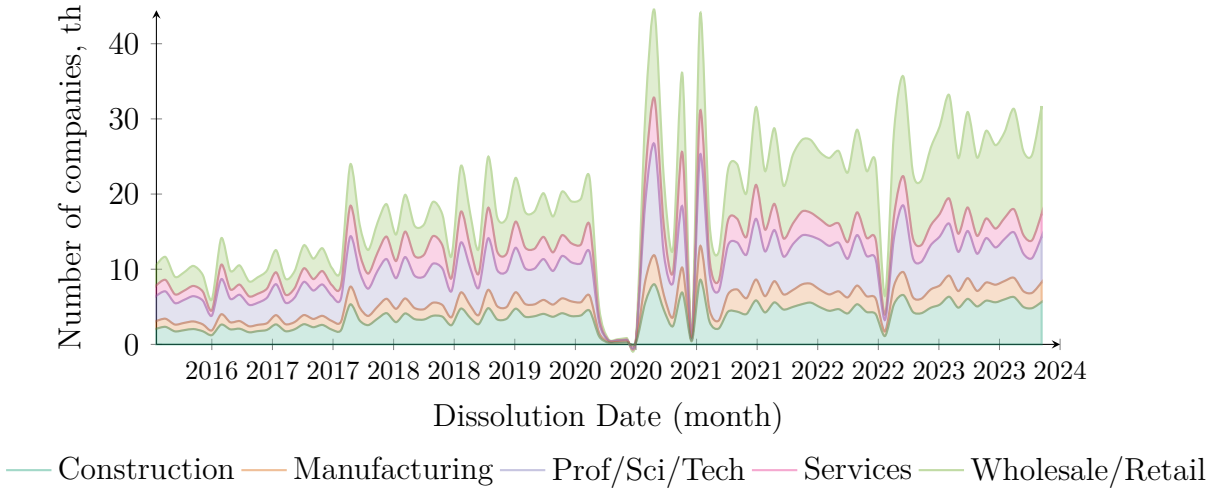
Notes: "Single-company" postcodes have 1 firm, "semi-single-company" 2–5 firms, and "multiple-company" more than 5 firms.

A.4.3 Sectoral decomposition

Our data offer the company-reported 5-digit SIC code. Figure A.8 plots the number of company incorporations and dissolutions for construction, manufacturing, Professional, Scientific and Technical Activities, Services, and Wholesale and Retail trade. The aggregation here is at the section level (the most aggregate industry level). Full breakdown is available upon request.



(a) Company incorporations



(b) Company dissolutions

Notes: We plot the dissolution by Section in dissolution. Section in dissolution may differ from Section in incorporation.

Figure A.8: Company incorporations and dissolution, by selected sectors

Notes: Company incorporations and dissolutions by Section (letter-digit) and month of incorporation for selected sectors. Sectors are matched on the *first* 5-digit SIC sector for each company. We plot here registrations from April 2016 onwards, because SIC recording improves after 2016. Construction are registrations in Section F; Manufacturing are registrations in Section C; Professional, Scientific and Technical Activities are Section M; Services refer to Other Service Activities and are registrations in Section S; Wholesale and Retail are registrations in Section G.
Source: Authors' elaboration based on CHRT data

A.5 Dynamics

A.5.1 Companies changing postcode and SIC codes

The dataset is first sorted by Company Registration Number (CRN) and by month of register release. For each company, we identify changes in postcode or primary SIC code between consecutive releases. Prior to 2020, the interval between available releases is

irregular and generally exceeds one month; from 2020 onward, releases are available at a monthly frequency.

Figure A.9 shows the average monthly rate at which active companies change postcode and/or primary SIC code. To account for varying gaps between register snapshots, we normalize by dividing the observed change rate by the number of months elapsed since the previous snapshot. This yields an average monthly change rate that is comparable across the entire time series. For example, if 6% of companies changed postcodes between two snapshots taken 6 months apart, the normalized rate would be 1% per month.

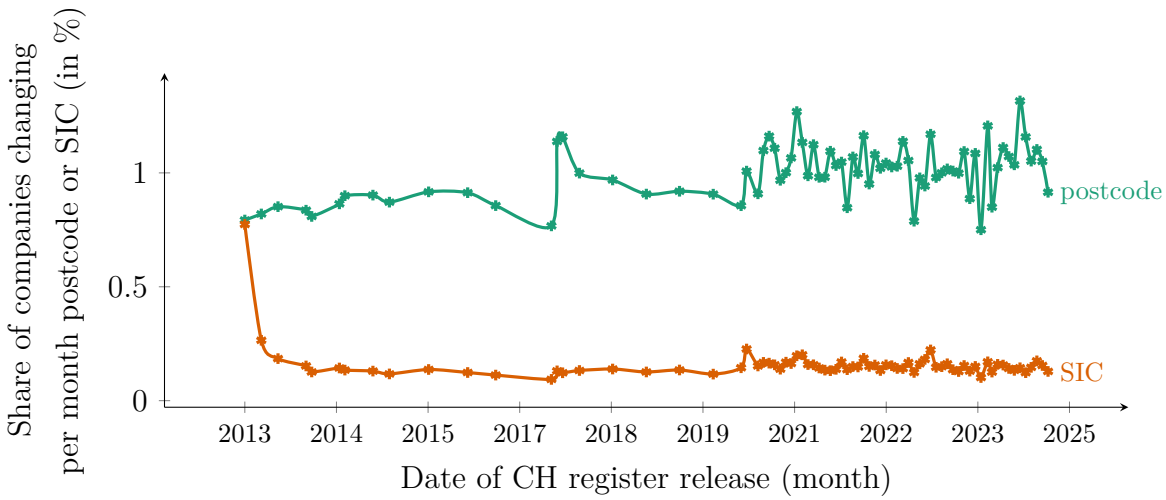


Figure A.9: Monthly average share of companies changing postcode or primary sector

Notes: We drop company-register release observations with missing values in Incorporation Date, postcode and primary SIC code. Each point represents the share of companies that changed their postcode (green line) or primary SIC code (orange line) between two adjacent CH register releasesbetween two adjacent CH register releases, divided by the number of months between releases to show the average monthly rate. For example, a value of 1% indicates that approximately 1% of active companies changed that attribute per month on average during the period. Both series are normalized to account for varying gaps between register snapshots. When registers are available monthly, the normalization has no effect.
Source: Authors' calculations based on Companies House Census data

A.6 Companies House Register Limitations

Companies House (CH) data are subject to several limitations relevant for analysis. First, there is a delay of approximately 2–3 months before CH removes from the register those companies that are (a) in the course of dissolution or (b) in liquidation. This delay may affect releases from 2017 onwards, when the publication frequency increased. To mitigate this issue, we rely on the *masterfile*, which records the exact dates of incorporation and dissolution.

Second, in response to the COVID-19 pandemic, CH implemented a temporary easement suspending voluntary strike-off action between March 2020 (the onset of the first national lockdown) and 10 September 2020. This policy was intended to prevent compa-

nies from being dissolved during the pandemic and results in fewer recorded dissolutions in our data.

Finally, the strike-off process was paused for five weeks from the end of October 2022 and subsequently resumed in a phased approach. According to the CH statistics team (via email correspondence), this was an operational decision taken in response to unforeseen circumstances.

B Other Data Sources of UK Business Dynamism

This section compares key UK data sources on business entry (birth) and exit (death). Definitions of a "business" differ across sources, so estimates are not directly comparable. Table B.1 contrasts definitions, disaggregation (geographic, industry, and other dimensions), and limitations. Table B.2 compares official statistics and microdata availability, as well as temporal coverage.

The ONS *Business Demography* series records only enterprise additions and removals from the IDBR. The Department for Business and Trade's *Business Population Estimates* (BPE) cover both registered and unregistered businesses, estimating the latter using self-employment data from the Labour Force Survey and tax returns. However, the BPE omits incorporated companies not registered for VAT or PAYE—these are assumed inactive unless listed on the IDBR. Department for Business and Trade (2023) propose a combined IDBR–Companies House approach to identify active incorporations excluded from the IDBR, suggesting that actual active business counts may be 4–32% higher than BPE figures.

Table B.1: Comparison of selected data sources: definitions, disaggregation, and limitations

	BIS Business Population Estimates Access here	**ONS UK Business: Activity, Size and Location** Access here	**ONS Business Demography** Annual release Experimental release	**Companies House Real-Time**
Measure & coverage	Population at start of calendar year (1 January), with associated employment and turnover.	Population at end of financial year (30 March).	Business births, deaths, and population 'active' at any point during the year.	Incorporated companies active at any point during the year.
"Business" definition	Enterprise and company[†,*]	Enterprise	Enterprise	Company
Business coverage	VAT and/or PAYE registered businesses plus estimated unregistered population.	VAT and/or PAYE registered businesses and local units	VAT and/or PAYE registered businesses.	Incorporated companies.
Data source	(a) ONS Inter-Departmental Business Register (IDBR), (b) ONS Labour Force Survey, (c) HMRC self-assessment tax data	ONS IDBR	ONS IDBR	Companies House registers, Companies House API, and Companies House accounts data
Timeliness	Released ∼10 months after reference date (January data in October).	Released ∼6 months after reference date (March data in Sept/Oct).	(a) Annual: ∼11 months after reference period, (b) Experimental: quarterly	API: daily; Accounts: almost daily; Registers: monthly
Geography	UK, with country and regional breakdowns.	UK, down to county/district and parliamentary constituency.	UK, down to county/district.	UK, down to full postcode.
Legal status	Private sector (companies, public corporations, sole proprietorships, partnerships).	Companies, public sector, nonprofits, partnerships, sole proprietors.	Companies, public corporations, nonprofits, partnerships, sole proprietors.	Companies formed under the Companies Act 2006 (public and private limited, limited by guarantee, unlimited, dormant, etc.).
Industry	SIC 2007, 3-digit level.	SIC 2007, 4-digit level.	SIC 2007, 4-digit level.	SIC 2007, 5-digit level.[‡]
Other disaggregations	Employee size band.	Age of business, turnover, and employment size band.	Survival of businesses.	(a) Survival, (b) Postcode and SIC panels (*company census*)
Exclusions	(a) Public administration, private households, extra-territorial entities, (b) Composite Management Service Companies, and firms registered at an Official Receiver's address	Composite Management Service Companies.	(a) Agriculture, public administration, (b) Composite Management Service Companies	None

[†] "Enterprises" are organisational units that may consist of multiple local units or plants. For consistency with Companies House entities, local units are referred to as "companies" here.

[*] The BPE excludes Companies House organisations not linked to the IDBR. Since 2018, this omission likely means BPE undercounts active private-sector businesses by 4–32%.

[‡] Companies House uses a condensed SIC list, adding two 5-digit codes for dormant (99999) and non-trading (74990) companies.

Source: Authors' elaboration based on https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/16418/guide_to_the_uk_business_population_and_demography.pdf.

Table B.2: Comparison of selected data sources: statistics, microdata, and time span

	BIS Business Population Estimates	**ONS UK Business: Activity, Size and Location**	**ONS Business Demography**	**Companies House Real-Time**
Statistics	Annual	Annual	Annual and quarterly (experimental)	Monthly*
Firm-level microdata	Restricted	Restricted	Restricted	Publicly available
Time span	Since 2010	Since 2014	Annual since 2013; quarterly since 2019	Registers since 2012; accounts since 2008

* Companies House official statistics are released quarterly.

Source: Authors' elaboration.